

Package ‘agridat’

November 30, 2017

Title Agricultural Datasets

Version 1.13

Date 2017-11-29

Type Package

Description Datasets from books, papers, and websites related to agriculture.

Example graphics and analyses are included. Data come from small-plot trials, multi-environment trials, uniformity trials, yield monitors, and more.

Suggests AER, agricolae, betareg, broom, car, coin, corrgram, desplot, dplyr, effects, equivalence, FrF2, gam, gge, ggplot2, gnm, gstat, HH, knitr, lattice, latticeExtra, lme4, lsmeans, lucid, mapproj, maps, MASS, MCMCglmm, metafor, mgcv, NADA, nlme, nullabor, ordinal, pbkrtest, pls, pscl, reshape2, rgdal, qicharts, qtl, sp, SpATS, survival, vcd, testthat

License CC BY-SA 4.0

LazyData yes

URL <https://github.com/kwstat/agridat>

BugReports <https://github.com/kwstat/agridat/issues>

VignetteBuilder knitr

RoxygenNote 6.0.1

NeedsCompilation no

Author Kevin Wright [aut, cre]

Maintainer Kevin Wright <kw.stat@gmail.com>

Repository CRAN

Date/Publication 2017-11-29 23:43:12 UTC

R topics documented:

| | |
|-------------------------------|----|
| aastveit.barley | 7 |
| acorsi.grayleafspot | 10 |
| adugna.sorghum | 12 |

| | |
|---------------------------------------|-----|
| agridat | 14 |
| allcroft.lodging | 22 |
| archbold.apple | 23 |
| ars.earlywhitecorn96 | 25 |
| australia.soybean | 26 |
| baker.barley.uniformity | 28 |
| batchelor.uniformity | 30 |
| battese.survey | 33 |
| beall.webworms | 35 |
| beaven.barley | 37 |
| becker.chicken | 38 |
| besag.bayesian | 39 |
| besag.beans | 41 |
| besag.elbatan | 43 |
| besag.endive | 45 |
| besag.met | 48 |
| besag.triticale | 53 |
| blackman.wheat | 55 |
| bliss.borers | 57 |
| bond.diallel | 58 |
| box.cork | 59 |
| bradley.multi.uniformity | 62 |
| brandle.rape | 64 |
| brandt.switchback | 66 |
| bridges.cucumber | 67 |
| broadbalk.wheat | 70 |
| burgueno.alpha | 72 |
| burgueno.rowcol | 76 |
| burgueno.unreplicated | 78 |
| butron.maize | 80 |
| byers.apple | 84 |
| caribbean.maize | 85 |
| carlson.germination | 87 |
| carmer.density | 88 |
| cate.potassium | 90 |
| chinloy.fractionalfactorial | 91 |
| christidis.competition | 93 |
| christidis.wheat.uniformity | 95 |
| cleveland.soil | 96 |
| cochran.beets | 97 |
| cochran.bib | 98 |
| cochran.crd | 99 |
| cochran.eelworms | 100 |
| cochran.factorial | 102 |
| cochran.latin | 103 |
| cochran.lattice | 104 |
| cochran.wireworms | 105 |
| connolly.potato | 106 |

| | |
|---------------------------------------|-----|
| cornelius.maize | 109 |
| corsten.interaction | 111 |
| cox.stripsplit | 112 |
| cramer.cucumber | 114 |
| crampton.pig | 116 |
| crossa.wheat | 118 |
| crowder.seeds | 120 |
| cullis.earlygen | 124 |
| darwin.maize | 128 |
| day.wheat.uniformity | 130 |
| denis.missing | 131 |
| denis.ryegrass | 132 |
| depalluel.sheep | 133 |
| digby.jointregression | 135 |
| diggle.cow | 137 |
| draper.safflower.uniformity | 139 |
| durban.competition | 141 |
| durban.rowcol | 142 |
| durban.splitplot | 145 |
| eden.nonnormal | 147 |
| eden.potato | 150 |
| engelstad.nitro | 152 |
| fan.stability | 153 |
| federer.diagcheck | 154 |
| federer.tobacco | 159 |
| fisher.barley | 160 |
| fisher.latin | 163 |
| foulley.calving | 164 |
| fox.wheat | 166 |
| garber.multi.uniformity | 167 |
| gartner.corn | 169 |
| gathmann.bt | 171 |
| gauch.soy | 173 |
| giles.wheat | 174 |
| gilmour.serpentine | 176 |
| gilmour.slatehall | 179 |
| gomez.fractionalfactorial | 181 |
| gomez.groupsplit | 183 |
| gomez.heteroskedastic | 184 |
| gomez.multilocsplitplot | 185 |
| gomez.nitrogen | 186 |
| gomez.nonnormal1 | 188 |
| gomez.nonnormal2 | 189 |
| gomez.nonnormal3 | 190 |
| gomez.rice.uniformity | 192 |
| gomez.seedrate | 193 |
| gomez.splitplot.subsample | 194 |
| gomez.splitsplit | 196 |

| | |
|--|-----|
| gomez.stripplot | 197 |
| gomez.stripsplitplot | 199 |
| gomez.wetdry | 200 |
| gotway.hessianfly | 202 |
| goulden.barley.uniformity | 204 |
| goulden.eggs | 205 |
| goulden.latin | 206 |
| goulden.splitsplit | 207 |
| graybill.heteroskedastic | 208 |
| gregory.cotton | 210 |
| gumpertz.pepper | 212 |
| hanks.sprinkler | 214 |
| hanover.whitepine | 216 |
| harris.multi.uniformity | 218 |
| harris.wateruse | 221 |
| harrison.priors | 225 |
| harvey.lsmmeans | 226 |
| harville.lamb | 228 |
| hayman.tobacco | 229 |
| hazell.vegetables | 232 |
| heady.fertilizer | 234 |
| henderson.milkfat | 236 |
| hernandez.nitrogen | 238 |
| hessling.argentina | 240 |
| hildebrand.systems | 242 |
| holland.arthropods | 245 |
| holshouser.splitstrip | 246 |
| holtsmark.timothy.uniformity | 248 |
| huehn.wheat | 249 |
| hughes.grapes | 251 |
| hunter.corn | 253 |
| ilri.sheep | 254 |
| immer.sugarbeet.uniformity | 256 |
| ivins.herbs | 258 |
| iyer.wheat.uniformity | 259 |
| jansen.apple | 261 |
| jansen.carrot | 262 |
| jansen.strawberry | 263 |
| jenkyn.mildew | 264 |
| john.alpha | 266 |
| johnson.blight | 270 |
| kadam.millet.uniformity | 271 |
| kalamkar.potato.uniformity | 273 |
| kalamkar.wheat.uniformity | 274 |
| kang.maize | 275 |
| kang.peanut | 277 |
| karcher.turfgrass | 278 |
| keen.potatodamage | 279 |

| | |
|--|-----|
| kempton.barley.uniformity | 281 |
| kempton.competition | 283 |
| kempton.rowcol | 284 |
| kempton.slatehall | 285 |
| kenward.cattle | 288 |
| khin.rice.uniformity | 291 |
| kiesselbach.oats.uniformity | 292 |
| kreusler.maize | 293 |
| kristensen.barley.uniformity | 296 |
| lambert.soiltemp | 298 |
| lasrosas.corn | 299 |
| lavoranti.eucalyptus | 301 |
| lee.potatobligh | 302 |
| lehner.soybeanmold | 305 |
| lessman.sorghum.uniformity | 307 |
| li.millet.uniformity | 309 |
| lillemo.wheat | 310 |
| lin.superiority | 312 |
| lin.unbalanced | 314 |
| little.splitblock | 315 |
| lonnquist.maize | 317 |
| love.cotton.uniformity | 319 |
| lu.stability | 320 |
| lucas.switchback | 322 |
| lyon.potato.uniformity | 324 |
| lyons.wheat | 325 |
| masood.rice.uniformity | 326 |
| mccllland.corn.uniformity | 327 |
| mconway.turnip | 328 |
| mcleod.barley | 330 |
| mead.cauliflower | 332 |
| mead.cowpeamaize | 333 |
| mead.germination | 335 |
| mead.lamb | 336 |
| mead.strawberry | 338 |
| mercerc.mangold.uniformity | 339 |
| mercerc.wheat.uniformity | 341 |
| minnesota.barley.weather | 342 |
| minnesota.barley.yield | 345 |
| montgomery.wheat.uniformity | 347 |
| moore.uniformity | 349 |
| nass.corn | 351 |
| nebraska.farmincome | 353 |
| nonnecke.corn.uniformity | 354 |
| nonnecke.peas.uniformity | 356 |
| odland.soybean.uniformity | 357 |
| omer.sorghum | 358 |
| onofri.winterwheat | 360 |

| | |
|---------------------------------------|-----|
| ortiz.tomato | 361 |
| pacheco.soybean | 363 |
| parker.orange.uniformity | 364 |
| patterson.switchback | 366 |
| pearce.apple | 368 |
| pearl.kernels | 369 |
| perry.springwheat | 370 |
| piepho.cocksfoot | 373 |
| polson.safflower.uniformity | 374 |
| ratkowsky.onions | 375 |
| reid.grasses | 376 |
| ridout.appleshoots | 378 |
| rothamsted.brussels | 380 |
| ryder.groundnut | 382 |
| salmon.bunt | 383 |
| sawyer.multi.uniformity | 385 |
| senshu.rice | 387 |
| shafii.rapeseed | 388 |
| silva.cotton | 390 |
| sinclair.clover | 393 |
| smith.beans.uniformity | 395 |
| smith.corn.uniformity | 397 |
| snedecor.asparagus | 399 |
| snijders.fusarium | 402 |
| stephens.sorghum.uniformity | 404 |
| steptoe.morex.pheno | 406 |
| stickler.sorghum.uniformity | 409 |
| stirret.borers | 411 |
| streibig.competition | 412 |
| stroup.nin | 413 |
| stroup.splitplot | 418 |
| student.barley | 421 |
| tai.potato | 423 |
| talbot.potato | 424 |
| theobald.barley | 426 |
| theobald.covariate | 427 |
| thompson.cornsoy | 428 |
| turner.herbicide | 431 |
| urquhart.feedlot | 432 |
| usgs.herbicides | 434 |
| vaneeuwijk.drymatter | 436 |
| vaneeuwijk.fusarium | 438 |
| vaneeuwijk.nematodes | 439 |
| vargas.txe | 441 |
| vargas.wheat1 | 444 |
| vargas.wheat2 | 446 |
| verbyla.lupin | 448 |
| vold.longterm | 452 |

| | |
|--|-----|
| vsn.lupin3 | 454 |
| wallace.iowaland | 457 |
| walsh.cottonprice | 459 |
| wassom.brome.uniformity | 460 |
| waynick.soil | 461 |
| wedderburn.barley | 462 |
| weiss.incblock | 464 |
| weiss.lattice | 465 |
| welch.bermudagrass | 467 |
| wheatley.carrot | 470 |
| wiebe.wheat.uniformity | 471 |
| wiedemann.safflower.uniformity | 473 |
| williams.barley.uniformity | 474 |
| williams.cotton.uniformity | 475 |
| williams.trees | 477 |
| yan.winterwheat | 478 |
| yang.barley | 479 |
| yates.missing | 480 |
| yates.oats | 482 |
| zuidhof.broiler | 485 |

Index**487**

| | |
|-----------------|--|
| aastveit.barley | <i>Barley heights and environmental covariates in Norway</i> |
|-----------------|--|

Description

Average height for 15 genotypes of barley in each of 9 years. Also 19 covariates in each of the 9 years.

Usage

```
data("aastveit.barley.covs")
data("aastveit.barley.height")
```

Format

The 'aastveit.barley.covs' dataframe has 9 observations on the following 20 variables.

```
year year
R1 avg rainfall (mm/day) in period 1
R2 avg rainfall (mm/day) in period 2
R3 avg rainfall (mm/day) in period 3
R4 avg rainfall (mm/day) in period 4
R5 avg rainfall (mm/day) in period 5
R6 avg rainfall (mm/day) in period 6
```

S1 daily solar radiation (ca/cm²) in period 1
S2 daily solar radiation (ca/cm²) in period 2
S3 daily solar radiation (ca/cm²) in period 3
S4 daily solar radiation (ca/cm²) in period 4
S5 daily solar radiation (ca/cm²) in period 5
S6 daily solar radiation (ca/cm²) in period 6
ST sowing date
T1 avg temp (deg Celsius) in period 1
T2 avg temp (deg Celsius) in period 2
T3 avg temp (deg Celsius) in period 3
T4 avg temp (deg Celsius) in period 4
T5 avg temp (deg Celsius) in period 5
T6 avg temp (deg Celsius) in period 6
value value of the covariate

The 'aastveit.barley.height' dataframe has 135 observations on the following 3 variables.

year year, 9

gen genotype, 15 levels

height height (cm)

Details

Experiments were conducted at As, Norway.

The height dataframe contains average plant height (cm) of 15 varieties of barley in each of 9 years.

The growth season of each year was divided into eight periods from sowing to harvest. Because the plant stop growing about 20 days after ear emergence, only the first 6 periods are included here.

Source

Aastveit, A. H. and Martens, H. (1986). ANOVA interactions interpreted by partial least squares regression. *Biometrics*, 42, 829–844. <http://doi.org/10.2307/2530697>

Used with permission of Harald Martens.

References

J. Chadoeuf and J. B. Denis (1991). Asymptotic variances for the multiplicative interaction model. *J. App. Stat.* 18, 331-353. <http://doi.org/10.1080/02664769100000032>

Examples

```

data("aastveit.barley.covs")
data("aastveit.barley.height")

if(require(reshape2) & require(pls)){

  # First, PCA of each matrix separately

  Z <- acast(aastveit.barley.height, year ~ gen, value.var="height")
  Z <- sweep(Z, 1, rowMeans(Z))
  Z <- sweep(Z, 2, colMeans(Z)) # Double-centered
  sum(Z^2)*4 # Total SS = 10165
  sv <- svd(Z)$d
  round(100 * sv^2/sum(sv^2),1) # Prop of variance each axis
  # Aastveit Figure 1. PCA of height
  biplot(prcomp(Z),
         main="aastveit.barley - height", cex=0.5)

  U <- aastveit.barley.covs
  rownames(U) <- U$year
  U$year <- NULL
  U <- scale(U) # Standardized covariates
  sv <- svd(U)$d
  round(100 * sv^2/sum(sv^2),1) # Proportion of variance on each axis

  ## Not run:
  # Now, PLS relating the two matrices
  m1 <- plsr(Z~U)
  loadings(m1)
  # Aastveit Fig 2a (genotypes), but rotated differently
  biplot(m1, which="y", var.axes=TRUE)
  # Fig 2b, 2c (not rotated)
  biplot(m1, which="x", var.axes=TRUE)

  ## End(Not run)

}

## Not run:
# Adapted from section 7.4 of Turner & Firth,
# "Generalized nonlinear models in R: An overview of the gnm package"
# who in turn reproduce the analysis of Chadoeuf & Denis (1991),
# "Asymptotic variances for the multiplicative interaction model"

require(agridat)
require(gnm)
data("aastveit.barley.height")
dath <- aastveit.barley.height
dath$year = factor(dath$year)

set.seed(1)

```

```

m2 <- gnm(height ~ year + gen + Mult(year, gen), data = dath)
# Turner: "To obtain parameterization of equation 1, in which sig_k is the
# singular value for component k, the row and column scores must be constrained
# so that the scores sum to zero and the squared scores sum to one.
# These contrasts can be obtained using getContrasts"
gamma <- getContrasts(m2, pickCoef(m2, "[.]y"),
                      ref = "mean", scaleWeights = "unit")
delta <- getContrasts(m2, pickCoef(m2, "[.]g"),
                      ref = "mean", scaleWeights = "unit")
# estimate & std err
gamma <- gamma$qvframe
delta <- delta$qvframe
# change sign of estimate
gamma[,1] <- -1 * gamma[,1]
delta[,1] <- -1 * delta[,1]
# conf limits based on asymptotic normality, Chadoeuf table 8, p. 350,
round(cbind(gamma[,1], gamma[, 1] +
            outer(gamma[, 2], c(-1.96, 1.96))) ,3)
round(cbind(delta[,1], delta[, 1] +
            outer(delta[, 2], c(-1.96, 1.96))) ,3)

## End(Not run)

```

acorsi.grayleafspot *Multi-environment trial evaluating 36 maize genotypes in 9 locations*

Description

Multi-environment trial evaluating 36 maize genotypes in 9 locations

Usage

```
data("acorsi.grayleafspot")
```

Format

A data frame with 324 observations on the following 3 variables.

gen genotype, 36 levels

env environment, 9 levels

rep replicate, 2 levels

y grey leaf spot severity

Details

Experiments conducted in 9 environments in Brazil in 2010-11. Each location had an RCB with 2 reps.

The response variable is the percentage of leaf area affected by gray leaf spot within each experimental unit (plot).

Acorsi et al. use this data to illustrate the fitting of a generalized AMMI model with non-normal data.

Source

C. R. L. Acorsi, T. A. Guedes, M. M. D. Coan, R. J. B. Pinto, C. A. Scapim, C. A. P. Pacheco, P. E. O. Guimaraes, C. R. Casela. (2016). Applying the generalized additive main effects and multiplicative interaction model to analysis of maize genotypes resistant to grey leaf spot. *Journal of Agricultural Science*. <http://doi.org/10.1017/S0021859616001015>

Electronic data and R code kindly provided by Marlon Coan.

References

None

Examples

```
data(acorsi.grayleafspot)
dat <- acorsi.grayleafspot

# Acorsi figure 2. Note: Acorsi used cell means
op <- par(mfrow=c(2,1), mar=c(5,4,3,2))
boxplot(y ~ env, dat, las=2,
        xlab="environment", ylab="GLS severity")
title("acorsi.grayleafspot")
boxplot(y ~ gen, dat, las=2,
        xlab="genotype", ylab="GLS severity")
par(op)

## Not run:

# GLM models

# glm main-effects model with logit u(1-u) and wedderburn u^2(1-u)^2
# variance functions
# glm1 <- glm(y~ env/rep + gen + env, data=dat, family=quasibinomial)
# glm2 <- glm(y~ env/rep + gen + env, data=dat, family=wedderburn)
# plot(glm2, which=1); plot(glm2, which=2)

# GAMMI models of Acorsi. See also section 7.4 of Turner
# "Generalized nonlinear models in R: An overview of the gnm package"

# full gnm model with wedderburn, seems to work
require(gnm)
```

```

set.seed(1)
gnm1 <- gnm(y ~ env/rep + env + gen + instances(Mult(env,gen),2),
            data=dat,
            family=wedderburn, iterMax =800)
deviance(gnm1) # 433.8548
# summary(gnm1)
# anova(gnm1, test ="F") # anodev, Acorsi table 4
##
##          Df Deviance Resid. Df Resid. Dev      F    Pr(>F)
## NULL
## env          8  1045.09      639   2310.4 68.4696 < 2.2e-16 ***
## env:rep       9   12.33      630   2298.1  0.7183  0.6923
## gen          35  1176.23      595   1121.9 17.6142 < 2.2e-16 ***
## Mult(env, gen, inst = 1) 42   375.94      553    745.9  4.6915 < 2.2e-16 ***
## Mult(env, gen, inst = 2) 40   312.06      513    433.9  4.0889 3.712e-14 ***

# maybe better, start simple and build up the model
gnm2a <- gnm(y ~ env/rep + env + gen,
             data=dat,
             family=wedderburn, iterMax =800)

# add first interaction term
res2a <- residSVD(gnm2a, env, gen, 2)
gnm2b <- update(gnm2a, . ~ . + Mult(env,gen,inst=1),
               start = c(coef(gnm2a), res2a[, 1]))
deviance(gnm2b) # 692.19

# add second interaction term
res2b <- residSVD(gnm2b, env, gen, 2)
gnm2c <- update(gnm2b, . ~ . + Mult(env,gen,inst=1) + Mult(env,gen,inst=2),
               start = c(coef(gnm2a), res2a[, 1], res2b[,1]))
deviance(gnm2c) # 433.8548
# anova(gnm2c) # weird error message

# note, to build the ammi biplot, use the first column of res2a to get
# axis 1, and the FIRST column of res2b to get axis 2. Slightly confusing
emat <- cbind(res2a[1:9, 1], res2b[1:9, 1])
rownames(emat) <- gsub("fac1.", "", rownames(emat))

gmat <- cbind(res2a[10:45, 1], res2b[10:45, 1])
rownames(gmat) <- gsub("fac2.", "", rownames(gmat))

# match Acorsi figure 4
biplot(gmat, emat, xlim=c(-2.2, 2.2), ylim=c(-2.2, 2.2), expand=2, cex=0.5,
       xlab="Axis 1", ylab="Axis 2",
       main="acorsi.grayleafspot - GAMMI biplot")

## End(Not run)

```

Description

Multi-environment trial of sorghum at 3 locations across 5 years

Format

A data frame with 289 observations on the following 6 variables.

gen genotype, 28 levels

trial trial, 2 levels

env environment, 13 levels

yield yield kg/ha

year year, 2001-2005

loc location, 3 levels

Details

Sorghum yields at 3 locations across 5 years. The trials were carried out at three locations in dry, hot lowlands of Ethiopia:

Melkassa (39 deg 21 min E, 8 deg 24 min N)

Mieso (39 deg 22 min E, 8 deg 41 min N)

Kobo (39 deg 37 min E, 12 deg 09 min N)

Trial 1 was 14 hybrids and one open-pollinated variety.

Trial 2 was 12 experimental lines.

Source

Asfaw Adugna (2008). Assessment of yield stability in sorghum using univariate and multivariate statistical approaches. *Hereditas*, 145, 28–37. <http://doi.org/10.1111/j.0018-0661.2008.2023.x>

Used with permission of Asfaw Adugna.

Examples

```
data(adugna.sorghum)
dat <- adugna.sorghum

require(lattice)
redblue <- colorRampPalette(c("firebrick", "lightgray", "#375997"))
levelplot(yield ~ env*gen, data=dat, main="adugna.sorghum gxe heatmap",
          col.regions=redblue)

# Genotype means match Adugna
tapply(dat$yield, dat$gen, mean)

# CV for each genotype. G1..G15 match, except for G2.
# The table in Adugna scrambles the means for G16..G28
if(require(reshape2)){
```

```

mat <- acast(dat, gen~env, value.var='yield')
round(sqrt(apply(mat, 1, var, na.rm=TRUE)) / apply(mat, 1, mean, na.rm=TRUE) * 100,2)

# Shukla stability. G1..G15 match Adugna. Can't match G16..G28.
dat1 <- droplevels(subset(dat, trial=="T1"))
mat1 <- acast(dat1, gen~env, value.var='yield')
w <- mat1; k=15; n=8 # k=p gen, n=q env
w <- sweep(w, 1, rowMeans(mat1, na.rm=TRUE))
w <- sweep(w, 2, colMeans(mat1, na.rm=TRUE))
w <- w + mean(mat1, na.rm=TRUE)
w <- rowSums(w^2, na.rm=TRUE)
sig2 <- k*w/((k-2)*(n-1)) - sum(w)/((k-1)*(k-2)*(n-1))
round(sig2/10000,1) # Genotypes in T1 are divided by 10000
}

```

agridat

Datasets from agricultural experiments

Description

This package contains datasets from publications relating to agriculture, including field crops, tree crops, animal studies, and a few others.

Details

If you use these data, please cite both the agridat package and the original source of the data.

Abbreviations in the 'other' column include: xy = coordinates, pls = partial least squares, rsm = response surface methodology, row-col = row-column design, ts = time series,

Uniformity trials with a single genotype

| name | dimensions | other | model |
|---|------------|-------------------|-----------|
| baker.barley.uniformity | 3 x 19 | xy, 10 years | |
| batchelor.apple.uniformity | 8 x 28 | xy | |
| batchelor.lemon.uniformity | 14 x 16 | xy | |
| batchelor.navell1.uniformity | 20 x 50 | xy | |
| batchelor.navell2.uniformity | 15 x 33 | xy | |
| batchelor.valencia.uniformity | 12 x 20 | xy | |
| batchelor.walnut.uniformity | 10 x 28 | xy | |
| bradley.multi.uniformity | 10 x 11 | xy | |
| christidis.wheat.uniformity | 12 x 24 | xy | |
| day.wheat.uniformity | 31 x 100 | xy | |
| draper.safflower.uniformity | 18 x 16 | xy, 2 expt | smith |
| garber.multi.uniformity | 45 x 6 | xy, 2 years/crops | |
| gomez.rice.uniformity | 18 x 36 | xy | aov,smith |
| goulden.barley.uniformity | 20 x 20 | xy | |
| harris.multi.uniformity | 2 x 23 | xy, 23 crops | corrgram |
| holtsmark.timothy.uniformity | 6 x 40 | xy | |

| | | | |
|------------------------------------|--------------|----------------------|---------------------|
| immer.sugarbeet.uniformity | 10 x 60 | xy, 3 traits | |
| iyer.wheat.uniformity | 25 x 80 | xy | |
| kadam.millet.uniformity | 8 x 20 | xy, 2 expts | |
| kalamkar.potato.uniformity | 6 x 96 | xy | |
| kalamkar.wheat.uniformity | 16 x 80 | xy, 2 traits | |
| kempton.barley.uniformity | 7 x 28 | xy | |
| khin.rice.uniformity | 30 x 36 | xy | |
| kiesselbach.oats.uniformity | 3 x 69 | xy | |
| kristensen.barley.uniformity | 22 x 11 | xy | |
| lessman.sorghum.uniformity | 46 x 60 | xy | |
| li.millet.uniformity | 6 x 100 | xy | |
| love.cotton.uniformity | 16 x 10 | xy | |
| lyon.potato.uniformity | 34 x 6 | xy | |
| masood.rice.uniformity | 12 x 24 | xy | |
| mcclelland.corn.uniformity | 2 x 219 | xy | |
| mercercor.mangold.uniformity | 10 x 20 | xy | |
| mercercor.wheat.uniformity | 25 x 20 | xy, 2 traits | spplot |
| montgomery.wheat.uniformity | 14 x 16 | xy, 2 years | lm |
| moore.polebean.uniformity | 12 x 12 | xy | |
| moore.bushbean.uniformity | 24 x 24 | xy | |
| moore.sweetcorn.uniformity | 24 x 12 | xy | |
| moore.carrots.uniformity | 24 x 12 | xy | |
| moore.springcauliflower.uniformity | 12 x 20 | xy | |
| moore.fallcauliflower.uniformity | 12 x 20 | xy | |
| nonnecke.corn.uniformity | 32 x 18 | xy, 3 loc | |
| nonnecke.peas.uniformity | 15 x 18 | xy, 2 traits | |
| odland.soybean.uniformity | 25 x 42 | xy | |
| odland.soyhay.uniformity | 28 x 55 | xy | |
| parker.orange.uniformity | 10 x 27 | xy, 6 yr | |
| polson.safflower.uniformity | 52 x 33 | xy | smith |
| sawyer.multi.uniformity | 8 x 6 | xy, 3 year | |
| smith.beans.uniformity | 18x12, 16x15 | xy, 2 yr, 2 crops | |
| smith.corn.uniformity | 6 x 20 | xy, 3 years | rgl |
| stephens.sorghum.uniformity | 100 x 20 | xy | |
| stickler.sorghum.uniformity | 20 x 20 | xy, 4 expts, 2 years | |
| wassom.brome.uniformity | 36 x 36 | xy, 3 expts | |
| wiebe.wheat.uniformity | 12 x 125 | xy | medianpolish, loess |
| wiedemann.safflower.uniformity | 54 x 33 | xy | smith |
| williams.barley.uniformity | 48 x 15 | xy | loess |
| williams.cotton.uniformity | 24 x 12 | xy | loess |

Yield monitor

| name | reps | years | trt | other | model |
|---------------|------|-------|-----|-------|-------|
| gartner.corn | | | | xy,ym | |
| lasrosas.corn | 3 | 2 | 6 | xy,ym | lm |

Animals

| name | gen | years | trt | other | model |
|----------------------|------|-------|-----|--------------|----------------|
| becker.chicken | 5,12 | | | heritability | lmer |
| crampton.pig | | | 5 | 2 cov | lm |
| brandt.switchback | 10 | | 2 | | aov |
| depalluel.sheep | 4 | | 4 | latin | |
| diggle.cow | | | 4 | ts | |
| foulley.calving | | | | ordinal | polr |
| goulden.eggs | | | | controlchart | |
| harvey.lsmeans | 3,3 | | | | lm |
| harville.lamb | 5 | | | | lmer |
| henderson.milkfat | | | | | nls,lm,glm,gam |
| holland.arthropods | 5 | | | | |
| ilri.sheep | 4 | 6 | | diallel | lmer, asreml |
| kenward.cattle | | | 2 | | asreml |
| lucas.switchback | 12 | | 3 | | aov |
| mead.lamb | 3 | | 3 | | glm |
| patterson.switchback | 12 | | 4 | | aov |
| urquhart.feedlot | 11 | | 3 | | lm |
| zuidhof.broiler | | | | ts | |

Trees

| name | gen | loc | reps | years | trt | other | model |
|----------------------|-----|-----|------|-------|-----|--------------|----------------|
| box.cork | | | | | | repeated | radial, asreml |
| harris.wateruse | 2 | | | | 2 | repeated | asreml,lme |
| hanover.whitepine | 7*4 | | 4 | | | heritability | lmer |
| lavoranti.eucalyptus | 70 | 7 | | | | | svd |
| pearce.apple | | | 4 | | 6 | cov | lm,lmer |
| williams.trees | 37 | 6 | | | 2 | | |

Field and horticulture crops

| name | gen | loc | reps | years | trt | other | model |
|----------------------|-----|-----|------|-------|-----|-----------------|-------------|
| acorsi.grayleafspot | 36 | 9 | 2 | 5 | | nonnormal | gnm,ammi |
| adugna.sorghum | 28 | 13 | | 5 | | | |
| aastveit.barley | 15 | | | 9 | | yr*gen~yr*trt | pls |
| allcroft.lodging | 32 | 7 | | | | percent | tobit |
| archbold.apple | 2 | | 5 | | 24 | split-split | lmer |
| ars.earlywhitecorn96 | 60 | 9 | | | | 6 traits | dotplot |
| australia.soybean | 58 | 4 | | 2 | | 4-way, 6 traits | biplot |
| battese.survey | | 12 | 1-5 | | 2 | | lmer |
| beall.webworms | | | 15 | | 2,2 | xy | glm poisson |
| beaven.barley | 8 | | 20 | | | xy | |
| besag.bayesian | 75 | | 3 | | | xy | asreml |

| | | | | | | |
|-----------------------------|-----|----|-----|---------|-------------------|-----------------------|
| besag.beans | 6 | | 4*6 | | xy | lm,competition |
| besag.elbatan | 50 | | 3 | | xy | lm, gam |
| besag.endive | | | | | xy,binary | autologistic |
| besag.met | 64 | 6 | 3 | | xy, incblock | asreml, lme |
| besag.triticale | 3 | | | 2,2,3 | xy | lm, asreml |
| bliss.borers | | | | 4 | | glm |
| blackman.wheat | 12 | 7 | | 2 | | biplot |
| bond.diallel | 6*6 | | | 9 | diallel | |
| bridges.cucumber | 4 | 2 | 4 | | xy, latin, hetero | asreml |
| brandle.rape | 5 | 9 | 4 | 3 | | lmer |
| burgueno.alpha | 15 | | 3 | | xy, alpha | asreml,lmer |
| burgueno.rowcol | 64 | | 2 | | xy, row-col | asreml,lmer |
| burgueno.unreplicated | 280 | | | | xy | asreml |
| butron.maize | 49 | 3 | | 2 | diallel,pedigree | biplot,asreml |
| caribbean.maize | | 17 | 4 | 3 | | |
| carmer.density | 8 | | | 4 | | nls,nlme |
| carlson.germination | 15 | | | 8 | | glm |
| chinloy.fractionalfactorial | | | 9 | 1/3 3^5 | xy | aov |
| christidis.competition | 9 | | 5 | | xy | |
| cochran.beets | | | 6 | 7 | | |
| cochran.bib | 13 | 13 | | | bib | aov, lme |
| cochran.crd | | | | 7 | xy, crd | aov |
| cochran.factorial | | | 2 | 4^2 | factorial | aov |
| cochran.latin | | | 6 | 6 | xy, latin | aov |
| cochran.lattice | | | 5 | 16 | xy, latin | lmer |
| cochran.wireworms | | | 5 | 5 | xy, latin | glm |
| cochran.eelworms | | | 4 | 5 | xy | aov |
| connolly.potato | 20 | | 4 | | xy, competition | lm |
| cornelius.maize | 9 | 20 | | | | svd |
| corsten.interaction | 20 | 7 | | | | |
| cramer.cucumber | | | 8 | | | pathcoef |
| crossa.wheat | 18 | 25 | | | | ammi |
| crowder.seeds | 2 | | 21 | 2 | | glm,jags |
| cox.stripssplit | | | 4 | 3,4,2 | split-block | aov |
| cullis.earlygen | 532 | | | | xy | asreml |
| darwin.maize | | | 12 | 2 | | t.test |
| denis.missing | 5 | 26 | | | | lme |
| denis.ryegrass | 21 | 7 | | | | aov |
| digby.jointregression | 10 | 17 | | 4 | | lm |
| durban.competition | 36 | | 3 | | xy, competition | lm |
| durban.rowcol | 272 | | 2 | | xy | lm, gam, asreml |
| durban.splitplot | 70 | | 4 | 2 | xy | lm, gam, asreml |
| eden.potato | | | 4 | 3 | 4-12 | xy, rcb, latin |
| eden.nonnormal | | | 4 | 4 | | aov |
| engelstad.nitro | | 2 | 5 | 6 | rsm1 | nls quadratic plateau |
| fan.stability | 13 | 10 | | 2 | 3-way | stability |
| federer.diagcheck | 122 | | | | xy | lm, lmer, asreml |
| federer.tobacco | | | 8 | 7 | xy | lm |

| | | | | | | | |
|---------------------------|-----|----|---|----|---------|------------------|-----------------|
| fisher.barley | 5 | 6 | | 2 | | | |
| fisher.latin | | | 5 | | 5 | xy,latin | lm |
| fox.wheat | 22 | 14 | | | | | lm |
| gathmann.bt | 2 | | 8 | | | | tost |
| gauch.soy | 7 | 7 | 4 | 12 | | | ammi |
| giles.wheat | 19 | | | 13 | | 2 traits | gnm |
| gilmour.serpentine | 108 | | 3 | | | xy, serpentine | asreml |
| gilmour.slatehall | 25 | | 6 | | | xy | asreml |
| gomez.fractionalfactorial | | | 2 | | 1/2 2^6 | xy | lm |
| gomez.groupsplit | 45 | | 3 | | 2 | xy, 3 gen groups | aov |
| gomez.heteroskedastic | 35 | | 3 | | | hetero | |
| gomez.multilocsplitplot | 2 | 3 | 3 | | | rsm1,nitro | aov, lmer |
| gomez.nitrogen | | | 4 | | 8 | | aov, contrasts |
| gomez.nonnormal1 | | | 4 | | 9 | log10 | lm |
| gomez.nonnormal2 | 14 | | 3 | | | sqrt | lm |
| gomez.nonnormal3 | 12 | | 3 | | | arcsin | lm |
| gomez.seedrate | | | 4 | | 6 | rate | lm |
| gomez.splitplot.subsample | | | 3 | | 8,4 | subsample | aov |
| gomez.splitsplit | 3 | | 3 | | | xy, nitro, mgmt | aov, lmer |
| gomez.stripplot | 6 | | 3 | | | xy, nitro | aov |
| gomez.stripsplitplot | 6 | | 3 | | | xy, nitro | aov |
| gomez.wetdry | | | 3 | 2 | 5 | nitro | lmer |
| gotway.hessianfly | 16 | | 4 | | | xy | lmer |
| goulden.latin | | | 5 | | 5 | xy, latin | lm |
| goulden.splitsplit | 2 | | 4 | | 2*5 | xy, split | aov |
| graybill.heteroskedastic | 4 | 13 | | | | hetero | |
| gregory.cotton | | | | 2 | 4*3*2*2 | polar | |
| gumpertz.pepper | | | | | | xy | glm |
| hanks.sprinkler | 3 | | 3 | | | xy | asreml |
| hayman.tobacco | 8 | | 2 | 2 | | diallel | asreml |
| hazell.vegetables | 4 | | | 6 | | | linprog |
| heady.fertilizer | | | 2 | | 9*9 | rsm2 | lm,rgl |
| hernandez.nitrogen | | 5 | 4 | | | rsm1 | lm, nls |
| hildebrand.systems | | 14 | | | 4 | | asreml |
| holshouser.splitstrip | 4 | | 4 | | 2*4 | rsm1,pop | lmer |
| huehn.wheat | 20 | 10 | | | | | huehn |
| hughes.grapes | | | 3 | | 6 | binomial | lmer, aod, glmm |
| hunter.corn | | 12 | | 3 | 1 | rsm1 | xyplot |
| ivins.herbs | 13 | | 6 | | | 2 traits | lm, friedman |
| jansen.apple | 3 | | 4 | | 3 | binomial | glmer |
| jansen.carrot | 16 | | 3 | | 2 | binomial | glmer |
| jansen.strawberry | 12 | | 4 | | | ordinal | mosaicplot |
| jenkyn.mildew | | | 9 | | 4 | | lm |
| john.alpha | 24 | | 3 | | | alpha | lm, lmer |
| johnson.blight | | | | 2 | | | logistic |
| kang.maize | 17 | 4 | | 3 | 2,4 | | |
| kang.peanut | 10 | 15 | 4 | | | | gge |
| karcher.turfgrass | | | 4 | | 2,4 | ordinal | polr |

| | | | | | | |
|--------------------------|-----|----|----|-------|-----------------------|------------------------|
| keen.potatodamage | 6 | 4 | | 2,3,8 | ordinal | mosaicplot |
| kempton.competition | 36 | 3 | | | xy, competition | lme AR1 |
| kempton.rowcol | 35 | 2 | | | xy, row-col | lmer |
| kempton.slatehall | 25 | 6 | | | xy | asreml, lmer |
| lee.potatobligh | 337 | 4 | 11 | | xy, ordinal, repeated | |
| lehner.soybeanmold | | 35 | 4 | 11 | | metafor, lmer |
| lillemo.wheat | 24 | 13 | 7 | | qq | medpolish, huehn |
| lin.superiority | 33 | 12 | | | | superiority |
| lin.unbalanced | 33 | 18 | | | | superiority |
| little.splitblock | | | 4 | 4,5 | xy, split-block | aov |
| lonnquist.maize | 11 | | | | diallel | asreml |
| lyons.wheat | | 12 | 4 | | | |
| lu.stability | 5 | 6 | | | | huehn |
| mconway.turnip | 2 | | 4 | 2,4 | hetero | aov, lme |
| mcleod.barley | | 8 | | 6 | | aggregate |
| mead.cauliflower | | | 2 | | poisson | glm |
| mead.cowpeamaize | 3,2 | | 3 | 4 | intercrop | |
| mead.germination | | | 4 | 4,4 | binomial | glm |
| mead.strawberry | 8 | 4 | | | | |
| minnesota.barley.weather | | 6 | 10 | | | |
| minnesota.barley.yield | 22 | 6 | 10 | | | dotplot |
| omer.sorghum | 18 | 2 | 4 | 3 | | jags |
| onofri.winterwheat | 8 | | 3 | 7 | | ammi |
| ortiz.tomato | 15 | 18 | | 16 | env*gen~env*cov | pls |
| pacheco.soybean | 18 | 11 | | | | ammi |
| perry.springwheat | 28 | 5 | 4 | | gain | lm,lmer,asreml |
| piepho.cocksfoot | 25 | | 7 | | | lmer |
| ratkowsky.onions | | | | | | lm |
| reid.grasses | 4 | | 3 | 21 | | nlme SSfpl |
| ridout.appleshoots | | | 30 | 2,4 | zip | zeroinfl |
| rothamsted.brussels | | | 4 | 6 | | |
| ryder.groundnut | 5 | | 4 | | xy, rcb | lm |
| salmon.bunt | 10 | | 2 | 20 | | betareg |
| senshu.rice | | 40 | | | | lm,Fieller |
| shafii.rapeseed | 6 | 14 | 3 | 3 | | biplot |
| silva.cotton | | | 5 | 5 | 5 traits | glm,poisson |
| sinclair.clover | | | | 5,5 | rsm2,mitzerlich | nls,rgl |
| snedecor.asparagus | | | 4 | 4 | | split-plot, antedepend |
| snijders.fusarium | 17 | | 3 | 4 | percent | glm/gnm,gammi |
| steptoe.morex.pheno | 152 | 16 | | | 10 traits | |
| steptoe.morex.geno | 150 | | | | 223 markers, qtl | |
| streibig.competition | 2 | | 3 | | | glm |
| stroup.nin | 56 | | 4 | | xy | asreml |
| stroup.splitplot | | | 4 | | | asreml, MCMCglmm |
| student.barley | 2 | 51 | 6 | | | lmer |
| tai.potato | 8 | 3 | 2 | | | tai |
| talbot.potato | 9 | 12 | | | gen*env~gen*trt | pls |
| theobald.barley | 3 | 5 | 2 | 5 | rsm1 | |

| | | | | | | | |
|----------------------|-----|----|-----|----|-------|-------------------------------|--------------|
| theobald.covariate | 10 | 7 | 5 | | | cov | jags |
| thompson.cornsoy | | 5 | | 33 | | repeated measures | aov |
| vaneeuwijk.fusarium | 20 | | | 4 | 7 | 3-way | aov |
| vaneeuwijk.drymatter | 6 | 4 | | 7 | | 3-way | aov,lmer |
| vaneeuwijk.nematodes | 11 | | | | | nonnormal,poisson | gnm, gammi |
| vargas.wheat1 | 7 | | | 6 | | gen*yr~gen*trt, yr*gen~yr*cov | pls |
| vargas.wheat2 | 8 | 7 | | | | env*gen~env*cov | pls |
| vargas.txe | | | | 10 | 24 | yr*trt~yr*cov | pls |
| verbyla.lupin | 9 | 8 | 2-3 | 2 | | rsm1, xy, density | asreml |
| vold.longterm | | | | 19 | 4 | rsm1 | nls,nlme |
| vsn.lupin3 | 336 | 3 | | | | xy | asreml |
| wedderburn.barley | 10 | 9 | | | | percent | glm/gnm |
| weiss.incblock | 31 | | 6 | | | xy,incblock | asreml |
| weiss.lattice | 49 | | 4 | | | xy,lattice | lm,asreml |
| welch.bermudagrass | | | | | 4,4,4 | rsm3, factorial | lm, jags |
| wheatley.carrot | | | 3 | | 11 | | glm-binomial |
| yan.winterwheat | 18 | 9 | | | | | gge,biplot |
| yang.barley | 6 | 18 | | | | | biplot |
| yates.missing | | | 10 | | 3^2 | factorial | lm, pca |
| yates.oats | 3 | | 6 | | | xy,split,nitro | lmer |

Time series

| name | years | trt | other | model |
|--------------------|-------|-----|--------------|-------|
| byers.apple | | | | lme |
| broadbalk.wheat | 74 | 17 | | |
| hessling.argentina | 30 | | temp,precip | |
| kreusler.maize | 4 | 5 | plant growth | |
| lambert.soiltemp | 1 | 7 | | |
| nass.barley | 146 | | | |
| nass.corn | 146 | | | |
| nass.cotton | 146 | | | |
| nass.hay | 104 | | | |
| nass.sorghum | 93 | | | |
| nass.wheat | 146 | | | |
| nass.rice | 117 | | | |
| nass.soybean | 88 | | | |
| walsh.cottonprice | 34 | | | cor |

Other

| name | model |
|---------------------|-------------|
| cate.potassium | cate-nelson |
| cleveland.soil | loess 2D |
| harrison.priors | nls, prior |
| nebraska.farmincome | choropleth |

| | |
|----------------------------------|-----------------------|
| pearl.kernels | chisq |
| stirret.borers | lm, 4 trt |
| turner.herbicide | glm, 4 trt |
| usgs.herbicides | non-detect |
| wallace.iowaland | lm, choropleth |
| waynick.soil | spatial, nitro/carbon |

Comments of the package purpose

Box (1957) said, "I had hoped that we had seen the end of the obscene tribal habit practiced by statisticians of continually exhuming and massaging dead data sets after their purpose in life has long since been forgotten and there was no possibility of doing anything useful as a result of this treatment."

Massaging these dead data sets will not lead to any of the genetics being released for commercial use. The value of this package is: 1. Validating published analyses. 2. Providing data for testing new analysis methods. 3. Illustrating (and validating) the use of R.

White and van Evert (2008) present some guidelines for publication of data.

Some of the examples use the `asreml` package since it is the `_only_ R` tool for fitting mixed models with complex variance structures to large datasets, and the best option for modelling AR1xAR1 residual variance structures. Commercial use of `asreml` requires a license: <http://www.vsni.co.uk/downloads/asreml>.

Comments on the package structure

A large portion of these datasets appear in electronic form here for the first time.

A tremendous amount of effort has been given to the curating process of identifying datasets, extracting the data from source materials, checking data values, and documenting the data. In effect, to make the data somewhat 'computable' (Wolfram 2017).

The original sources for these data use several different words to refer to genotypes including *accession*, *breed*, *cultivar*, *genotype*, *hybrid*, *line*, *progeny*, *stock*, *type*, and *variety*. For consistency, these datasets mostly use `gen` (genotype).

Also for consistency `row` and `col` are usually used for the field coordinates.

In dataframes, 'block', 'rep', and similar terms are almost always coded like B1, B2, B3 instead of 1, 2, 3. This causes R to treat the data as a factor instead of a numeric covariate (which is a good thing).

Almost all of the data are presented as 'tidy' dataframes with 'observations' in rows and 'variables' in columns.

Although using `data()` is not necessary to access the data files, the example sections do include the use of `data()` because `devtools::run_examples()` needs it.

Please report any bugs to the package author or at the package github site.

Author(s)

Kevin Wright

The author is grateful to the many people who granted permission to include their data in this package.

References

- G. E. P. Box (1957), Integration of Techniques in Process Development, *Transactions of the American Society for Quality Control*.
- J. White and Frits van Evert. (2008). Publishing Agronomic Data. *Agron J.* 100, 1396-1400. <http://doi.org/10.2134/agronj2008.0080F>
- Stephen Wolfram (2017). *Launching the Wolfram Data Repository: Data Publishing that Really Works*. <http://blog.stephenwolfram.com/2017/04/launching-the-wolfram-data-repository-data-publishing-that-really-works/>

allcroft.lodging

Multi-environment trial of cereal with lodging data

Description

Percent lodging is given for 32 genotypes at 7 environments.

Format

A data frame with 224 observations on the following 3 variables.

env environment, 1-7

gen genotype, 1-32

y percent lodged

Details

This data is for the first year of a three-year study.

Source

D. J. Allcroft and C. A. Glasbey, 2003. Analysis of crop lodging using a latent variable model. *Journal of Agricultural Science*, 140, 383–393. <http://doi.org/10.1017/S0021859603003332>

Used with permission of Chris Glasbey.

Examples

```
data("allcroft.lodging")
dat <- allcroft.lodging

# Transformation
dat$sy <- sqrt(dat$y)
# Variety 4 has no lodging anywhere, so add a small amount
dat[dat$env=='E5' & dat$gen=='G04',]$sy <- .01

require(lattice)
dotplot(env~y|gen, dat, as.table=TRUE,
```

```

      xlab="Percent lodged (by genotype)", ylab="Variety",
      main="allcroft.lodging")

# Tobit model
if(require(AER)){
  m3 <- tobit(sy ~ 1 + gen + env, left=0, right=100, data=dat)

  # Table 2 trial/variety means
  preds <- expand.grid(gen=levels(dat$gen), env=levels(dat$env))
  preds$pred <- predict(m3, newdata=preds)
  round(tapply(preds$pred, preds$gen, mean),2)
  round(tapply(preds$pred, preds$env, mean),2)

}

```

 archbold.apple

Split-split plot experiment on apple trees

Description

Split-split plot experiment on apple trees with different spacing, root stock, and cultivars.

Format

A data frame with 120 observations on the following 10 variables.

rep block, 5 levels

row row

pos position within each row

spacing spacing between trees, 6,10,14 feet

stock rootstock, 4 levels

gen genotype, 2 levels

yield yield total, kg/tree from 1975-1979

trt treatment code

Details

In rep 1, the 10-foot-spacing main plot was split into two non-contiguous pieces. This also happened in rep 4. In the analysis of Cornelius and Archbold, they consider each row x within-row-spacing to be a distinct main plot. (Also true for the 14-foot row-spacing, even though the 14-foot spacing plots were contiguous.)

The treatment code is defined as $100 * \text{spacing} + 10 * \text{stock} + \text{gen}$, where stock=0,1,6,7 for Seedling,MM111,MM106,M0007 and gen=1,2 for Redspur,Golden, respectively.

Source

D Archbold and G. R. Brown and P. L. Cornelius. (1987). Rootstock and in-row spacing effects on growth and yield of spur-type delicious and Golden delicious apple. *Journal of the American Society for Horticultural Science*, 112, 219-222.

References

Cornelius, PL and Archbold, DD, 1989. Analysis of a split-split plot experiment with missing data using mixed model equations. *Applications of Mixed Models in Agriculture and Related Disciplines*. Pages 55-79.

Examples

```
data(archbold.apple)
dat <- archbold.apple

# Define main plot and subplot
dat <- transform(dat, rep=factor(rep), spacing=factor(spacing), trt=factor(trt),
                 mp = factor(paste(row,spacing,sep="")),
                 sp = factor(paste(row,spacing,stock,sep="")))

# Due to 'spacing', the plots are different sizes, but the following layout
# shows the relative position of the plots and treatments. Note that the
# 'spacing' treatments are not contiguous in some reps.
if(require(desplot)){
  desplot(spacing~row*pos, dat,
          col=stock, cex=1, num=gen, # aspect unknown
          main="archbold.apple")
}

## Not run:
require("lme4")
require("lucid")

m1 <- lmer(yield ~ -1 + trt + (1|rep/mp/sp), dat)

vc(m1) # Variances/means on Cornelius, page 59
##      grp      var1 var2  vcov sdcov
## sp:(mp:rep) (Intercept) <NA> 193.3 13.9
##      mp:rep (Intercept) <NA> 203.8 14.28
##      rep (Intercept) <NA> 197.3 14.05
##      Residual      <NA> <NA> 1015 31.86

## End(Not run)
```

ars.earlywhitecorn96 *Multi-environment trial of early white food corn*

Description

Multi-environment trial of early white food corn for 60 white hybrids.

Format

A data frame with 540 observations on the following 9 variables.

loc location, 9 levels

gen gen, 60 levels

yield yield, bu/ac

stand stand, percent

rootlodge root lodging, percent

stalklodge stalk lodging, percent

earht ear height, inches

flower days to flower

moisture moisture, percent

Details

Data are the average of 3 replications.

Yields were measured for each plot and converted to bushels / acre and adjusted to 15.5 percent moisture.

Stand is expressed as a percentage of the optimum plant stand.

Lodging is expressed as a percentage of the total plants for each hybrid.

Ear height was measured from soil level to the top ear leaf collar. Heights are expressed in inches.

Days to flowering is the number of days from planting to mid-tassel or mid-silk.

Moisture of the grain was measured at harvest.

Source

L. Darrah, R. Lundquist, D. West, C. Poneleit, B. Barry, B. Zehr, A. Bockholt, L. Maddux, K. Ziegler, and P. Martin. (1996). *White Food Corn 1996 Performance Tests*. Agricultural Research Service Special Report 502.

Examples

```

data(ars.earlywhitecorn96)
dat <- ars.earlywhitecorn96

require(lattice)
# These views emphasize differences between locations
dotplot(gen~yield, dat, group=loc, auto.key=list(columns=3),
        main="ars.earlywhitecorn96")
## dotplot(gen~stalklodge, dat, group=loc, auto.key=list(columns=3),
##        main="ars.earlywhitecorn96")
splom(~dat[,3:9], group=dat$loc, auto.key=list(columns=3),
      main="ars.earlywhitecorn96")

# MANOVA
m1 <- manova(cbind(yield,earht,moisture) ~ gen + loc, dat)
m1
summary(m1)

```

australia.soybean *Multi-environment trial of soybean in Australia*

Description

Yield and other traits of 58 varieties of soybeans, grown in four locations across two years in Australia. This is four-way data of Year x Loc x Gen x Trait.

Format

A data frame with 464 observations on the following 10 variables.

env environment, 8 levels, first character of location and last two characters of year

loc location

year year

gen genotype of soybeans, 1-58

yield yield, metric tons / hectare

height height (meters)

lodging lodging

size seed size, (millimeters)

protein protein (percentage)

oil oil (percentage)

Details

Measurements are available from four locations in Queensland, Australia in two consecutive years 1970, 1971.

The 58 different genotypes of soybeans consisted of 43 lines (40 local Australian selections from a cross, their two parents, and one other which was used as a parent in earlier trials) and 15 other lines of which 12 were from the US.

Lines 1-40 were local Australian selections from Mamloxi (CPI 172) and Avoyelles (CPI 15939).

| No. | Line |
|------|--------------------------------|
| 1-40 | Local selections |
| 41 | Avoyelles (CPI 15939) Tanzania |
| 42 | Heron 49 (CPI 15948) Tanzania |
| 43 | Mamloxi (CPI 172) Nigeria |
| 44 | Dorman USA |
| 45 | Hampton USA |
| 46 | Hill USA |
| 47 | Jackson USA |
| 48 | Leslie USA |
| 49 | Semstar Australia |
| 50 | Wills USA |
| 51 | C26673 Morocco |
| 52 | C26671 Morocco |
| 53 | Bragg USA |
| 54 | Delmar USA |
| 55 | Lee USA |
| 56 | Hood USA |
| 57 | Ogden USA |
| 58 | Wayne USA |

Note on the data in Basford and Tukey book. The values for line 58 for Nambour 1970 and Redland Bay 1971 are incorrectly listed on page 477 as 20.490 and 15.070. They should be 17.350 and 13.000, respectively. In the data set made available here, these values have been corrected.

Source

Basford, K. E., and Tukey, J. W. (1999). Graphical analysis of multiresponse data illustrated with a plant breeding trial. Chapman and Hall/CRC.

Retrieved from: <http://three-mode.leidenuniv.nl/data/soybeaninf.htm>

Used with permission of Kaye Basford, Pieter Kroonenberg.

References

K E Basford. 1982. The Use of Multidimensional Scaling in Analysing Multi-Attribute Genotype Response Across Environments, *Aust J Agric Res*, 33, 473–480.

Kroonenberg, P. M., & Basford, K. E. B. (1989). An investigation of multi-attribute genotype response across environments using three-mode principal component analysis. *Euphytica*, 44, 109–123.

Marcin Kozak (2010). Use of parallel coordinate plots in multi-response selection of interesting genotypes. *Communications in Biometry and Crop Science*, 5, 83-95.

Examples

```

data(australia.soybean)
dat <- australia.soybean

if(require(reshape2)){
dm <- melt(dat, id.var=c('env', 'year', 'loc', 'gen'))

# Joint plot of genotypes & traits. Similar to Figure 1 of Kroonenberg 1989
dmat <- acast(dm, gen~variable, fun=mean)
dmat <- scale(dmat)
biplot(princomp(dmat), main="australia.soybean trait x gen biplot", cex=.75)
}

# Figure 1 of Kozak 2010, lines 44-58
## Not run:
require(agridat); require(reshape2) ; require(lattice); require(latticeExtra)
data(australia.soybean)
dat <- australia.soybean
dat <- melt(dat, id.var=c('env', 'year', 'loc', 'gen'))
dat <- acast(dat, gen~variable, fun=mean)
dat <- scale(dat)
dat <- as.data.frame(dat)[,c(2:6,1)]
dat$gen <- rownames(dat)
# data for the graphic by Kozak
dat2 <- dat[44:58,]
dat3 <- subset(dat2, is.element(gen, c("G48", "G49", "G50", "G51")))

parallelplot( ~ dat3[,1:6]|dat3$gen, main="australia.soybean",
  as.table=TRUE, horiz=FALSE) +
  parallelplot( ~ dat2[,1:6], horiz=FALSE, col="gray80") +
  parallelplot( ~ dat3[,1:6]|dat3$gen,
    as.table=TRUE, horiz=FALSE, lwd=2)

## End(Not run)

```

Description

Uniformity trials of barley at Davis, California, 1925-1935, 10 years on same ground.

Format

A data frame with 570 observations on the following 4 variables.

row row

col column

year year

yield yield, pounds/acre

Details

Ten years of uniformity trials were sown on the same ground. Baker (1952) shows a map of the field.

Field width: 19 plots = 827 ft

Field length: 3 plots * 161 ft + 2 alleys * 15 feet = 513 ft

Source

Baker, GA and Huberty, MR and Veihmeyer, FJ. (1952) A uniformity trial on unirrigated barley of ten years' duration. *Agronomy Journal*, 44, 267-270. <http://doi.org/10.2134/agronj1952.00021962004400050011x>

Examples

```
data(baker.barley.uniformity)
dat <- baker.barley.uniformity

# Ten-year average
dat2 <- aggregate(yield ~ row*col, data=dat, FUN=mean, na.rm=TRUE)

if(require(desplot)){
  desplot(yield~col*row|year, data=dat,
          aspect = 513/827, # true aspect
          main="baker.barley.uniformity - heatmaps by year")
}
if(require(desplot)){
  desplot(yield~col*row, data=dat2,
          aspect = 513/827, # true aspect
          main="baker.barley.uniformity - heatmap of 10-year average")
  # Note low yield in upper right, slanting to left a bit due to sandy soil
  # as shown in Baker figure 1.
}

# Baker fig 2, stdev vs mean
dat3 <- aggregate(yield ~ row*col, data=dat, FUN=sd, na.rm=TRUE)
plot(dat2$yield, dat3$yield, xlab="Mean yield", ylab="Std Dev yield",
     main="baker.barley.uniformity")
```

```
# Baker table 4, correlation of plots across years
if(require(reshape2)){
mat <- acast(dat, row+col~year)
round(cor(mat, use='pair'),2)
}
```

batchelor.uniformity *Uniformity trials of apples, lemons, oranges, and walnuts*

Description

Uniformity trials of apples, lemons, oranges, and walnuts, in California & Utah, 1915-1918.

Format

Each dataset has the following format

```
row row
col column
yield yield per tree (pounds)
```

Details

A few of the trees affected by disease were eliminated and the yield was replaced by the average of the eight surrounding trees.

The following details are from Batchelor (1918).

Jonathan Apples

"The apple records were obtained from a 10-year old Jonathan apple orchard located at Providence, Utah. The surface soil of this orchard is very uniform to all appearances except on the extreme eastern edge, where the percentage of gravel increases slightly. The trees are planted 16 feet apart, east and west, and 30 feet apart north and south."

Note: The orientation of the field is not given in the paper, but all other fields in the paper have north at the top, so that is assumed to be true for this field as well.

Field length: 28 rows * 30 feet = 840 feet

Field width: 8 trees * 16 feet = 128 feet

Eureka Lemon

The lemon (*Citrus limonia*) tree yields were obtained from a grove of 364 23-year-old trees, located at Upland, California. The records extend from October 1, 1915, to October 1, 1916. The grove consists of 14 rows of 23-year-old trees, extending north and south, with 26 trees in a row, planted 24 by 24 feet apart. This grove presents the most uniform appearance of any under consideration [in this paper]. The land is practically level, and the soil is apparently uniform in texture. The records show a grouping of several low-yielding trees; yet a field observation gives one the impression that the grove as a whole is remarkably uniform.

Field width: 14 trees * 24 feet = 336 feet

Field length: 26 trees * 24 feet = 624 feet

Navel 1 (Arlington)

These records were of the 1915-16 yields of one thousand 24-year-old navel-orange trees near Arlington station, Riverside, California. The grove consists of 20 rows of trees from north to south, with 50 trees in a row, planted 22 by 22 feet. A study of the records shows certain distinct high- and low-yielding areas. The northeast corner and the south end contain notably high-yielding trees. The north two-thirds of the west side contains a large number of low-yielding trees. These areas are apparently correlated with soil variation. Variations from tree to tree also occur, the cause of which is not evident. These variations, which are present in every orchard, bring uncertainty into the results of field experiments.

Field length: 50 trees * 22 feet = 1100 feet

Field width: 20 trees * 22 feet = 440 feet

Navel 2 (Antelope)

The navel-orange grove later referred to as the Antelope Heights navels is a plantation of 480 ten-year-old trees planted 22 by 22 feet, located at Naranjo, California. The yields are from 1916. The general appearance of the trees gives a visual impression of uniformity greater than a comparison of the individual tree production substantiates.

Field width: 15 trees * 22 feet = 330 feet

Field length: 33 trees * 22 feet = 726 feet

Valencia Orange

The Valencia orange grove is composed of 240 15-year-old trees, planted 21 feet 6 inches by 22 feet 6 inches, located at Villa Park, California. The yields were obtained in 1916.

Field length: 20 rows * 22 feet = 440 feet

Field width: 12 rows * 22 feet = 264 feet

Walnut

The walnut (*Juglans regia*) yields were obtained during the seasons of 1915 and 1916 from a 24-year-old Santa Barbara softshell seedling grove, located at Whittier, California. [Note, The yields here appear to be the 1915 yields.] The planting is laid out 10 trees wide and 32 trees long, entirely surrounded by additional walnut plantings, except on a part of one side which is adjacent to an orange grove. The trees are planted on the square system, 50 feet apart.

Field width: 10 trees * 50 feet = 500 feet

Field length: 32 trees * 50 feet = 1600 feet

Source

Batchelor, LD and Reed, HS. 1918. Relation of the variability of yields of fruit trees to the accuracy of field trials. *J. Agric. Res.*, 12, 245–283. <http://books.google.com/books?id=Lil6AAAAMAAJ&lr&pg=PA245>

References

McCullagh, P. and Clifford, D., (2006). Evidence for conformal invariance of crop yields, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 462, 2119–2143. <http://doi.org/10.1098/rspa.2006.1667>

Examples

```
if(require(desplot)){
  # Apple
  data(batchelor.apple.uniformity)
  desplot(yield~col*row, data=batchelor.apple.uniformity,
          aspect=840/128, tick=TRUE, # true aspect
          main="batchelor.apple.uniformity")
}

if(require(desplot)){
  # Lemon
  data(batchelor.lemon.uniformity)
  desplot(yield~col*row, data=batchelor.lemon.uniformity,
          aspect=624/336, # true aspect
          main="batchelor.lemon.uniformity")
}

if(require(desplot)){
  # Navel1 (Arlington)
  data(batchelor.navel1.uniformity)
  desplot(yield~col*row, batchelor.navel1.uniformity,
          aspect=1100/440, # true aspect
          main="batchelor.navel1.uniformity - Arlington")
}

if(require(desplot)){
  # Navel2 (Antelope)
  data(batchelor.navel2.uniformity)
  desplot(yield~col*row, data=batchelor.navel2.uniformity,
          aspect=726/330, # true aspect
          main="batchelor.navel2.uniformity - Antelope")
}

if(require(desplot)){
  # Valencia
  data(batchelor.valencia.uniformity)
  desplot(yield~col*row, data=batchelor.valencia.uniformity,
          aspect=440/264, # true aspect
          main="batchelor.valencia.uniformity")
}

if(require(desplot)){
  # Walnut
  data(batchelor.walnut.uniformity)
  desplot(yield~col*row, data=batchelor.walnut.uniformity,
          aspect=1600/500, # true aspect
          main="batchelor.walnut.uniformity")
}
```

`battese.survey`*Survey and satellite data for corn and soy areas in Iowa*

Description

Survey and satellite data for corn and soy areas in Iowa

Usage

```
data("battese.survey")
```

Format

A data frame with 37 observations on the following 9 variables.

`county` county name

`segment` sample segment number (within county)

`countysegs` number of segments in county

`cornhect` hectares of corn in segment

`soyhect` hectares of soy

`cornpix` pixels of corn in segment

`soypix` pixels of soy

`cornmean` county mean of corn pixels per segment

`soymean` county mean of soy pixels per segment

Details

The data are for 12 counties in north-central Iowa in 1978.

The USDA determined the area of soybeans in 37 area sampling units (called 'segments'). Each segment is about one square mile (about 259 hectares). The number of pixels of that were classified as corn and soybeans came from Landsat images obtained in Aug/Sep 1978. Each pixel represents approximately 0.45 hectares.

Data originally compiled by USDA.

This data is also available in R packages: 'rsae::landsat' and 'JoSAE::landsat'.

Source

Battese, George E and Harter, Rachel M and Fuller, Wayne A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *emphJournal of the American Statistical Association*, 83, 28-36. <http://doi.org/10.2307/2288915>

Battese (1982) preprint version. http://www.une.edu.au/__data/assets/pdf_file/0017/15542/emetwp15.pdf

References

Pushpal K Mukhopadhyay and Allen McDowell. (2011). Small Area Estimation for Survey Data Analysis Using SAS Software SAS Global Forum 2011.

Examples

```

data(battese.survey)
dat <- battese.survey

# Battese fig 1 & 2. Corn plot shows outlier in Hardin county
library("lattice")
dat <- dat[order(dat$cornpix),]
xyplot(cornhect ~ cornpix, data=dat, group=county, type=c('p','l'),
       main="battese.survey", xlab="Pixels of corn", ylab="Hectares of corn",
       auto.key=list(columns=3))

## Not run:
dat <- dat[order(dat$soypix),]
xyplot(soyhect ~ soypix, data=dat, group=county, type=c('p','l'),
       main="battese.survey", xlab="Pixels of soy", ylab="Hectares of soy",
       auto.key=list(columns=3))

## End(Not run)

## Not run:
require("lme4")
require("lucid")

# Fit the models of Battese 1982, p.18. Results match
m1 <- lmer(cornhect ~ 1 + cornpix + (1|county), data=dat)
fixef(m1)
## (Intercept)    cornpix
##  5.4661899    0.3878358
vc(m1)
##      grp      var1 var2  vcov  sdcov
## county (Intercept) <NA> 62.83 7.926
## Residual          <NA> <NA> 290.4 17.04
m2 <- lmer(soyhect ~ 1 + soypix + (1|county), data=dat)
fixef(m2)
## (Intercept)    soypix
## -3.8223566    0.4756781
vc(m2)
##      grp      var1 var2  vcov  sdcov
## county (Intercept) <NA> 239.2 15.47
## Residual          <NA> <NA> 180  13.42

# Predict for Humboldt county as in Battese 1982 table 2
5.4662+.3878*290.74
# 118.2152 # mu_i^0
5.4662+.3878*290.74+ -2.8744
# 115.3408 # mu_i^gamma
(185.35+116.43)/2

```

```

# 150.89 # y_i bar

# Survey regression estimator of Battese 1988

# Delete the outlier
dat2 <- subset(dat, !(county=="Hardin" & soyhect < 30))

# Results match top-right of Battese 1988, p. 33
m3 <- lmer(cornhect ~ cornpix + soypix + (1|county), data=dat2)
fixef(m3)
## (Intercept)      cornpix      soypix
## 51.0703979    0.3287217   -0.1345684
vc(m3)
##      grp      var1 var2  vcov sdcor
## county (Intercept) <NA> 140  11.83
## Residual          <NA> <NA> 147.3 12.14
m4 <- lmer(soyhect ~ cornpix + soypix + (1|county), data=dat2)
fixef(m4)
## (Intercept)      cornpix      soypix
## -15.59027098    0.02717639    0.49439320
vc(m4)
##      grp      var1 var2  vcov sdcor
## county (Intercept) <NA> 247.5 15.73
## Residual          <NA> <NA> 190.5 13.8

## End(Not run)

```

beall.webworms

Counts of webworms in a beet field, with insecticide treatments.

Description

Counts of webworms in a beet field, with insecticide treatments.

Usage

```
data("beall.webworms")
```

Format

A data frame with 1300 observations on the following 7 variables.

```

row row
col column
y count of webworms
block block
trt treatment
spray spray treatment yes/no
lead lead treatment yes/no

```

Details

The beet webworm lays egg masses as small as 1 egg, seldom exceeding 5 eggs. The larvae can move freely, but usually mature on the plant on which they hatch.

Each plot contained 25 unit areas, each 1 row by 3 feet long. The row width is 22 inches. The arrangement of plots within the blocks seems certain, but the arrangement of the blocks/treatments is not certain, since the authors say "since the plots were 5 units long and 5 wide it is only practicable to combine them into groups of 5 in one direction or the other".

Treatment 1 = None. Treatment 2 = Contact spray. Treatment 3 = Lead arsenate. Treatment 4 = Both spray, lead arsenate.

Source

Beall, Geoffrey (1940). The fit and significance of contagious distributions when applied to observations on larval insects. *Ecology*, 21, 460-474. Table 6. <http://doi.org/10.2307/1930285>

Examples

```
data(beall.webworms)
dat <- beall.webworms

# Match Beall table 1
with(dat, table(y,trt))

require(lattice)
histogram(~y|trt, data=dat, layout=c(1,4), as.table=TRUE,
          main="beall.webworms")

# Visualize Beall table 6. Block effects may exist, but barely.
if(require(desplot)){
  grays <- colorRampPalette(c("white", "#252525"))
  desplot(y ~ col*row, data=dat,
          col.regions=grays(10),
          at=0:10-0.5,
          out1=block, out2=trt, num=trt, flip=TRUE, # aspect unknown
          main="beall.webworms (count of worms)")
}

# Following plot suggests interaction is needed
# with(dat, interaction.plot(spray, lead, y))

# Just a simple model here...
m1 <- glm(y ~ block+spray*lead, data=dat, family="poisson")
summary(m1)
anova(m1)
```

`beaven.barley`*Yields of 8 barley varieties in 1913 as used by Student.*

Description

Yields of 8 barley varieties in 1913.

Usage

```
data("beaven.barley")
```

Format

A data frame with 160 observations on the following 4 variables.

```
row row
```

```
col column
```

```
gen genotype
```

```
yield yield (grams)
```

Details

Eight races of barley were grown on a regular pattern of plots.

These data were prepared from Richey (1926) because the text was cleaner.

Each plot was planted 40 inches on a side, but only the middle square 36 inches on a side was harvested.

Field length: 5 plots * 3 feet = 15 feet

Field width: 32 plots * 3 feet = 96 feet.

Source

Student. (1923). On testing varieties of cereals. *Biometrika*, 271-293.

<http://doi.org/10.1093/biomet/15.3-4.271>

References

Frederick D. Richey (1926). The moving average as a basis for measuring correlated variation in agronomic experiments. *Jour. Agr. Research*, 32, 1161-1175.

Examples

```

data(beaven.barley)
dat <- beaven.barley

# Match the means shown in Richey table IV
tapply(dat$yield, dat$gen, mean)
##      a      b      c      d      e      f      g      h
## 298.080 300.710 318.685 295.260 306.410 276.475 304.605 271.820

# Compare to Student 1923, diagram I,II
if(require(desplot)){
  desplot(yield ~ col*row, data=dat,
          aspect=15/96, # true aspect
          main="beaven.barley - variety trial", text=gen)
}

```

becker.chicken

Mating crosses of chickens

Description

Mating crosses of chickens

Usage

```
data("becker.chicken")
```

Format

A data frame with 45 observations on the following 3 variables.

male male parent

female female parent

weight weight (g) at 8 weeks

Details

From a large flock White Rock chickens, five male sires were chosen and mated to each of three female dams, producing 3 female progeny. The data are body weights at eight weeks of age.

Becker (1984) used these data to demonstrate the calculation of heritability.

Source

Walter A. Becker (1984). *Manual of Quantitative Genetics*, 4th ed. Page 83.

Prepared in electronic form by Kevin Wright.

References

None.

Examples

```

data(becker.chicken)
dat <- becker.chicken

if(require(lattice)){
  dotplot(weight ~ female, data=dat, group=male,
    main="becker.chicken - progeny weight by M*F",
    xlab="female parent",ylab="progeny weight",
    auto.key=list(columns=5))
}

## Not run:
# Sums match Becker
sum(dat$weight)
aggregate(weight ~ male + female, dat, FUN=sum)

# Variance components
require(lme4)
require(lucid)
m1 <- lmer(weight ~ (1|male) + (1|female), data=dat)
as.data.frame(lme4::VarCorr(m1))
## grp      var1 var2      vcov      sdcov
## 1  female (Intercept) <NA> 1095.6296 33.10030
## 2   male (Intercept) <NA>  776.7543 27.87031
## 3 Residual          <NA> <NA> 5524.4000 74.32631

# Calculate heritabilities
s2m <- 776 # variability for males
s2f <- 1095 # variability for females
s2w <- 5524 # variability within crosses
vp <- s2m + s2f + s2w # 7395
4*s2m/vp # .42 male heritability
4*s2f/vp # .59 female heritability

## End(Not run)

```

besag.bayesian

Spring barley in United Kingdom

Description

An experiment with 75 varieties of barley, planted in 3 reps.

Format

A data frame with 225 observations on the following 4 variables.

col column (also blocking factor)

row row

yield yield

gen variety/genotype

Details

RCB design, each column is one rep.

Source

Besag, J. E., Green, P. J., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statistical Science*, 10, 3-66. <http://www.jstor.org/stable/2246224>

Used with permission of David Higdon.

References

Davison, A. C. (2003). *Statistical Models*. Cambridge University Press. Pages 534-535.

Examples

```
data(besag.bayesian)
dat <- besag.bayesian

# Yield values were scaled to unit variance
var(dat$yield, na.rm=TRUE)

# Besag Fig 2. Reverse row numbers to match Besag, Davison
dat$rrow <- 76 - dat$row
require("lattice")
xyplot(yield ~ rrow|col, dat, layout=c(1,3), type='s',
       xlab="row", ylab="yield", main="besag.bayesian")

# -----

## Not run:
# asreml3
require(asreml)

# Use asreml to fit a model with AR1 gradient in rows

dat <- transform(dat, cf=factor(col), rf=factor(rrow))
m1 <- asreml(yield ~ -1 + gen, data=dat, random=~ar1v(rf))
m1 <- update(m1)

# Visualize trends, similar to Besag figure 2.
```



```

dat$res <- m1$residuals
dat$geneff <- coef(m1)$fixed[as.numeric(dat$gen)]
dat <- transform(dat, fert=yield-geneff-res)
xyplot(geneff ~ rrow|col, dat, layout=c(1,3), type='s',
       main="Variety effects", ylim=c(5,15 ))
xyplot(fert ~ rrow|col, dat, layout=c(1,3), type='s',
       main="Fertility", ylim=c(-2,2))
xyplot(res ~ rrow|col, dat, layout=c(1,3), type='s',
       main="Residuals", ylim=c(-4,4))

## End(Not run)

# -----

## Not run:
## require(asreml4)

## # Use asreml to fit a model with AR1 gradient in rows

## dat <- transform(dat, cf=factor(col), rf=factor(rrow))
## m1 <- asreml(yield ~ -1 + gen, data=dat, random=~ar1v(rf))
## m1 <- update(m1)
## m1 <- update(m1)
## m1 <- update(m1)
## m1 <- update(m1)
## m1 <- update(m1)
## m1 <- update(m1)
## m1 <- update(m1)

## # Visualize trends, similar to Besag figure 2.
## dat$res <- resid(m1)
## dat$geneff <- coef(m1)$fixed[as.numeric(dat$gen)]
## dat <- transform(dat, fert=yield-geneff-res)
## xyplot(geneff ~ rrow|col, dat, layout=c(1,3), type='s',
##        main="Variety effects", ylim=c(5,15 ))
## xyplot(fert ~ rrow|col, dat, layout=c(1,3), type='s',
##        main="Fertility", ylim=c(-2,2))
## xyplot(res ~ rrow|col, dat, layout=c(1,3), type='s',
##        main="Residuals", ylim=c(-4,4))

## End(Not run)

```

Description

Competition experiment in beans with height measurements

Usage

```
data("besag.beans")
```

Format

A data frame with 152 observations on the following 6 variables.

```
gen genotype / variety
height plot height, cm
yield plot yield, g
row row / block
rep replicate factor
col column
```

Details

Field beans of regular height were grown beside shorter varieties. In each block, each variety occurred once as a left-side neighbor and once as a right-side neighbor of every variety (including itself). Border plots were placed at the ends of each block. Each block with 38 adjacent plots. Each plot was one row, 3 meters long with 50 cm spacing between rows. No gaps between plots. Spacing between plants was 6.7 cm. Four blocks (rows) were used, each with six replicates.

Plot yield and height was recorded.

Kempton and Lockwood used models that adjusted yield according to the difference in height of neighboring plots.

Field length: 4 plots * 3m = 12m

Field width: 38 plots * 0.5 m = 19m

Source

Julian Besag and Rob Kempton (1986). Statistical Analysis of Field Experiments Using Neighbouring Plots. *Biometrics*, 42, 231-251. Table 6. <http://doi.org/10.2307/2531047>

References

Kempton, RA and Lockwood, G. (1984). Inter-plot competition in variety trials of field beans (*Vicia faba* L.). *The Journal of Agricultural Science*, 103, 293–302.

Examples

```
data(besag.beans)
dat = besag.beans

if(require(desplot)){
  desplot(yield ~ col*row, data=dat,
          aspect=12/19, out1=row, out2=rep, num=gen, cex=1, # true aspect
          main="besag.beans")
}
```

```

}

if(require(reshape2)){
  # Add a covariate = excess height of neighbors
  mat <- acast(dat, row~col, value.var='height')
  mat2 <- matrix(NA, nrow=4, ncol=38)
  mat2[,2:37] <- (mat[,1:36] + mat[,3:38] - 2*mat[,2:37])
  dat2 <- melt(mat2)
  colnames(dat2) <- c('row', 'col', 'cov')
  dat <- merge(dat, dat2)

  # Drop border plots
  dat <- subset(dat, rep != 'R0')

  require(lattice)
  # Plot yield vs neighbors height advantage
  xyplot(yield~cov, data=dat, group=gen,
         main="besag.beans",
         xlab="Mean excess heights of neighbor plots",
         auto.key=list(columns=3))

  # Trial mean.
  mean(dat$yield) # 391 matches Kempton table 3

  # Mean excess height of neighbors for each genotype
  tapply(dat$cov, dat$gen, mean)/2 # Matches Kempton table 4

  # Variety means, matches Kempton table 4 mean yield
  m1 <- lm(yield ~ -1 + gen, dat)
  coef(m1)

  # Full model used by Kempton, eqn 5. Not perfectly clear.
  # Appears to include rep term, perhaps within block
  dat$blk <- factor(dat$row)
  dat$blkrep <- factor(paste(dat$blk, dat$rep))
  m2 <- lm(yield ~ -1 + gen + blkrep + cov, data=dat)
  coef(m2) # slope 'cov' = -.72, while Kempton says -.79
}

```

besag.elbatan

RCB experiment of 50 varieties of wheat in 3 blocks with strong spatial trend.

Description

RCB experiment of 50 varieties of wheat in 3 blocks with strong spatial trend.

Format

A data frame with 150 observations on the following 4 variables.

```
yield yield of wheat
gen genotype, factor with 50 levels
col column/block
row row
```

Details

RCB experiment on wheat at El Batan, Mexico. There are three single-column replicates with 50 varieties in each replicate.

Plot dimensions are not given by Besag.

Source

Julian Besag and D Higdon, 1999. Bayesian Analysis of Agricultural Field Experiments, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61, 691–746. Table 1.

<http://doi.org/10.1111/1467-9868.00201>

Retrieved from <http://web.archive.org/web/19991008143232/www.stat.duke.edu/~higdon/trials/elbatan.dat>

Used with permission of David Higdon.

References

Wilkinson 1984.

Besag & Seheult 1989.

Examples

```
data(besag.elbatan)
dat <- besag.elbatan

if(require(desplot)){
  desplot(yield~col*row, dat,
          num=gen, # aspect unknown
          main="besag.elbatan wheat yields")
}

## Not run:
# Besag figure 1
library(lattice)
xyplot(yield~row|col, dat, type=c('l'),
       layout=c(1,3), main="besag.elbatan wheat yields")

## End(Not run)

# RCB
```

```

m1 <- lm(yield ~ 0 + gen + factor(col), dat)
p1 <- coef(m1)[1:50]

## Not run:
# Formerly used gam package, but as of R 3.1, Rcmd check --as-cran
# is complaining
# Calls: plot.gam ... model.matrix.gam -> predict -> predict.gam -> array
# but it works perfectly in interactive mode !!!
# Remove the FALSE to run the code below
if(is.element("gam", search())) detach(package:gam)
library(mgcv)
m2 <- mgcv::gam(yield ~ -1 + gen + factor(col) + s(row), data=dat)
plot(m2, residuals=TRUE, main="besag.elbatan")
pred <- cbind(dat, predict(m2, dat, type="terms"))
# Need to correct for the average loess effect, which is like
# an overall intercept term.
adjlo <- mean(pred$"s(row)")
p2 <- coef(m2)[1:50] + adjlo

# Compare estimates
lims <- range(c(p1,p2))
plot(p1, p2, xlab="RCB prediction",
      ylab="RCB with smooth trend (predicted)",
      type='n', xlim=lims, ylim=lims,
      main="besag.elbatan")
text(p1, p2, 1:50, cex=.5)
abline(0,1,col="gray")

## End(Not run)

```

besag.endive

Presence of footroot disease in an endive field

Description

Presence of footroot disease in an endive field

Format

A data frame with 2506 observations on the following 3 variables.

col column

row row

disease plant is diseased, Y=yes,N=no

Details

In a field of endives, does each plant have footrot, or not? Data are binary on a lattice of 14 x 179 plants.

Modeled as an autologistic distribution.

We assume the endives are a single genotype.

Besag (1978) may have had data taken at 4 time points. This data was extracted from Friel and Pettitt. It is not clear what, if any, time point was used.

Friel does not give the dimensions. Besag is not available.

Source

J Besag (1978). Some Methods of Statistical Analysis for Spatial Data. *Bulletin of the International Statistical Institute*, 47, 77-92.

References

N Friel & A. N Pettitt (2004). Likelihood Estimation and Inference for the Autologistic Model. *Journal of Computational and Graphical Statistics*, 13:1, 232-246. <http://doi.org/10.1198/1061860043029>

Examples

```
data(besag.endive)
dat <- besag.endive

# Incidence map. Figure 2 of Friel and Pettitt
if(require(desplot)){
  grays <- colorRampPalette(c("#d9d9d9", "#252525"))
  desplot(disease~col*row, dat,
          col.regions=grays(2),
          aspect = 0.5, # aspect unknown
          main="besag.endive - Disease incidence")
}

# -----

## Not run:

# Besag (2000) "An Introduction to Markov Chain Monte Carlo" suggested
# that the autologistic model is not a very good fit for this data.
# We try it anyway. No idea if this is correct or how to interpret...

require(ngspatial)
A = adjacency.matrix(179,14)
X = cbind(x=dat$col, y=dat$row)
Z = as.numeric(dat$disease=="Y")
m1 <- autologistic(Z ~ 0+X, A=A, control=list(confint="none"))

summary(m1)
## Coefficients:
```

```

##      Estimate Lower Upper MCSE
## Xx -0.007824    NA    NA    NA
## Xy -0.144800    NA    NA    NA
## eta  0.806200    NA    NA    NA

## End(Not run)

# -----

## Not run:
# asreml3
require(asreml)

# Now try an AR1xAR1 model.
dat2 <- transform(dat, xf=factor(col), yf=factor(row),
                  pres=as.numeric(disease=="Y"))

m2 <- asreml(pres ~ 1, data=dat2, rcov= ~ar1(xf):ar1(yf))
# The 0/1 response is arbitrary, but there is some suggestion
# of auto-correlation in the x (.17) and y (.10) directions,
# suggesting the pattern is more 'patchy' than just random noise,
# but is it meaningful?

require(lucid)
vc(m2)
##      effect component std.error z.ratio constr
## R!variance  0.1301  0.003798   34      pos
## R!xf.cor   0.1699  0.01942   8.7  uncon
## R!yf.cor   0.09842  0.02038   4.8  uncon

## End(Not run)

# -----

## Not run:
## require(asreml4)

## # Now try an AR1xAR1 model.
## dat2 <- transform(dat, xf=factor(col), yf=factor(row),
##                   pres=as.numeric(disease=="Y"))

## m2 <- asreml(pres ~ 1, data=dat2,
##              resid = ~ar1(xf):ar1(yf))
## # The 0/1 response is arbitrary, but there is some suggestion
## # of auto-correlation in the x (.17) and y (.10) directions,
## # suggesting the pattern is more 'patchy' than just random noise,
## # but is it meaningful?

## require(lucid)
## vc(m2)
## ##      effect component std.error z.ratio bound
## ##      xf:yf(R)  0.1301  0.003798   34      P   0

```

```
## ## xf:yf!xf!cor 0.1699 0.01942 8.7 U 0
## ## xf:yf!yf!cor 0.09842 0.02038 4.8 U 0
```

```
## End(Not run)
```

besag.met

Multi-environment trial of corn laid out in incomplete-blocks

Description

Multi-environment trial of corn laid out in incomplete-blocks at each location.

Format

A data frame with 1152 observations on the following 7 variables.

```
county county
row row
col column
rep rep
block incomplete block
yield yield
gen genotype, 1-64
```

Details

Multi-environment trial of 64 corn hybrids in six counties in North Carolina. Each location had 3 replicates in in incomplete-block design with an 18x11 lattice of plots whose length-to-width ratio was about 2:1.

Note: In the original data, each county had 6 missing plots. This data has rows for each missing plot that uses the same county/block/rep to fill-out the row, sets the genotype to G01, and sets the yield to missing. These missing values were added to the data so that asreml could more easily do AR1xAR1 analysis using rectangular regions.

Each location/panel is:

Field length: 18 rows * 2 units = 36 units.

Field width: 11 plots * 1 unit = 11 units.

Source

Julian Besag and D Higdon, 1999. Bayesian Analysis of Agricultural Field Experiments, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61, 691–746. Table 1.

<http://doi.org/10.1111/1467-9868.00201>

Retrieved from <http://web.archive.org/web/19990505223413/www.stat.duke.edu/~higdon/trials/nc.dat>

Used with permission of David Higdon.

Examples

```

data(besag.met)
dat <- besag.met

if(require(desplot)){
  desplot(yield ~ col*row|county, dat,
          aspect=36/11, # true aspect
          out1=rep, out2=block,
          main="besag.met")
}

# Average reps
datm <- aggregate(yield ~ county + gen, data=dat, FUN=mean)

# Sections below fit heteroskedastic variance models (variance for each variety)
# asreml takes 1 second, lme 73 seconds, SAS PROC MIXED 30 minutes

# -----

## Not run:
# lme
require(nlme)
m1l <- lme(yield ~ -1 + gen, data=datm, random=~1|county,
          weights = varIdent(form=~ 1|gen))
m1l$sigma^2 * c(1, coef(m1l$modelStruct$varStruct, unc = FALSE))^2
##      G02      G03      G04      G05      G06      G07      G08
## 91.90 210.75 63.03 112.05 28.39 237.36 72.72 42.97
## ... etc ...

## End(Not run)

# -----

## Not run:
# asreml3
require(asreml)

# asreml Using 'rcov' ALWAYS requires sorting the data

datm <- datm[order(datm$gen),]
m1a <- asreml(yield ~ gen, data=datm,
             random = ~ county,
             rcov = ~ at(gen):units,
             predict=asreml::predict.asreml(classify="gen"))

require(lucid)
vc(m1a)[1:7,]
##      effect component std.error z.ratio constr
## county!county.var 1324      838.2    1.6    pos
## gen_G01!variance  91.93    58.82    1.6    pos
## gen_G02!variance 210.7    133.9    1.6    pos

```

```

##   gen_G03!variance    63.03    40.53    1.6   pos
##   gen_G04!variance   112.1    71.53    1.6   pos
##   gen_G05!variance    28.39    18.63    1.5   pos
##   gen_G06!variance   237.4    150.8    1.6   pos

# We get the same results from asreml & lme
plot(m1a$gammas[-1],
     m1l$sigma^2 * c(1, coef(m1l$modelStruct$varStruct, unc = FALSE))^2)

# The following example shows how to construct a GxE biplot
# from the FA2 model.

dat <- besag.met
dat <- transform(dat, xf=factor(col), yf=factor(row))
dat <- dat[order(dat$county, dat$xf, dat$yf), ]

# First, AR1xAR1
m1 <- asreml(yield ~ county, data=dat,
             random = ~ gen:county,
             rcov = ~ at(county):ar1(xf):ar1(yf))

# Add FA1.
# For ASEExtras:::summary.fa, use fa(county,1):gen, NOT gen:fa(county,1)
m2 <- update(m1, random=~fa(county,1):gen)
# FA2
m3 <- update(m2, random=~fa(county,2):gen)

# Use the loadings to make a biplot
vars <- vc(m3)
psi <- vars[grepl(".var$", vars$effect), "component"]
la1 <- vars[grepl(".fa1$", vars$effect), "component"]
la2 <- vars[grepl(".fa2$", vars$effect), "component"]
mat <- as.matrix(data.frame(psi, la1, la2))
rot <- svd(mat[,-1])$v # rotation matrix
lam <- mat[,-1]
colnames(lam) <- c("load1", "load2")

co3 <- coef(m3)$random # Scores are the GxE coefficients
ix1 <- grepl("_Comp1:gen", rownames(co3))
ix2 <- grepl("_Comp2:gen", rownames(co3))
sco <- matrix(c(co3[ix1], co3[ix2]), ncol=2, byrow=FALSE)
sco <- sco
dimnames(sco) <- list(levels(dat$gen), c('load1','load2'))
rownames(lam) <- levels(dat$county)
sco[,1] <- -1 * sco[,1]
lam[,1] <- -1 * lam[,1]
biplot(sco, lam, cex=.5, main="FA2 coefficient biplot")
# G variance matrix
gvar <- lam

# Now get predictions and make an ordinary biplot
p3 <- predict(m3, data=dat, classify="county:gen")

```

```

p3 <- p3$pred$pval
require("gge")
bi3 <- gge(predicted.value ~ gen*county, data=p3, scale=FALSE)
dev.new()
# Very similar to the coefficient biplot
biplot(bi3, stand=FALSE, # what does 'stand' do?
       main="SVD biplot of FA2 predictions")

# latent factor plots and more
if(FALSE) {
  library(ASEExtras)
  ASEExtras::summary.fa(m3, g.list=c("G01", "G02", "G03", "G04", "G05", "G06", "G07", "G08"))
  out <- ASEExtras::summary.fa(m3, uniplot=0, blups=1, regplot=1, addedplot=0, heatmap=0)
  loads <- as.data.frame(out$gammas$`fa(county, 2):gen`$`rotated loads`)
  loads$county <- rownames(loads)

  regdat <- out$blups$`fa(county, 2):gen`$blups.inmet
  regdat <- merge(regdat, loads)

  library(latticeExtra)
  xyplot(blup ~ fac_1|gen, data=regdat, as.table=TRUE)+
    xyplot(regblup ~ fac_1|gen, data=regdat,
          as.table=TRUE, type='r')
}

## End(Not run)

# -----

## Not run:
## require(asreml4)

## # asreml Using 'rcov' ALWAYS requires sorting the data
## datm <- datm[order(datm$gen),]

## m1 <- asreml(yield ~ gen, data=datm,
##             random = ~ county,
##             resid = ~ dsum( ~ units|gen))
## #summary(m1)$varcomp[1:7,]
## require(lucid)
## vc(m1)[1:7,]
## ##   effect component std.error z.ratio bound
## ##   county      1324      838.2      1.6      P 0
## ##   gen_G01(R)    91.95      58.85      1.6      P 0
## ##   gen_G02(R)   210.7      133.8      1.6      P 0.1
## ##   gen_G03(R)    63.04      40.55      1.6      P 0
## ##   gen_G04(R)   112.1      71.54      1.6      P 0
## ##   gen_G05(R)    28.38      18.61      1.5      P 0.1
## ##   gen_G06(R)   237.4      150.8      1.6      P 0

## # We get the same results from asreml & lme
## plot(m1$vparameters[-1],

```

```

##      m1$sigma^2 * c(1, coef(m1$modelStruct$varStruct, unc = FALSE))^2)

## # The following example shows how to construct a GxE biplot
## # from the FA2 model.

## dat <- besag.met
## dat <- transform(dat, xf=factor(col), yf=factor(row))
## dat <- dat[order(dat$county, dat$xf, dat$yf), ]

## # First, AR1xAR1
## m1 <- asreml(yield ~ county, data=dat,
##             random = ~ gen:county,
##             resid = ~ dsum( ~ ar1(xf):ar1(yf)|county))

## # Add FA1
## m2 <- update(m1, random=~gen:fa(county,1))
## # FA2
## m3 <- update(m2, random=~gen:fa(county,2))

## # Use the loadings to make a biplot
## vars <- vc(m3)
## psi <- vars[grepl("!var$", vars$effect), "component"]
## la1 <- vars[grepl("!fa1$", vars$effect), "component"]
## la2 <- vars[grepl("!fa2$", vars$effect), "component"]
## mat <- as.matrix(data.frame(psi, la1, la2))
## rot <- svd(mat[,-1])$v # rotation matrix
## lam <- mat[,-1]
## colnames(lam) <- c("load1", "load2")

## co3 <- coef(m3)$random # Scores are the GxE coefficients
## ix1 <- grepl("_Comp1$", rownames(co3))
## ix2 <- grepl("_Comp2$", rownames(co3))
## sco <- matrix(c(co3[ix1], co3[ix2]), ncol=2, byrow=FALSE)
## sco <- sco
## dimnames(sco) <- list(levels(dat$gen) , c('load1','load2'))
## rownames(lam) <- levels(dat$county)
## sco[,1] <- -1 * sco[,1]
## lam[,1] <- -1 * lam[,1]
## biplot(sco, lam, cex=.5, main="FA2 coefficient biplot")
## # G variance matrix
## gvar <- lam

## # Now get predictions and make an ordinary biplot
## p3 <- predict(m3, data=dat, classify="county:gen")
## p3 <- p3$pvals
## require("gge")
## bi3 <- gge(predicted.value ~ gen*county, data=p3, scale=FALSE)
## dev.new()
## # Very similar to the coefficient biplot
## biplot(bi3, stand=FALSE, main="SVD biplot of FA2 predictions")

## End(Not run)

```

`besag.triticale`*Four-way factorial agronomic experiment in triticale*

Description

Four-way factorial agronomic experiment in triticale

Usage

```
data("besag.triticale")
```

Format

A data frame with 54 observations on the following 7 variables.

yield yield, g/m²

row row

col column

gen genotype / variety, 3 levels

rate seeding rate, kg/ha

nitro nitrogen rate, kw/ha

regulator growth regulator, 3 levels

Details

Experiment conducted as a factorial on the yields of triticale. Fully randomized. Plots were 1.5m x 5.5m, but the orientation is not clear.

Besag and Kempton show how accounting for neighbors changes non-significant genotype differences into significant differences.

Source

Julian Besag and Rob Kempton (1986). Statistical Analysis of Field Experiments Using Neighbouring Plots. *Biometrics*, 42, 231-251. Table 2. <http://doi.org/10.2307/2531047>

References

None.

Examples

```

data(besag.triticale)
dat <- besag.triticale

if(require(desplot)){
  desplot(yield ~ col*row, data=dat,
          # aspect unknown
          main="besag.triticale")
}

## Not run:
# Besag & Kempton are not perfectly clear on the model, but
# indicate that there was no evidence of any two-way interactions.
# A reduced, main-effect model had genotype effects that were
# "close to significant" at the five percent level.
# The model below has p-value of gen at .04, so must be slightly
# different than their model.
dat <- transform(dat, rate=factor(rate), nitro=factor(nitro))
dat <- transform(dat, xf=factor(col), yf=factor(row))
m2 <- lm(yield ~ gen + rate + nitro + regulator + yf, data=dat)
anova(m2)

# Similar, but not exact, to Besag figure 5
dat$res <- resid(m2)
require(lattice)
xyplot(res ~ col|as.character(row), data=dat,
        as.table=TRUE, type="s", layout=c(1,3),
        main="besag.triticale")

## End(Not run)

# -----

## Not run:
# asreml3
require(asreml)

# Besag uses an adjustment based on neighboring plots.
# This analysis fits the standard AR1xAR1 residual model

dat <- dat[order(dat$xf, dat$yf), ]
m3 <- asreml(yield ~ gen + rate + nitro + regulator +
             gen:rate + gen:nitro + gen:regulator +
             rate:nitro + rate:regulator +
             nitro:regulator + yf, data=dat,
             rcov = ~ ar1(xf):ar1(yf))
anova(m3) # Strongly significant gen, rate, regulator

## End(Not run)

```

```

# -----
## Not run:
## require(asreml4)

## # Besag uses an adjustment based on neighboring plots.
## # This analysis fits the standard AR1xAR1 residual model

## dat <- dat[order(dat$xf, dat$yf), ]
## m3 <- asreml(yield ~ gen + rate + nitro + regulator +
##             gen:rate + gen:nitro + gen:regulator +
##             rate:nitro + rate:regulator +
##             nitro:regulator + yf, data=dat,
##             resid = ~ ar1(xf):ar1(yf))
## wald(m3) # Strongly significant gen, rate, regulator
## ##           Df Sum of Sq Wald statistic Pr(Chisq)
## ## (Intercept)    1  1288255      103.971 < 2.2e-16 ***
## ## gen            2   903262      72.899 < 2.2e-16 ***
## ## rate           1   104774      8.456  0.003638 **
## ## nitro          1     282      0.023  0.880139
## ## regulator     2   231403     18.676 8.802e-05 ***
## ## yf            2     3788      0.306  0.858263
## ## gen:rate      2     1364      0.110  0.946461
## ## gen:nitro     2     30822      2.488  0.288289
## ## gen:regulator 4     37269      3.008  0.556507
## ## rate:nitro    1     1488      0.120  0.728954
## ## rate:regulator 2     49296      3.979  0.136795
## ## nitro:regulator 2     41019      3.311  0.191042
## ## residual (MS)           12391

## End(Not run)

```

| | |
|----------------|---|
| blackman.wheat | <i>Multi-environment trial of conventional and semi-dwarf wheat varieties</i> |
|----------------|---|

Description

Yield for conventional and semi-dwarf wheat varieties at 7 locs with low/high fertilizer levels.

Format

A data frame with 168 observations on the following 5 variables.

gen genotype

loc location

nitro nitrogen fertilizer, low/high

yield yield (g/m²)

type type factor, conventional/semi-dwarf

Details

Conducted in U.K. in 1975. Each loc had three reps, two nitrogen treatments.

Locations were Begbroke, Boxworth, Crafts Hill, Earith, Edinburgh, Fowlmere, Trumpington.

At the two highest-yielding locations, Earith and Edinburgh, yield was `_lower_` for the high-nitrogen treatment. Blackman et al. say "it seems probable that effects on development and structure of the crop were responsible for the reductions in yield at high nitrogen".

Source

Blackman, JA and Bingham, J. and Davidson, JL (1978). Response of semi-dwarf and conventional winter wheat varieties to the application of nitrogen fertilizer. *The Journal of Agricultural Science*, 90, 543–550. <http://doi.org/10.1017/S0021859600056070>

References

Gower, J. and Lubbe, S.G. and Gardner, S. and Le Roux, N. (2011). *Understanding Biplots*, Wiley.

Examples

```
data(blackman.wheat)
dat <- blackman.wheat

require(lattice)

# Semi-dwarf generally higher yielding than conventional
# bwplot(yield~type|loc,dat, main="blackman.wheat")

# Peculiar interaction--Ear/Edn locs have reverse nitro response
dotplot(gen~yield|loc, dat, group=nitro, auto.key=TRUE,
        main="blackman.wheat: yield for low/high nitrogen")

# Height data from table 6 of Blackman. Height at Trumpington loc.
# Shorter varieties have higher yields, greater response to nitro.
heights <- data.frame(gen=c("Cap", "Dur", "Fun", "Hob", "Hun", "Kin",
                           "Ran", "Spo", "T64", "T68", "T95", "Tem"),
                     ht=c(101,76,76,80,98,88,98,81,86,73,78,93))
dat$height <- heights$ht[match(dat$gen, heights$gen)]
xyplot(yield~height|loc,dat,group=nitro,type=c('p','r'),
      main="blackman.wheat",
      subset=loc=="Tru", auto.key=TRUE)

## Not run:
require(reshape2)
# AMMI-style biplot Fig 6.4 of Gower 2011
dat$env <- factor(paste(dat$loc,dat$nitro,sep="-"))
datm <- acast(dat, gen~env, value.var='yield')
datm <- sweep(datm, 1, rowMeans(datm))
datm <- sweep(datm, 2, colMeans(datm))
biplot(prcomp(datm), main="blackman.wheat AMMI-style biplot")
```



```
## End(Not run)
```

| | |
|--------------|---|
| bliss.borers | <i>Corn borer infestation under four treatments</i> |
|--------------|---|

Description

Corn borer infestation under four treatments

Format

A data frame with 48 observations on the following 3 variables.

borers number of borers per hill

treat treatment factor

freq frequency of the borer count

Details

Four treatments to control corn borers. Treatment 1 is the control.

In 15 blocks, for each treatment, 8 hills of plants were examined, and the number of corn borers present was recorded. The data here are aggregated across blocks.

Bliss mentions that the level of infestation varied significantly between the blocks.

Source

C. Bliss and R. A. Fisher. (1953). Fitting the Negative Binomial Distribution to Biological Data. *Biometrics*, 9, 176–200. Table 3. <http://doi.org/10.2307/3001850>

Geoffrey Beall. 1940. The Fit and Significance of Contagious Distributions when Applied to Observations on Larval Insects. *Ecology*, 21, 460-474. Page 463. <http://doi.org/10.2307/1930285>

Examples

```
data(bliss.borers)
dat <- bliss.borers

# Add 0 frequencies
dat0 <- expand.grid(borers=0:26, treat=c('T1','T2','T3','T4'))
dat0 <- merge(dat0,dat, all=TRUE)
dat0$freq[is.na(dat0$freq)] <- 0

# Expand to individual (non-aggregated) counts for each hill
dd <- data.frame(borers = rep(dat0$borers, times=dat0$freq),
                 treat = rep(dat0$treat, times=dat0$freq))

require(lattice)
```

```

histogram(~borers|treat, dd, type='count', breaks=0:27-.5,
          layout=c(1,4), main="bliss.borers", xlab="Borers per hill")

if(require(MASS)){
  m1 <- glm.nb(borers~0+treat, data=dd)
  # Bliss, table 3, presents treatment means, which are matched by:
  exp(coef(m1)) # 4.033333 3.166667 1.483333 1.508333
  # Bliss gives treatment values k = c(1.532,1.764,1.333,1.190).
  # The mean of these is 1.45, similar to this across-treatment estimate
  m1$theta # 1.47
}

# Plot observed and expected distributions for treatment 2
if(require(latticeExtra)){
  xx <- 0:26
  yy <- dnbinom(0:26, mu=3.17, size=1.47)*120 # estimates are from glm.nb
  histogram(~borers, dd, type='count', subset=treat=='T2',
            main="bliss.borers - trt T2 observed and expected",
            breaks=0:27-.5) +
    xyplot(yy~xx, col='navy', type='b')
}

# "Poissonness"-type plot
if(require(vcd)) {
  dat2 <- droplevels(subset(dat, treat=='T2'))
  vcd::distplot(dat2$borers, type = "nbinomial",
                main="bliss.borers neg binomialness plot")
  # Better way is a rootogram
  g1 <- vcd::goodfit(dat2$borers, "nbinomial")
  plot(g1, main="bliss.borers - Treatment 2")
}

```

bond.diallel

Diallel cross of winter beans

Description

Diallel cross of winter beans

Format

A data frame with 36 observations on the following 3 variables.

female female parent

male male parent

yield yield, grams/plot

stems stems per plot
 nodes podded nodes per stem
 pods pods per podded node
 seeds seeds per pod
 weight weight (g) per 100 seeds
 height height (cm) in April
 width width (cm) in April
 flower mean flowering date in May

Details

Yield in grams/plot for diallel crosses between inbred lines of winter beans. Values are means over two years.

Source

D. A. Bond (1966). Yield and components of yield in diallel crosses between inbred lines of winter beans (*Vicia faba*). *The Journal of Agricultural Science*, 67, 325–336. <http://doi.org/10.1017/S0021859600017329>

References

Peter John, *Statistical Design and Analysis of Experiments*, p. 85.

Examples

```

data(bond.diallel)
dat <- bond.diallel

require(lattice)
splom(dat[,3:11], main="bond.diallel")

# Needs an example. Bond says yield heterosis of F1 hybrids over parent
# means is 22.56, but I cannot match.

# See man page for FDdata in R package sommer

```

 box.cork

Weight of cork samples on four sides of trees

Description

The cork data gives the weights of cork borings of the trunk for 28 trees on the north (N), east (E), south (S) and west (W) directions.

Format

Data frame with 28 observations on the following 5 variables.

```
tree tree number
dir direction N,E,S,W
y weight of cork deposit (centigrams), north direction
```

Source

C.R. Rao (1948). Tests of significance in multivariate analysis. *Biometrika*, 35, 58-79. <http://doi.org/10.2307/2332629>

References

K.V. Mardia, J.T. Kent and J.M. Bibby (1979) *Multivariate Analysis*, Academic Press.
 Russell D Wolfinger, (1996). Heterogeneous Variance: Covariance Structures for Repeated Measures. *Journal of Agricultural, Biological, and Environmental Statistics*, 1, 205-230.

Examples

```
data(box.cork)
dat <- box.cork

if(require(lattice) & require(reshape2)){
  dat2 <- acast(dat, tree ~ dir, value.var='y')
  splom(dat2, pscales=3,
        prepanel.limits = function(x) c(25,100),
        main="box.cork", xlab="Cork yield on side of tree",
        panel=function(x,y,...){
          panel.splom(x,y,...)
          panel.abline(0,1,col="gray80")
        })
}

# -----

## Not run:
require(plotrix)
require(reshape2)
## Each tree is one line
dat2 <- acast(dat, tree ~ dir, value.var='y')
radial.plot(dat2, start=pi/2, rp.type='p', clockwise=TRUE,
            radial.lim=c(0,100), main="box.cork",
            lwd=2, labels=c('North','East','South','West'),
            line.col=rep(c("royalblue","red","#009900","dark orange",
                          "#999999","#a6761d","deep pink"),
                       length=nrow(dat2)))

## End(Not run)

# -----
```

```

## Not run:
# asreml3
require(asreml)

data(box.cork)
dat <- box.cork

# Unstructured covariance
dat$dir <- factor(dat$dir)
dat$tree <- factor(dat$tree)

dat <- dat[order(dat$tree, dat$dir), ]

# Unstructured covariance matrix
m1 <- asreml(y~dir, data=dat,
             rcov = ~ tree:us(dir, init=rep(200,10)))
## Note: 'rcor' is a personal function to extract the correlation
## round(rcor(m1)$dir, 2)
##      E      N      S      W
## E 219.93 223.75 229.06 171.37
## N 223.75 290.41 288.44 226.27
## S 229.06 288.44 350.00 259.54
## W 171.37 226.27 259.54 226.00

# Factor Analytic with different specific variances
# Note: Wolfinger used a common diagonal variance
m2 <- update(m1, rcov=~tree:facv(dir,1))
## round(rcor(m2)$dir, 2)
##      E      N      S      W
## E 219.94 209.46 232.85 182.27
## N 209.46 290.41 291.82 228.43
## S 232.85 291.82 349.99 253.94
## W 182.27 228.43 253.94 225.99

## End(Not run)

# -----

## Not run:
## require(asreml4)

## data(box.cork)
## dat <- box.cork

## # Unstructured covariance
## dat$dir <- factor(dat$dir)
## dat$tree <- factor(dat$tree)

## dat <- dat[order(dat$tree, dat$dir), ]

## # Unstructured covariance matrix

```

```

## m1 <- asreml(y~dir, data=dat,
##             resid = ~ tree:us(dir, init=rep(200,10)))

## library(lucid)
## vc(m1)

## # Note: 'rcor' is a personal function to extract the correlation
## # round(rcor(m1)$dir, 2)
## #      E      N      S      W
## # E 219.93 223.75 229.06 171.37
## # N 223.75 290.41 288.44 226.27
## # S 229.06 288.44 350.00 259.54
## # W 171.37 226.27 259.54 226.00

## # Factor Analytic with different specific variances
## # Note: Wolfinger used a common diagonal variance
## # FIXME - does not work with asreml4
## m2 <- update(m1, resid = ~tree:fa(dir,1))
## # round(rcor(m2)$dir, 2)
## #      E      N      S      W
## # E 219.94 209.46 232.85 182.27
## # N 209.46 290.41 291.82 228.43
## # S 232.85 291.82 349.99 253.94
## # W 182.27 228.43 253.94 225.99

## End(Not run)

```

bradley.multi.uniformity

Uniformity trial of 4 crops on the same land

Description

Uniformity trial of 4 crops on the same land in Trinidad.

Usage

```
data("bradley.multi.uniformity")
```

Format

A data frame with 440 observations on the following 5 variables.

row row

col column

yield yield, pounds per plot

season season

crop crop

Details

Experiments conducted in Trinidad.

Plots were marked in May 1939 in Fields 1, 2, and 3. Prior to 1939 it was difficult to obtain significant results on this land.

Plots were 1/40 acre each, 33 feet square. Discard between blocks (the rows) was 7 feet and between plots (the columns) was 4 feet. For roadways, a gap of 14 feet is between blocks 10 and 11 and a gap of 10 feet between plots E/F (which we call columns 5/6).

Data was collected for 4 crops. Two other crops had poor germination and were omitted.

Field width: $10 \text{ plots} * 33 \text{ feet} + 8 \text{ gaps} * 4 \text{ feet} + 1 \text{ gap} * 10 = 372 \text{ feet}$

Field length: $11 \text{ blocks (plots)} * 33 \text{ feet} + 9 \text{ gaps} * 7 \text{ feet} + 1 \text{ gap} * 14 \text{ feet} = 440 \text{ feet}$

Crop 1. Woolly Pyrol. Crop cut at flowering and weighed in pounds. Note, woolly pyrol appears to be a bean also called black gram, *phaseolus mungo*.

Crop 2. Woolly Pyrol. Crop cut at flowering and weighed in pounds.

Crop 3. Maize. Net weight of cobs in pounds. Source document also has number of cobs.

Crop 4. Yams. Weights in pounds. Source document has weight to 1/4 pound, which has here been rounded to the nearest pound. (Half pounds were rounded to nearest even pound.) Source document also has number of yams.

Notes by Bradley.

The edges of the field tended to be slightly higher yielding. Thought to be "due to the heavier cultivation which the edges receive" (p. 18).

The plot in row 9, col 7 (9G in Bradley) is higher yielding than its neighbors, thought to be the site of a saman tree dug up and burned when the field was plotted. Bits of charcoal were still in the soil.

Bradley also examined soil samples on selected plots and looked at nutrients, moisture, texture, etc. The selected plots were 4 high-yielding plots and 4 low-yielding plots. Little difference was observed. Unexpectedly, yams gave higher yield on plots with more compaction.

Source

P. L. Bradley (1941). A study of the variation in productivity over a number of fixed plots in field 2. Dissertation: The University of the West Indies. Appendix 1a, 1b, 1c, 1d. <http://hdl.handle.net/2139/41264>

References

None

Examples

```
data(bradley.multi.uniformity)

# figures similar to Bradley, pages 11-15
dat1 <- subset(bradley.multi.uniformity, season==1)
if(require(desplot)){
  desplot(yield ~ col*row, dat1,
          flip=TRUE, aspect=433/366, # true aspect (omits roadways))
}
```

```

        main="bradley.multi.uniformity - season 1, woolly pyrol")
    }

dat2 <- subset(bradley.multi.uniformity, season==2)
if(require(desplot)){
  desplot(yield ~ col*row, dat2,
          flip=TRUE, aspect=433/366, # true aspect (omits roadways)
          main="bradley.multi.uniformity - season 2, woolly pyrol")
}

dat3 <- subset(bradley.multi.uniformity, season==3)
if(require(desplot)){
  desplot(yield ~ col*row, dat3,
          flip=TRUE, aspect=433/366, # true aspect (omits roadways)
          main="bradley.multi.uniformity - season 3, maize")
}

dat4 <- subset(bradley.multi.uniformity, season==4)
if(require(desplot)){
  desplot(yield ~ col*row, dat4,
          flip=TRUE, aspect=433/366, # true aspect (omits roadways)
          main="bradley.multi.uniformity - season 4, yams")
}

## Not run:
# to combine plots across seasons, each yield value was converted to percent
# of maximum yield in that season. Same as Bradley, page 17.
dat1$percent <- dat1$yield / max(dat1$yield) * 100
dat2$percent <- dat2$yield / max(dat2$yield) * 100
dat3$percent <- dat3$yield / max(dat3$yield) * 100
dat4$percent <- dat4$yield / max(dat4$yield) * 100
# make sure data is in same order, then combine
dat1 <- dat1[order(dat1$col, dat1$row),]
dat2 <- dat2[order(dat2$col, dat2$row),]
dat3 <- dat3[order(dat3$col, dat3$row),]
dat4 <- dat4[order(dat4$col, dat4$row),]
dat14 <- dat1[,c('row', 'col')]
dat14$fertility <- dat1$percent + dat2$percent + dat3$percent + dat4$percent
if(require(desplot)){
  desplot(fertility ~ col*row, dat14,
          tick=TRUE, flip=TRUE, aspect=433/366, # true aspect (omits roadways)
          main="bradley.multi.uniformity - fertility")
}

## End(Not run)

```


Description

Rape seed yields for 5 genotypes, 3 years, 9 locations.

Format

A data frame with 135 observations on the following 4 variables.

gen genotype
 year year, numeric
 loc location, 9 levels
 yield yield, kg/ha

Details

The yields are the mean of 4 reps.

Note, in table 2 of Brandle, the value of Triton in 1985 at Bagot is shown as 2355, but should be 2555 to match the means reported in the paper.

Source

Brandle, JE and McVetty, PBE. (1988). Genotype x environment interaction and stability analysis of seed yield of oilseed rape grown in Manitoba. *Canadian Journal of Plant Science*, 68, 381–388.

Used with permission of P. McVetty.

Examples

```
data(brandle.rape)
dat <- brandle.rape

require(lattice)
dotplot(gen~yield|loc, dat, group=year, auto.key=list(columns=3),
        main="brandle.rape, yields per location", ylab="Genotype")

# Matches table 4 of Brandle
round(tapply(dat$yield, dat$gen, mean),0)

## Not run:
# Brandle reports variance components:
# sigma^2_g1: 9369 gy: 14027 g: 72632 resid: 150000
# Brandle analyzed rep-level data, so the residual variance is different.
# The other components are matched by the following analysis.

require(lme4)
require(lucid)
dat$year <- factor(dat$year)
m1 <- lmer(yield ~ year + loc + year:loc + (1|gen) +
          (1|gen:loc) + (1|gen:year), data=dat)
vc(m1)
##      grp      var1 var2 vcov sdcov
```

```
## gen:loc (Intercept) <NA> 9363 96.76
## gen:year (Intercept) <NA> 14030 118.4
##      gen (Intercept) <NA> 72630 269.5
## Residual          <NA> <NA> 75010 273.9
```

```
## End(Not run)
```

```
brandt.switchback      Switchback trial of milk yield for two feed mixtures in cattle
```

Description

Switchback trial of milk yield for two feed mixtures in cattle

Usage

```
data("brandt.switchback")
```

Format

A data frame with 30 observations on the following 5 variables.

```
group group: A,B
cow cow, 10 levels
trt treatment, 2 levels
period period, 3 levels
yield milk yield, pounds
```

Details

In this experiment, 10 cows were selected from the Iowa State College Holstein-Friesian herd and divided into two equal groups. Care was taken to have the groups as nearly equal as possible with regard to milk production, stage of gestation, body weight, condition and age. These cows were each given 10 pounds of timothy hay and 30 pounds of corn silage daily but were fed different grain mixtures. Treatment T1, then, consisted of feeding a grain mixture of 1 part of corn and cob meal to 1 part of ground oats, while treatment T2 consisted of feeding a grain mixture of 4 parts corn and cob meal, 4 parts of ground oats and 3 parts of gluten feed. The three treatment periods covered 105 days – three periods of 35 days each. The yields for the first 7 days of each period were not considered because of the possible effect of the transition from one treatment to the other. The data, together with sums and differences which aid in the calculations incidental to testing, are given in table 2.

It seems safe to conclude that the inclusion of gluten feed in the grain mixture fed in a timothy hay ration to Holstein-Friesian cows increased the production of milk. The average increase was 21.7 pounds per cow for a 28-day period.

Source

A.E. Brandt (1938). Tests of Significance in Reversal or Switchback Trials Iowa State College, Agricultural Research Bulletins. Bulletin 234. Book 22. http://lib.dr.iastate.edu/ag_researchbulletins/22/

Examples

```
data(brandt.switchback)
dat <- brandt.switchback

# In each period, treatment 2 is slightly higher
# bwplot(yield~trt|period,dat, layout=c(3,1), main="brandt.switchback",
#       xlab="Treatment", ylab="Milk yield")

# Yield at period 2 (trt T2) is above the trend in group A,
# below the trend (trt T1) in group B.
# Equivalently, treatment T2 is above the trend line
require(lattice)
xyplot(yield~period|group, data=dat, group=cow, type=c('l','r'),
       auto.key=list(columns=5), main="brandt.switchback",
       xlab="Period. Group A: T1,T2,T1. Group B: T2,T1,T2",
       ylab="Milk yield (observed and trend) per cow")

# Similar to Brandt Table 10
m1 <- aov(yield~period+group+cow:group+period:group, data=dat)
anova(m1)
```

bridges.cucumber

Multi-environment trial of cucumbers in a latin square design

Description

Cucumber yields in latin square design at two locs.

Format

A data frame with 32 observations on the following 5 variables.

loc location

gen genotype/cultivar

row row

col column

yield weight of marketable fruit per plot

Details

Conducted at Clemson University in 1985. four cucumber cultivars were grown in a latin square design at Clemson, SC, and Tifton, GA.

Separate variances are modeled each location.

Plot dimensions are not given.

Bridges (1989) used this data to illustrate fitting a heterogeneous mixed model.

Source

William Bridges, 1989. Analysis of a plant breeding experiment with heterogeneous variances using mixed model equations. *Applications of mixed models in agriculture and related disciplines, S. Coop. Ser. Bull.*, 45–51.

Used with permission of William Bridges.

Examples

```
data(bridges.cucumber)
dat <- bridges.cucumber
dat <- transform(dat, rowf=factor(row), colf=factor(col))

if(require(desplot)){
  desplot(yield~col*row|loc, data=dat,
          # aspect unknown
          text=gen, cex=1,
          main="bridges.cucumber")
}

# -----

## Not run:
# Graphical inference test for heterogenous variances
require(nullabor)
# Create a lineup of datasets
fun <- null_permute("loc")
dat20 <- lineup(fun, dat, n=20, pos=9)

# Now plot
library(lattice)
bwplot(yield ~ loc|factor(.sample), dat20,
       main="bridges.cucumber - graphical inference")

## End(Not run)

# -----

## Not run:
# asreml3
require(asreml)
## Random row/col/resid. Same as Bridges 1989, p. 147
```

```

m1 <- asreml(yield ~ 1 + gen + loc + loc:gen,
             random = ~ rowf:loc + colf:loc, data=dat)

require(lucid)
vc(m1)
##           effect component std.error z.ratio constr
## rowf:loc!rowf.var    31.62    23.02    1.4    pos
## colf:loc!colf.var    18.08    15.32    1.2    pos
##           R!variance    31.48    12.85    2.4    pos

## Random row/col/resid at each loc. Matches p. 147
m2 <- asreml(yield ~ 1 + gen + loc + loc:gen,
             random = ~ at(loc):rowf + at(loc):colf, data=dat,
             rcov = ~at(loc):units)

vc(m2)
##           effect component std.error z.ratio constr
## at(loc, Clemson):rowf!rowf.var    32.32    36.58    0.88    pos
## at(loc, Tifton):rowf!rowf.var    30.92    28.63    1.1    pos
## at(loc, Clemson):colf!colf.var    22.55    28.78    0.78    pos
## at(loc, Tifton):colf!colf.var    13.62    14.59    0.93    pos
##           loc_Clemson!variance    46.85    27.05    1.7    pos
##           loc_Tifton!variance    16.11    9.299    1.7    pos

predict(m2, data=dat, classify='loc:gen')$predictions$pvals
##   loc   gen Predicted Std Err   Status
## Clemson Dasher    45.55  5.043 Estimable
## Clemson Guardian    31.62  5.043 Estimable
## Clemson Poinsett    21.42  5.043 Estimable
## Clemson Sprint    25.95  5.043 Estimable
## Tifton Dasher    50.48  3.894 Estimable
## Tifton Guardian    38.72  3.894 Estimable
## Tifton Poinsett    33.01  3.894 Estimable
## Tifton Sprint    39.18  3.894 Estimable

# Is a heterogeneous model justified? Maybe not.
m1$loglik
## -67.35585
m2$loglik
## -66.35621

## End(Not run)

# -----

## Not run:
## require(asreml4)
## ## Random row/col/resid. Same as Bridges 1989, p. 147
## m1 <- asreml(yield ~ 1 + gen + loc + loc:gen,
##             random = ~ rowf:loc + colf:loc, data=dat)

## require(lucid)
## vc(m1)

```

```

## ## effect component std.error z.ratio bound
## ## rowf:loc 31.62 23.02 1.4 P 0
## ## colf:loc 18.08 15.32 1.2 P 0
## ## units(R) 31.48 12.85 2.4 P 0

## ## Random row/col/resid at each loc. Matches p. 147
## m2 <- asreml(yield ~ 1 + gen + loc + loc:gen,
##             random = ~ at(loc):rowf + at(loc):colf, data=dat,
##             resid = ~ dsum( ~ units|loc))
## vc(m2)
## ## effect component std.error z.ratio bound
## ## at(loc, Clemson):rowf 32.32 36.58 0.88 P 0
## ## at(loc, Tifton):rowf 30.92 28.63 1.1 P 0
## ## at(loc, Clemson):colf 22.55 28.78 0.78 P 0
## ## at(loc, Tifton):colf 13.62 14.59 0.93 P 0
## ## loc_Clemson(R) 46.85 27.05 1.7 P 0
## ## loc_Tifton(R) 16.11 9.299 1.7 P 0

## predict(m2, data=dat, classify='loc:gen')$pvals
## ## loc gen predicted.value std.error status
## ## 1 Clemson Dasher 45.6 5.04 Estimable
## ## 2 Clemson Guardian 31.6 5.04 Estimable
## ## 3 Clemson Poinsett 21.4 5.04 Estimable
## ## 4 Clemson Sprint 26 5.04 Estimable
## ## 5 Tifton Dasher 50.5 3.89 Estimable
## ## 6 Tifton Guardian 38.7 3.89 Estimable
## ## 7 Tifton Poinsett 33 3.89 Estimable
## ## 8 Tifton Sprint 39.2 3.89 Estimable

## # Is a heterogeneous model justified? Maybe not.
## m1$loglik
## ## -67.35585
## m2$loglik
## ## -66.35621

## End(Not run)

```

broadbalk.wheat

Long term wheat yields on Broadbalk fields at Rothamsted.

Description

Long term wheat yields on Broadbalk fields at Rothamsted.

Format

A data frame with 1258 observations on the following 4 variables.

year year

plot plot
 grain grain yield, tonnes
 straw straw yield, tonnes

Details

Rothamsted Experiment station conducted wheat experiments on the Broadbalk Fields beginning in 1844 with data for yields of grain and straw collected from 1852 to 1925. Ronald Fisher was hired to analyze data from the agricultural trials. Organic manures and inorganic fertilizer treatments were applied in various combinations to the plots.

N1 is 48kg, N1.5 is 72kg, N2 is 96kg, N4 is 192kg nitrogen.

| Plot | Treatment |
|------|--|
| 2b | manure |
| 3 | No fertilizer or manure |
| 5 | P K Na Mg (No N) |
| 6 | N1 P K Na Mg |
| 7 | N2 P K Na Mg |
| 8 | N3 P K Na Mg |
| 9 | N1* P K Na Mg since 1894; 9A and 9B received different treatments 1852-93 |
| 10 | N2 |
| 11 | N2 P |
| 12 | N2 P Na* |
| 13 | N2 P K |
| 14 | N2 P Mg* |
| 15 | N2 P K Na Mg (timing of N application different to other plots, see below) |
| 16 | N4 P K Na Mg 1852-64; unmanured 1865-83; N2*P K Na Mg since 1884 |
| 17 | N2 applied in even years; P K Na Mg applied in odd years |
| 18 | N2 applied in odd years; P K Na Mg applied in even years |
| 19 | N1.5 P and rape cake 1852-78, 1879-1925 rape cake only |

Source

D.F. Andrews and A.M. Herzberg. 1985. *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. Springer.

Retrieved from <http://lib.stat.cmu.edu/datasets/Andrews/>

References

Broadbalk Winter Wheat Experiment. <http://www.era.rothamsted.ac.uk/index.php?area=home&page=index&dataset=4>

Examples

```
data(broadbalk.wheat)
dat <- broadbalk.wheat
```

```

require(lattice)
## xyplot(grain~straw|plot, dat, type=c('p','smooth'), as.table=TRUE,
##       main="broadbalk.wheat")
xyplot(grain~year|plot, dat, type=c('p','smooth'), as.table=TRUE,
       main="broadbalk.wheat") # yields are decreasing

# See the treatment descriptions to understand the patterns
redblue <- colorRampPalette(c("firebrick", "lightgray", "#375997"))
levelplot(grain~year*plot, dat, main="broadbalk.wheat: Grain", col.regions=redblue)

```

 burgueno.alpha

Incomplete block alpha design

Description

Incomplete block alpha design

Usage

```
data("burgueno.alpha")
```

Format

A data frame with 48 observations on the following 6 variables.

rep rep, 3 levels

block block, 12 levels

row row

col column

gen genotype, 16 levels

yield yield, numeric

Details

A field experiment with 3 reps, 4 blocks per rep, laid out as an alpha design.

The plot size is not given.

Source

J Burgueno, A Cadena, J Crossa, M Banziger, A Gilmour, B Cullis. 2000. *User's guide for spatial analysis of field variety trials using ASREML*. CIMMYT. https://books.google.com/books?id=PR_tYCFyLCYC&pg=PA1

Electronic version of the data obtained from CropStat software.

Used with permission of Juan Burgueno.

Examples

```

data(burgueno.alpha)
dat <- burgueno.alpha

if(require(desplot)){
  desplot(yield~col*row, dat,
          out1=rep, out2=block, # aspect unknown
          text=gen, cex=1,shorten="none",
          main='burgueno.alpha')
}

## Not run:
require(lme4)
require(lucid)
# Inc block model
m0 <- lmer(yield ~ gen + (1|rep/block), data=dat)
vc(m0) # Matches Burgueno p. 26
##      grp      var1 var2  vcov sdcor
## block:rep (Intercept) <NA> 86900 294.8
##      rep (Intercept) <NA> 200900 448.2
## Residual      <NA> <NA> 133200 365

## End(Not run)

# -----

## Not run:
# asreml3
require(asreml)

dat <- transform(dat, xf=factor(col), yf=factor(row))
dat <- dat[order(dat$xf, dat$yf),]

# Sequence of models on page 36

m1 <- asreml(yield ~ gen, data=dat)
m1$loglik # -232.13

m2 <- asreml(yield ~ gen, data=dat,
            random = ~ rep)
m2$loglik # -223.48

# Inc Block model
m3 <- asreml(yield ~ gen, data=dat,
            random = ~ rep/block)
m3$loglik # -221.42
m3$coef$fixed # Matches solution on p. 27

# AR1xAR1 model
m4 <- asreml(yield ~ 1 + gen, data=dat,

```

```

      rcov = ~ar1(xf):ar1(yf))
m4$loglik # -221.47
plot(variogram(m4), main="burgueno.alpha") # Figure 1

m5 <- asreml(yield ~ 1 + gen, data=dat,
             random= ~ yf, rcov = ~ar1(xf):ar1(yf))
m5$loglik # -220.07

m6 <- asreml(yield ~ 1 + gen + pol(yf,-2), data=dat,
             rcov = ~ar1(xf):ar1(yf))
m6$loglik # -204.64 vs. 203.69

m7 <- asreml(yield ~ 1 + gen + lin(yf), data=dat,
             random= ~ spl(yf), rcov = ~ar1(xf):ar1(yf))
m7$loglik # -212.51

m8 <- asreml(yield ~ 1 + gen + lin(yf), data=dat,
             random= ~ spl(yf))
m8$loglik # -213.91

# Polynomial model with predictions
m9 <- asreml(yield ~ 1 + gen + pol(yf,-2) + pol(xf,-2), data=dat,
             random= ~ spl(yf), rcov = ~ar1(xf):ar1(yf))
m9$loglik # -191.44 vs -189.61
#p9 <- predict(m9, classify="gen:xf:yf", levels=list(xf=1,yf=1))
#p9$predictions

m10 <- asreml(yield ~ 1 + gen + lin(yf)+lin(xf), data=dat,
              rcov = ~ar1(xf):ar1(yf))
m10$loglik # -211.56

m11 <- asreml(yield ~ 1 + gen + lin(yf)+lin(xf), data=dat,
              random= ~ spl(yf), rcov = ~ar1(xf):ar1(yf))
m11$loglik # -208.90

m12 <- asreml(yield ~ 1 + gen + lin(yf)+lin(xf), data=dat,
              random= ~ spl(yf)+spl(xf), rcov = ~ar1(xf):ar1(yf))
m12$loglik # -206.82

m13 <- asreml(yield ~ 1 + gen + lin(yf)+lin(xf), data=dat,
              random= ~ spl(yf)+spl(xf))
m13$loglik # -207.52

## End(Not run)

# -----

## Not run:
## require(asreml4)

## dat <- transform(dat, xf=factor(col), yf=factor(row))
## dat <- dat[order(dat$xf, dat$yf),]

```

```
## # Sequence of models on page 36

## m1 <- asreml(yield ~ gen, data=dat)
## m1$loglik # -232.13

## m2 <- asreml(yield ~ gen, data=dat,
##             random = ~ rep)
## m2$loglik # -223.48

## # Inc Block model
## m3 <- asreml(yield ~ gen, data=dat,
##             random = ~ rep/block)
## m3$loglik # -221.42
## m3$coef$fixed # Matches solution on p. 27

## # AR1xAR1 model
## m4 <- asreml(yield ~ 1 + gen, data=dat,
##             resid = ~ar1(xf):ar1(yf))
## m4$loglik # -221.47
## plot(varioGram(m4), main="burgueno.alpha") # Figure 1

## m5 <- asreml(yield ~ 1 + gen, data=dat,
##             random= ~ yf, resid = ~ar1(xf):ar1(yf))
## m5$loglik # -220.07

## m6 <- asreml(yield ~ 1 + gen + pol(yf,-2), data=dat,
##             resid = ~ar1(xf):ar1(yf))
## m6$loglik # -204.64 vs. 203.69

## m7 <- asreml(yield ~ 1 + gen + lin(yf), data=dat,
##             random= ~ spl(yf), resid = ~ar1(xf):ar1(yf))
## m7$loglik # -212.51

## m8 <- asreml(yield ~ 1 + gen + lin(yf), data=dat,
##             random= ~ spl(yf))
## m8$loglik # -213.91

## # Polynomial model with predictions
## m9 <- asreml(yield ~ 1 + gen + pol(yf,-2) + pol(xf,-2), data=dat,
##             random= ~ spl(yf),
##             resid = ~ar1(xf):ar1(yf))
## m9 <- update(m9)
## m9$loglik # -191.44 vs -189.61
## p9 <- predict(m9, classify="gen:xf:yf", levels=list(xf=1,yf=1))
## p9

## m10 <- asreml(yield ~ 1 + gen + lin(yf)+lin(xf), data=dat,
##             resid = ~ar1(xf):ar1(yf))
## m10$loglik # -211.56

## m11 <- asreml(yield ~ 1 + gen + lin(yf)+lin(xf), data=dat,
##             random= ~ spl(yf),
```

```

##           resid = ~ar1(xf):ar1(yf))
## m11$loglik # -208.90

## m12 <- asreml(yield ~ 1 + gen + lin(yf)+lin(xf), data=dat,
##           random= ~ spl(yf)+spl(xf),
##           resid = ~ar1(xf):ar1(yf))
## m12$loglik # -206.82

## m13 <- asreml(yield ~ 1 + gen + lin(yf)+lin(xf), data=dat,
##           random= ~ spl(yf)+spl(xf))
## m13$loglik # -207.52

## End(Not run)

```

| | |
|-----------------|--------------------------|
| burgueno.rowcol | <i>Row-column design</i> |
|-----------------|--------------------------|

Description

Row-column design

Usage

```
data("burgueno.rowcol")
```

Format

A data frame with 128 observations on the following 5 variables.

```

rep rep, 2 levels
row row
col column
gen genotype, 64 levels
yield yield, tons/ha

```

Details

A field experiment with two contiguous replicates in 8 rows, 16 columns.
The plot size is not given.

Source

J Burgueno, A Cadena, J Crossa, M Banziger, A Gilmour, B Cullis. 2000. *User's guide for spatial analysis of field variety trials using ASREML*. CIMMYT.

Electronic version of the data obtained from CropStat software.

Used with permission of Juan Burgueno.

Examples

```

data(burgueno.rowcol)
dat <- burgueno.rowcol

# Two contiguous reps in 8 rows, 16 columns
if(require(desplot)){
  desplot(yield ~ col*row, data=dat,
          out1=rep, # aspect unknown
          text=gen, shorten="none", cex=.75,
          main="burgueno.rowcol")
}

## Not run:
require(lme4)
require(lucid)

# Random rep, row and col within rep
m1 <- lmer(yield ~ gen + (1|rep) + (1|rep:row) + (1|rep:col), data=dat)
vc(m1) # Match components of Burgueno p. 40
##      grp      var1 var2  vcov  sdcor
## rep:col (Intercept) <NA> 0.2189 0.4679
## rep:row (Intercept) <NA> 0.1646 0.4057
##      rep (Intercept) <NA> 0.1916 0.4378
## Residual          <NA> <NA> 0.1796 0.4238

## End(Not run)

# -----

## Not run:
# asreml3
require(asreml)
# AR1 x AR1 with linear row/col effects, random spline row/col
dat <- transform(dat, xf=factor(col), yf=factor(row))
dat <- dat[order(dat$xf,dat$yf),]
m2 <- asreml(yield ~ gen + lin(yf) + lin(xf), data=dat,
            random = ~ spl(yf) + spl(xf),
            rcov = ~ ar1(xf):ar1(yf))
m2 <- update(m2) # More iterations

# Scaling of spl components has changed in asreml from old versions
require(lucid)
vc(m2) # Match Burgueno p. 42
##      effect component std.error z.ratio constr
## spl(yf) 0.09077 0.08252 1.1 pos
## spl(xf) 0.08108 0.0821 0.99 pos
## R!variance 0.1482 0.03119 4.8 pos
## R!xf.cor 0.1152 0.2269 0.51 uncon
## R!yf.cor 0.009436 0.2414 0.039 uncon
## plot(variogram(m2), main="burgueno.rowcol")

```

```
## End(Not run)

# -----

## Not run:
## require(asreml4)
## # AR1 x AR1 with linear row/col effects, random spline row/col
## dat <- transform(dat, xf=factor(col), yf=factor(row))
## dat <- dat[order(dat$xf,dat$yf),]
## m2 <- asreml(yield ~ gen + lin(yf) + lin(xf), data=dat,
##             random = ~ spl(yf) + spl(xf),
##             resid = ~ ar1(xf):ar1(yf))
## m2 <- update(m2) # More iterations

## # Scaling of spl components has changed in asreml from old versions
## require(lucid)
## vc(m2) # Match Burgueno p. 42
## ##          effect component std.error z.ratio bound
## ##          spl(yf)  0.09077   0.08252   1.1      P 0
## ##          spl(xf)  0.08107   0.08209   0.99     P 0
## ##          xf:yf(R) 0.1482    0.03119   4.8      P 0
## ##          xf:yf!xf!cor 0.1152   0.2269   0.51     U 0.1
## ##          xf:yf!yf!cor 0.009467 0.2414   0.039    U 0.9

## plot(varioGram(m2), main="burgueno.rowcol")

## End(Not run)
```

burgueno.unreplicated *Field experiment with unreplicated genotypes plus one repeated check.*

Description

Field experiment with unreplicated genotypes plus one repeated check.

Usage

```
data("burgueno.unreplicated")
```

Format

A data frame with 434 observations on the following 4 variables.

gen genotype, 281 levels

col column

row row

yield yield, tons/ha

Details

A field experiment with 280 new genotypes. A check genotype is planted in every 4th column.
The plot size is not given.

Source

J Burgueno, A Cadena, J Crossa, M Banziger, A Gilmour, B Cullis. 2000. *User's guide for spatial analysis of field variety trials using ASREML*. CIMMYT.

Electronic version of the data obtained from CropStat software.

Used with permission of Juan Burgueno.

Examples

```
data(burgueno.unreplicated)
dat <- burgueno.unreplicated

# Define a 'check' variable for colors
dat$check <- ifelse(dat$gen=="G000", 2, 1)
# Every fourth column is the 'check' genotype
if(require(desplot)){
  desplot(yield ~ col*row, data=dat,
          col=check, num=gen, #text=gen, cex=.3, # aspect unknown
          main="burgueno.unreplicated")
}

# -----

## Not run:
# asreml3
require(asreml)
require(lucid)

# AR1 x AR1 with random genotypes
dat <- transform(dat, xf=factor(col), yf=factor(row))
dat <- dat[order(dat$xf,dat$yf),]
m2 <- asreml(yield ~ 1, data=dat, random = ~ gen,
            rcov = ~ ar1(xf):ar1(yf))
vc(m2)
# Note the strong saw-tooth pattern in the variogram. Seems to
# be column effects.
plot(variogram(m2), xlim=c(0,15), ylim=c(0,9), zlim=c(0,0.5),
     main="burgueno.unreplicated - AR1xAR1")
# library(lattice) # Show how odd columns are high
# bwplot(resid(m2) ~ col, data=dat, horizontal=FALSE)

# Define an even/odd column factor as fixed effect
# dat$oddcoll <- factor(dat$col
# The modulus operator throws a bug, so do it the hard way.
dat$oddcoll <- factor(dat$col - floor(dat$col / 2) *2 )
```

```

m3 <- update(m2, yield ~ 1 + oddcol)
m3$loglik # Matches Burgueno table 3, line 3
plot(variogram(m3), xlim=c(0,15), ylim=c(0,9), zlim=c(0,0.5),
     main="burgueno.unreplicated - AR1xAR1 + Even/Odd")
# Much better-looking variogram

## End(Not run)

# -----

## Not run:
## require(asreml4)
## require(lucid)

## # AR1 x AR1 with random genotypes
## dat <- transform(dat, xf=factor(col), yf=factor(row))
## dat <- dat[order(dat$xf,dat$yf),]
## m2 <- asreml(yield ~ 1, data=dat, random = ~ gen,
##             resid = ~ ar1(xf):ar1(yf))
## vc(m2)
## ##          effect component std.error z.ratio bound
## ##          gen      0.9122  0.127      7.2      P 0
## ##          xf:yf(R)  0.4993  0.05601   8.9      P 0
## ##          xf:yf!xf!cor -0.2431  0.09156  -2.7      U 0
## ##          xf:yf!yf!cor  0.1255  0.07057   1.8      U 0.1

## # Note the strong saw-tooth pattern in the variogram. Seems to
## # be column effects.
## plot(varioGram(m2), xlim=c(0,15), ylim=c(0,9), zlim=c(0,0.5),
##      main="burgueno.unreplicated - AR1xAR1")
## # library(lattice) # Show how odd columns are high
## # bwplot(resid(m2) ~ col, data=dat, horizontal=FALSE)

## # Define an even/odd column factor as fixed effect
## # dat$soddcol <- factor(dat$col
## # The modulus operator throws a bug, so do it the hard way.
## # dat$soddcol <- factor(dat$col - floor(dat$col / 2) * 2 )

## m3 <- update(m2, yield ~ 1 + oddcol)
## m3$loglik # Matches Burgueno table 3, line 3

## plot(varioGram(m3), xlim=c(0,15), ylim=c(0,9), zlim=c(0,0.5),
##      main="burgueno.unreplicated - AR1xAR1 + Even/Odd")
## # Much better-looking variogram

## End(Not run)

```


Description

Maize yields in a multi-environment trial. Pedigree included.

Format

A data frame with 245 observations on the following 5 variables.

gen genotype

male male parent

female female parent

env environment

yield yield, Mg/ha

Details

Ten inbreds were crossed to produce a diallel without reciprocals. The 45 F1 crosses were evaluated along with 4 checks in a triple-lattice 7x7 design. Pink stem borer infestation was natural.

Experiments were performed in 1995 and 1996 at three sites in northwestern Spain: Pontevedra (42 deg 24 min N, 8 deg 38 min W, 20 m over sea), Pontecaldelas (42 deg 23 N, 8 min 32 W, 300 m above sea), Ribadumia (42 deg 30 N, 8 min 46 W, 50 m above sea).

A two-letter location code and the year are concatenated to define the environment.

The average number of larvae per plant in each environment:

| Env | Larvae |
|------|--------|
| pc95 | 0.54 |
| pc96 | 0.91 |
| ri96 | 1.78 |
| pv95 | 2.62 |
| pv96 | 3.35 |

Source

Butron, A and Velasco, P and Ordás, A and Malvar, RA. 2004. Yield evaluation of maize cultivars across environments with different levels of pink stem borer infestation. *Crop science*, 44, 741-747. <http://doi.org/10.2135/cropsci2004.7410>

Used with permission of Ana Butron.

Examples

```
data(butron.maize)
dat <- butron.maize
```

```
# -----
```

```
if(require(reshape2)){
```

```

mat <- acast(dat, gen~env, value.var='yield')
mat <- sweep(mat, 2, colMeans(mat))
mat.svd <- svd(mat)
# Calculate PC1 and PC2 scores as in Table 4 of Butron
# Comment out to keep Rcmd check from choking on '%*%'
# round(mat.svd$u[,1:2] %*% diag(sqrt(mat.svd$d[1:2])) %*% diag(c(-1,1)),3)

biplot(princomp(mat), main="butron.maize", cex=.7) # Figure 1 of Butron
}

# -----

## Not run:

# Here we see if including pedigree information is helpful for a
# multi-environment model
# Including the pedigree provided little benefit

# Create the pedigree
ped <- dat[, c('gen','male','female')]
ped <- ped[!duplicated(ped),] # remove duplicates
unip <- unique(c(ped$male, ped$female)) # Unique parents
unip <- unip[!is.na(unip)]
# We have to define parents at the TOP of the pedigree
ped <- rbind(data.frame(gen=c("Dent","Flint"), # genetic groups
                        male=c(0,0),
                        female=c(0,0)),
            data.frame(gen=c("A509","A637","A661","CM105","EP28",
                            "EP31","EP42","F7","PB60","Z77016"),
                        male=rep(c('Dent','Flint'),each=5),
                        female=rep(c('Dent','Flint'),each=5)),
            ped)
ped[is.na(ped$male),'male'] <- 0
ped[is.na(ped$female),'female'] <- 0

# View the pedigree. Can't use kinship2 ... plants can be both male/female
# Not the best view...too much overplotting
require(synbreed)
pe <- with(ped[1:57,], create.pedigree(gen, male, female, gener=NULL))
windows(9,6)
synbreed::plot.pedigree(pe, vertex.size=10, vertex.label.cex=.5, asp=.5) #

## End(Not run)

# -----

## Not run:
# asreml3
require(asreml)
ped.ainv <- asreml.Ainverse(ped)$ginv

m0 <- asreml(yield ~ 1+env, random = ~ gen, data=dat)
m1 <- asreml(yield ~ 1+env, random = ~ ped(gen), ginverse=list(gen=ped.ainv), data=dat)

```

```

m2 <- update(m1, random = ~ id(env):ped(gen))
m3 <- update(m2, random = ~ diag(env):ped(gen))
m4 <- update(m3, random = ~ fa(env,1):ped(gen))
## AIC(m0,m1,m2,m3,m4)
##   df      AIC
## m0  2 229.4037
## m1  2 213.2487
## m2  2 290.6156
## m3  6 296.8061
## m4 11 218.1568

p0 <- predict(m0, data=dat, classify="gen")$pred$pvals
p4 <- predict(m4, data=dat, classify="gen")$pred$pvals
p4par <- p4[1:12,] # parents
p4 <- p4[-c(1:12),] # hybrids
# Careful! Need to manually sort the predictions
p0 <- p0[order(as.character(p0$gen)),]
p4 <- p4[order(as.character(p4$gen)),]

# lims <- range(c(p0$pred, p4$pred)) * c(.95,1.05)
lims <- c(6,8.25) # zoom in on the higher-yielding hybrids
plot(p0$predicted.value, p4$predicted.value,
     pch="", xlim=lims, ylim=lims, main="butron.maize",
     xlab="BLUP w/o pedigree", ylab="BLUP with pedigree")
abline(0,1,col="lightgray")
text(x=p0$predicted.value, y=p4$predicted.value, p0$gen, cex=.5, srt=-45)
text(x=min(lims), y=p4par$predicted.value, p4par$gen, cex=.5)

## End(Not run)

# -----

## Not run:

## require(asreml4)
## ped.ainv <- ainverse(ped)

## if(FALSE){
##   # asreml4 does not use ginverse
##   m0 <- asreml(yield ~ 1+env, data=dat, random = ~ gen)
##   m1 <- asreml(yield ~ 1+env, random = ~ ped(gen), ginverse=list(gen=ped.ainv), data=dat)
##   m2 <- update(m1, random = ~ id(env):ped(gen))
##   m3 <- update(m2, random = ~ diag(env):ped(gen))
##   m4 <- update(m3, random = ~ fa(env,1):ped(gen))
##   ## AIC(m0,m1,m2,m3,m4)
##   ##   df      AIC
##   ## m0  2 229.4037
##   ## m1  2 213.2487
##   ## m2  2 290.6156
##   ## m3  6 296.8061
##   ## m4 11 218.1568

##   p0 <- predict(m0, data=dat, classify="gen")$pred$pvals

```

```

## p4 <- predict(m4, data=dat, classify="gen")$pred$pvals
## p4par <- p4[1:12,] # parents
## p4 <- p4[-c(1:12),] # hybrids
## # Careful! Need to manually sort the predictions
## p0 <- p0[order(as.character(p0$gen)),]
## p4 <- p4[order(as.character(p4$gen)),]

## # lims <- range(c(p0$pred, p4$pred)) * c(.95,1.05)
## lims <- c(6,8.25) # zoom in on the higher-yielding hybrids
## plot(p0$predicted.value, p4$predicted.value,
##      pch="", xlim=lims, ylim=lims, main="butron.maize",
##      xlab="BLUP w/o pedigree", ylab="BLUP with pedigree")
## abline(0,1,col="lightgray")
## text(x=p0$predicted.value, y=p4$predicted.value, p0$gen, cex=.5, srt=-45)
## text(x=min(lims), y=p4par$predicted.value, p4par$gen, cex=.5)
## # Including the pedigree provided little benefit
## } # if(FALSE)

## End(Not run) # dontrun

```

byers.apple

Diameters of apples

Description

Measurements of the diameters of apples

Format

A data frame with 480 observations on the following 6 variables.

tree tree, 10 levels

apple apple, 24 levels

size size of apple

appleid unique id number for each apple

time time period, 1-6 = (week/2)

diameter diameter, inches

Details

Experiment conducted at the Winchester Agricultural Experiment Station of Virginia Polytechnic Institute and State University. Twentyfive apples were chosen from each of ten apple trees.

Of these, there were 80 apples in the largest size class, 2.75 inches in diameter or greater.

The diameters of the apples were recorded every two weeks over a 12-week period.

Source

Schabenberger, Oliver and Francis J. Pierce. 2002. *Contemporary Statistical Models for the Plant and Soil Sciences*. CRC Press, Boca Raton, FL.

Examples

```
data(byers.apple)
dat <- byers.apple

require(lattice)
xyplot(diameter ~ time | factor(appleid), data=dat, type=c('p','l'),
       strip=strip.custom(par.strip.text=list(cex=.7)),
       main="byers.apple")

## Not run:
# Overall fixed linear trend, plus random intercept/slope deviations
# for each apple. Observations within each apple are correlated.
require(nlme)
require(lucid)
m1 <- lme(diameter ~ 1 + time, data=dat,
         random = ~ time|appleid, method='ML',
         cor = corAR1(0, form=~ time|appleid),
         na.action=na.omit)

vc(m1)
##      effect  variance  stddev corr
## (Intercept) 0.007354  0.08575  NA
##      time 0.00003632 0.006027 0.83
##      Residual 0.0004555 0.02134  NA

## End(Not run)
```

 caribbean.maize

Multi-environment trial of maize with fertilization

Description

Maize fertilization trial on Antigua and St. Vincent.

Format

A data frame with 612 observations on the following 7 variables.

isle island, 2 levels
 site site
 block block
 plot plot, numeric
 trt treatment factor

ears number of ears harvested
 yield yield in kilograms

Details

Antigua is a coral island in the Caribbean with sufficient level land for experiments and a semi-arid climate, while St. Vincent is volcanic and level areas are uncommon, but the rainfall can be seasonally heavy.

Plots were 16 feet by 18 feet. A central area 12 feet by 12 feet was harvested and recorded.

The number of ears harvested was only recorded on the isle of Antigua.

The digits of the treatment represent the levels of nitrogen, phosphorus, and potassium fertilizer, respectively.

The TEAN site suffered damage from goats on plot 27, 35 and 36. The LFAN site suffered damage from cattle on one boundary—plots 9, 18, 27, 36.

Source

D.F. Andrews and A.M. Herzberg. 1985. *Data: A Collection of Problems from Many Fields for the Student and Research Worker*.

Retrieved from <http://lib.stat.cmu.edu/datasets/Andrews/>

References

Also in the DAAG package as data sets `antigua`, `stVincent`.

Examples

```
data(caribbean.maize)
dat <- caribbean.maize

# Yield and ears are correlated
require(lattice)
xyplot(yield~ears|site, dat, ylim=c(0,10), subset=isle=="Antigua",
       main="caribbean.maize - Antigua")

# Some locs show large response to nitrogen (as expected), e.g. UISV, OOSV
dotplot(trt~yield|site, data=dat, main="caribbean.maize treatment response")

# The pattern is a bit hard to see, so we split the treatment factor
# into separate factors, and group sites by island
dat <- transform(dat, N=factor(substring(trt,2,2)),
                P=factor(substring(trt,3,3)),
                K=factor(substring(trt,4,4)))
dat <- transform(dat, env=paste(substring(isle,1,1),site,sep="-"))
# Now we can see the strong N*site interaction
bwplot(yield~N|env, dat,
       main="caribbean.maize", xlab="nitrogen")
```

carlson.germination *Germination of alfalfa seeds at various salt concentrations*

Description

Germination of alfalfa seeds at various salt concentrations

Usage

```
data("carlson.germination")
```

Format

A data frame with 120 observations on the following 3 variables.

gen genotype factor, 15 levels

germ germination percent, 0-100

nacl salt concentration percent, 0-2

Details

Data are means averaged over 5, 10, 15, and 20 day counts. Germination is expressed as a percent of the no-salt control to account for differences in germination among the cultivars.

Source

Carlson, JR and Ditterline, RL and Martin, JM and Sands, DC and Lund, RE. (1983). Alfalfa Seed Germination in Antibiotic Agar Containing NaCl. *Crop science*, 23, 882-885. <http://doi.org/10.2135/cropsci1983.0011183X>

Examples

```
data(carlson.germination)
dat <- carlson.germination
dat$germ <- dat$germ/100 # Convert to percent

# Separate response curve for each genotype.
# Really, we should use a glmm with random int/slope for each genotype
m1 <- glm(germ~ 0 + gen*nacl, data=dat, family=quasibinomial)

# Plot data and fitted model
if(require(latticeExtra)){
  newd <- data.frame(expand.grid(gen=levels(dat$gen), nacl=seq(0,2,length=100)))
  newd$pred <- predict(m1, newd, type="response")
  xyplot(germ~nacl|gen, dat, as.table=TRUE, main="carlson.germination",
         xlab="Percent NaCl", ylab="Fraction germinated") +
  xyplot(pred~nacl|gen, newd, type='l', grid=list(h=1,v=0))
}
```

```

# Calculate LD50 values. Note, Carlson et al used quadratics, not glm.
# MASS::dose.p cannot handle multiple slopes, so do a separate fit for
# each genotype. Results are vaguely similar to Carlson table 5.
if(require(MASS)){
  for(ii in unique(dat$gen)){
    cat("\n", ii, "\n")
    mm <- glm(germ ~ 1 + nacl, data=dat, subset=gen==ii, family=quasibinomial(link="probit"))
    print(dose.p(mm))
  }
  ##           Dose           SE
  ## Anchor   1.445728  0.05750418
  ## Apollo   1.305804  0.04951644
  ## Baker    1.444153  0.07653989
  ## Drylander 1.351201  0.03111795
  ## Grimm    1.395735  0.04206377
}

```

carmer.density

Nonlinear maize yield-density model

Description

Nonlinear maize yield-density model.

Format

A data frame with 32 observations on the following 3 variables.

gen genotype/hybrid, 8 levels

pop population (plants)

yield yield, pounds per hill

Details

Eight single-cross hybrids were in the experiment—Hy2xOh7 and WF9xC103 were included because it was believed they had optimum yields at relatively high and low populations. Planted in 1963. Plots were thinned to 2, 4, 6, 8 plants per hill, giving densities 8, 16, 24, 32 thousand plants per acre. Hills were in rows 40 inches apart. One hill = 1/4000 acre. Split-plot design with 5 reps, density is main plot and subplot was hybrid.

Source

S G Carmer and J A Jackobs (1965). An Exponential Model for Predicting Optimum Plant Density and Maximum Corn Yield. *Agronomy Journal*, 57, 241–244. <http://doi.org/10.2134/agronj1965.00021962005700030003x>

Examples

```

data(carmer.density)
dat <- carmer.density
dat$gen <- factor(dat$gen, levels=c('Hy2x0h7', 'WF9xC103', 'R61x187-2',
                                   'WF9x38-11', 'WF9xB14', 'C103xB14',
                                   '0h43xB37', 'WF9xH60'))

# Separate analysis for each hybrid
# Model:  $y = x * a * k^x$ . Table 1 of Carmer and Jackobs.
out <- data.frame(a=rep(NA,8), k=NA)
preds <- NULL
rownames(out) <- levels(dat$gen)
newdat <- data.frame(pop=seq(2,8,by=.1))
for(i in levels(dat$gen)){
  print(i)
  dati <- subset(dat, gen==i)
  mi <- nls(yield ~ pop * a * k^pop, data=dati, start=list(a=10,k=1))
  out[i, ] <- mi$m$getPars()
  # Predicted values
  pi <- cbind(gen=i, newdat, pred= predict(mi, newdat=newdat))
  preds <- rbind(preds, pi)
}
# Optimum plant density is  $-1/\log(k)$ 
out$pop.opt <- -1/log(out$k)
round(out, 3)
##           a      k pop.opt
## Hy2x0h7  0.782 0.865  6.875
## WF9xC103  1.039 0.825  5.192
## R61x187-2 0.998 0.798  4.441
## WF9x38-11 1.042 0.825  5.203
## WF9xB14   1.067 0.806  4.647
## C103xB14  0.813 0.860  6.653
## 0h43xB37  0.673 0.862  6.740
## WF9xH60   0.858 0.854  6.358

# Fit an overall fixed-effect with random deviations for each hybrid.
require(nlme)
m1 <- nlme(yield ~ pop * a * k^pop,
           fixed = a + k ~ 1,
           random = a + k ~ 1|gen,
           data=dat, start=c(a=10,k=1))
# summary(m1) # Random effect for 'a' probably not needed

if(require(latticeExtra)){
  # Plot Data, fixed-effect prediction, random-effect prediction.
  pdat <- expand.grid(gen=levels(dat$gen), pop=seq(2,8,length=50))
  pdat$pred <- predict(m1, pdat)
  pdat$predf <- predict(m1, pdat, level=0)
}

```

```

xyplot(yield~pop|gen, dat, pch=16, as.table=TRUE,
       main="carmer.density models",
       key=simpleKey(text=c("Data", "Fixed effect", "Random effect"),
                    col=c("blue", "red", "darkgreen"), columns=3, points=FALSE)) +
xyplot(predf~pop|gen, pdat, type='l', as.table=TRUE, col="red") +
xyplot(pred~pop|gen, pdat, type='l', col="darkgreen", lwd=2)
}

```

cate.potassium

Relative cotton yield for different soil potassium concentrations

Description

Relative cotton yield for different soil potassium concentrations

Format

A data frame with 24 observations on the following 2 variables.

yield Relative yield

potassium Soil potassium, ppm

Details

Cate & Nelson used this data to determine the minimum optimal amount of soil potassium to achieve maximum yield.

Note, Fig 1 of Cate & Nelson does not match the data from Table 2. It sort of appears that points with high-concentrations of potassium were shifted left to a truncation point. Also, the calculations below do not quite match the results in Table 1. Perhaps the published data were rounded?

Source

Cate, R.B. and Nelson, L.A. (1971). A simple statistical procedure for partitioning soil test correlation data into two classes. *Soil Science Society of America Journal*, 35, 658–660. <http://doi.org/10.2136/sssaj1971.036159950>

Examples

```

data(cate.potassium)
dat <- cate.potassium
names(dat) <- c('y', 'x')

CateNelson <- function(dat){
  dat <- dat[order(dat$x),] # Sort the data by x
  x <- dat$x
  y <- dat$y

  # Create a data.frame to store the results

```

```

out <- data.frame(x=NA, mean1=NA, css1=NA, mean2=NA, css2=NA, r2=NA)

css <- function(x) { var(x) * (length(x)-1) }
tcss <- css(y) # Total corrected sum of squares

for(i in 2:(length(y)-2)){
  y1 <- y[1:i]
  y2 <- y[-(1:i)]

  out[i, 'x'] <- x[i]
  out[i, 'mean1'] <- mean(y1)
  out[i, 'mean2'] <- mean(y2)
  out[i, 'css1'] <- css1 <- css(y1)
  out[i, 'css2'] <- css2 <- css(y2)
  out[i, 'r2'] <- ( tcss - (css1+css2)) / tcss
}
return(out)
}

cn <- CateNelson(dat)
ix <- which.max(cn$r2)
with(dat, plot(y~x, ylim=c(0,110), xlab="Potassium", ylab="Yield"))
title("cate.potassium - Cate-Nelson analysis")
abline(v=dat$x[ix], col="skyblue")
abline(h=(dat$y[ix] + dat$y[ix+1])/2, col="skyblue")

## Not run:
# another approach with similar results
# https://joe.org/joe/2013october/tt1.php
require("rcompanion")
cateNelson(dat$x, dat$y)

## End(Not run)

```

chinloy.fractionalfactorial

Sugarcane fractional factorial 1/3 3⁵

Description

Sugarcane fractional factorial 1/3 3⁵.

Usage

```
data("chinloy.fractionalfactorial")
```

Format

A data frame with 81 observations on the following 10 variables.

yield yield

block block

row row position

col column position

trt treatment code

N nitrogen treatment: 0, 1, 2

P phosphorous treatment: 0, 1, 2

K potassium treatment: 0, 1, 2

B bagasse treatment: 0, 1, 2

F filter press mud treatment: 0, 1, 2

Details

An experiment grown in 1949 at the Worthy Park Estate in Jamaica.

Nitrogen was applied as sulphate of ammonia at 0, 3, 6 hundred-weight per acre.

Phosphorous was applied as superphosphate at 0, 4, 8 hundred-weight per acre.

Potassium was applied as muriate of potash at 0, 1, 2 hundred-weight per acre.

Bagasse applied pre-plant at 0, 20, 40 tons per acre.

Filter press mud applied pre-plant at 0, 10, 20 tons per acre.

Each plot was 18 yards long by 6 yards (3 rows) wide. Plots were arranged in nine columns of nine, a 2-yard space separating plots along the rows and two guard rows separating plots across the rows.

Field length: 18 yards * 9 plots + 2 yards * 8 gaps = 178 yards

Field width: 6 yards * 9 plots + 4 yards * 8 gaps = 86 yards

Source

T. Chinloy, R. F. Innes and D. J. Finney. (1953). An example of fractional replication in an experiment on sugar cane manuring. *Journ Agricultural Science*, 43, 1-11. <https://doi.org/10.1017/S0021859600044567>

References

None

Examples

```
data(chinloy.fractionalfactorial)
dat <- chinloy.fractionalfactorial
```

```
# Treatments are coded with levels 0,1,2. Make sure they are factors
dat <- transform(dat,
                 N=factor(N), P=factor(P), K=factor(K), B=factor(B), F=factor(F))
```

```

# Experiment layout
if(require(desplot)){
  desplot(yield ~ col*row, dat, out1=block, text=trt, shorten="no", cex=0.6,
          aspect=178/86,
          main="chinloy.fractionalfactorial")
}

# Main effect and some two-way interactions. These match Chinloy table 6.
# Not sure how to code terms like P^2K=B^2F
m1 <- aov(yield ~ block + N + P + K + B + F + N:P + N:K + N:B + N:F, dat)
anova(m1)

```

christidis.competition

Competition between varieties in cotton

Description

Competition between varieties in cotton, measurements taken for each row.

Usage

```
data("christidis.competition")
```

Format

A data frame with 270 observations on the following 8 variables.

```

plot plot
plotrow row within plot
block block
row row, only 1 row
col column
gen genotype
yield yield, kg
height height, cm

```

Details

Nine genotypes/varieties of cotton were used in a variety test. The plots were 100 meters long and 2.40 meters wide, each plot having 3 rows 0.80 meters apart. The layout was an RCB of 5 blocks, each block having 2 replicates of every variety (with the original intention of trying 2 seed treatments). Each row was harvested/weighed separately. After the leaves of the plants had dried up and fallen, the mean height of each row was measured.

Christidis found significant competition between varieties, but not due to height differences. Crude analysis.

TODO: Find a better analysis of this data which incorporates field trends AND competition effects, maybe including a random effect for border rows of all genotype pairs (as neighbors)?

Source

Christidis, Basil G (1935). Intervarietal competition in yield trials with cotton. *The Journal of Agricultural Science*, 25, 231-237. Table 1. <https://doi.org/10.1017/S0021859600009710>

References

None

Examples

```
data(christidis.competition)
dat <- christidis.competition

# Match Christidis Table 2 means
# aggregate(yield ~ gen, aggregate(yield ~ gen+plot, dat, sum), mean)

# Each RCB block has 2 replicates of each genotype
# with(dat, table(block,gen))

if(require(lattice)){

  # Tall plants yield more
  # xyplot(yield ~ height|gen, data=dat)

  # Huge yield variation across field. Also heterogeneous variance.
  xyplot(yield ~ col, dat, group=gen, auto.key=list(columns=5),
         main="christidis.competition")
}

## Not run:
require(lattice)
require(mgcv)
if(is.element("package:gam", search())) detach("package:gam")
# Simple non-competition model to remove main effects
m1 <- gam(yield ~ gen + s(col), data=dat)
p1 <- as.data.frame(predict(m1, type="terms"))
names(p1) <- c('geneff', 'coleff')
dat2 <- cbind(dat, p1)
dat2 <- transform(dat2, res=yield-geneff-coleff)
xyplot(res ~ col, data=dat2, group=gen,
       main="christidis.competition - residuals")

## End(Not run)
```

christidis.wheat.uniformity
Uniformity trial of wheat

Description

Uniformity trial of wheat at Cambridge, UK in 1931.

Usage

```
data("christidis.wheat.uniformity")
```

Format

A data frame with 288 observations on the following 3 variables.

```
row row
col column
yield yield
```

Details

Two blocks, 24 rows each. In block A, each 90-foot row was divided into 12 units, each unit 7.5 feet long. Rows were 8 inches wide.

Field length: 24 rows * 8 inches = 16 feet

Field width: 12 units * 7.5 feet = 90 feet

Source

Christidis, Basil G (1931). The importance of the shape of plots in field experimentation. *The Journal of Agricultural Science*, 21, 14-37. Table VI, p. 28. <http://dx.doi.org/10.1017/S0021859600007942>

References

None

Examples

```
data(christidis.wheat.uniformity)
dat <- christidis.wheat.uniformity

# sum(dat$yield) # Matches Christidis

if(require(desplot)){
  desplot(yield ~ col*row, data=dat,
          flip=TRUE, aspect=16/90, # true aspect
          main="christidis.wheat.uniformity")
}
```

```
}

```

| | |
|----------------|------------------------------------|
| cleveland.soil | <i>Soil resistivity in a field</i> |
|----------------|------------------------------------|

Description

Soil resistivity in a field

Format

A data frame with 8641 observations on the following 5 variables.

```
northing y ordinate
easting x ordinate
resistivity Soil resistivity, ohms
is.ns Indicator of north/south track
track Track number
```

Details

Resistivity is related to soil salinity.

Source

William Cleveland, (1993), *Visualizing Data*. Electronic version from StatLib: <http://lib.stat.cmu.edu/datasets/>
 Cleaned version from Luke Tierney <http://homepage.stat.uiowa.edu/~luke/classes/248/examples/soil>

Examples

```
data(cleveland.soil)
dat <- cleveland.soil

# Similar to Cleveland fig 4.64
## require(latticeExtra)
## redblue <- colorRampPalette(c("firebrick", "lightgray", "#375997"))
## levelplot(resistivity ~ easting + northing, data = dat,
##           col.regions=redblue,
##           panel=panel.levelplot.points,
##           aspect=2.4, xlab= "Easting (km)", ylab= "Northing (km)",
##           main="cleveland")

# 2D loess plot. Cleveland fig 4.68
sg1 <- expand.grid(easting = seq(.15, 1.410, by = .02),
                  northing = seq(.150, 3.645, by = .02))
lo1 <- loess(resistivity~easting*northing, data=dat, span = 0.1, degree = 2)
```



```

fit1 <- predict(lo1, sg1)
require(lattice)
redblue <- colorRampPalette(c("firebrick", "lightgray", "#375997"))
levelplot(fit1 ~ sg1$easting * sg1$northing,
          col.regions=redblue,
          cuts = 9,
          aspect=2.4, xlab = "Easting (km)", ylab = "Northing (km)",
          main="cleveland.soil - 2D smooth of Resistivity")

# -----

## Not run:
# 3D loess plot with data overlaid
require(rgl)
bg3d(color = "white")
clear3d()
points3d(dat$easting, dat$northing, dat$resistivity / 100,
        col = rep("gray50", nrow(dat)))
rgl::surface3d(seq(.15, 1.410, by = .02),
              seq(.150, 3.645, by = .02),
              fit1/100, alpha=0.9, col=rep("wheat", length(fit1)),
              front="fill", back="fill")

rgl.close()

## End(Not run)

```

cochran.beets

Yield and number of plants in a sugarbeet fertilizer experiment

Description

Yield and number of plants in a sugarbeet fertilizer experiment.

Usage

```
data("cochran.beets")
```

Format

A data frame with 42 observations on the following 4 variables.

fert fertilizer treatment

block block

yield yield (tons/acres)

plants number of plants per plot

Details

Yield (tons/acre) and number of beets per plot. Fertilizer treatments combine superphosphate (P), muriate of potash (K), and sodium nitrate (N).

Source

George Snedecor (1946). *Statistical Methods*, 4th ed. Table 12.13, p. 332.

References

H. Fairfield Smith (1957). Interpretation of Adjusted Treatment Means and Regressions in Analysis of Covariance. *Biometrics*, 13, 282-308. <http://doi.org/10.2307/2527917>

Examples

```
data(cochran.beets)
dat = cochran.beets

# P has strong effect
require(lattice)
xyplot(yield ~ plants|fert, dat, main="cochran.beets")
```

cochran.bib

Multi-environment trial of corn, balanced incomplete block design

Description

Balanced incomplete block design in corn

Format

A data frame with 52 observations on the following 3 variables.

loc location/block, 13 levels

gen genotype/line, 13 levels

yield yield, pounds/plot

Details

Incomplete block design. Each loc/block has 4 genotypes/lines. The blocks are planted at different locations.

Conducted in 1943 in North Carolina.

Source

North Carolina Agricultural Experiment Station, United States Department of Agriculture.

References

Cochran, W.G. and Cox, G.M. (1957), *Experimental Designs*, 2nd ed., Wiley and Sons, New York, p. 448.

Examples

```

data(cochran.bib)
dat <- cochran.bib

# Show the incomplete-block structure
require(lattice)
redblue <- colorRampPalette(c("firebrick", "lightgray", "#375997"))
levelplot(yield~loc*gen, dat,
          col.regions=redblue,
          xlab="loc (block)", main="cochran.bib - incomplete blocks")

with(dat, table(gen,loc))
rowSums(as.matrix(with(dat, table(gen,loc))))
colSums(as.matrix(with(dat, table(gen,loc))))

m1 = aov(yield ~ gen + Error(loc), data=dat)
summary(m1)

require(nlme)
m2 = lme(yield ~ -1 + gen, data=dat, random=~1|loc)

```

 cochran.crd

Potato scab infection with sulfur treatments

Description

Potato scab infection with sulfur treatments

Format

A data frame with 32 observations on the following 5 variables.

inf infection percent
 trt treatment factor
 row row
 col column

Details

The experiment was conducted to investigate the effect of sulfur on controlling scab disease in potatoes. There were seven treatments. Control, plus spring and fall application of 300, 600, 1200 lbs/acre of sulfur. The response variable was infection as a percent of the surface area covered with scab. A completely randomized design was used with 8 replications of the control and 4 replications of the other treatments.

Although the original analysis did not show significant differences in the sulfur treatments, including a polynomial trend in the model uncovered significant differences (Tamura, 1988).

Source

W.G. Cochran and G. Cox, 1957. *Experimental Designs*, 2nd ed. John Wiley, New York.

References

Tamura, R.N. and Nelson, L.A. and Naderman, G.C., (1988). An investigation of the validity and usefulness of trend analysis for field plot data. *Agronomy Journal*, 80, 712-718.

<http://doi.org/10.2134/agronj1988.00021962008000050003x>

Examples

```
data(cochran.crd)
dat <- cochran.crd

# Field plan
if(require(desplot)){
  desplot(inf~col*row, data=dat,
          text=trt, cex=1, # aspect unknown
          main="cochran.crd")
}

# CRD anova. Table 6 of Tamura 1988
contrasts(dat$trt) <- cbind(c1=c(1,1,1,-6,1,1,1), # Control vs Sulf
                          c2=c(-1,-1,-1,0,1,1,1)) # Fall vs Sp
m1 <- aov(inf ~ trt, data=dat)
anova(m1)
summary(m1, split=list(trt=list("Control vs Sulf"=1, "Fall vs Spring"=2)))

# Quadratic polynomial for columns...slightly different than Tamura 1988
m2 <- aov(inf ~ trt + poly(col,2), data=dat)
anova(m2)
summary(m2, split=list(trt=list("Control vs Sulf"=1, "Fall vs Spring"=2)))
```

cochran.eelworms

Counts of eelworms before and after fumigant treatments

Description

Counts of eelworms before and after fumigant treatments

Format

A data frame with 48 observations on the following 7 variables.

block block factor, 4 levels

row row

col column
 fumigant fumigant factor
 dose dose, Numeric 0,1,2. Maybe should be a factor?
 initial count of eelworms pre-treatment
 final count of eelworms post-treatment

Details

In the original experiment plan (as shown in Bailey 2008), columns 9, 10, 11 are shifted up slightly. Treatment codes: Con = Control, Chl = Chlorodinitrobenzen, Cym = Cymag, Car = Carbon Disulphide jelly, See = Seekay.

Experiment was conducted in 1935 at Rothamsted Experiment Station. In early March 400 grams of soil were sampled and the number of eelworm cysts were counted. Fumigants were added to the soil, oats were sown and later harvested. In October, the plots were again sampled and the final count of cysts recorded.

Source

Cochran and Cox, 1950. *Experimental Designs*. Table 3.1.

References

R. A. Bailey, 2008. *Design of Comparative Experiments*. Cambridge.

Examples

```
data(cochran.eelworms)
dat <- cochran.eelworms

# Very strong spatial trends
if(require(desplot)){
  desplot(initial ~ col*row, data=dat,
           flip=TRUE, # aspect unknown
           main="cochran.eelworms")
}

# final counts are strongly related to initial counts
require(lattice)
xyplot(final~initial|factor(dose), data=dat, group=fumigant,
       main="cochran.eelworms - by dose (panel) & fumigant",
       xlab="Initial worm count",
       ylab="Final worm count", auto.key=list(columns=5))

# One approach...log transform, use 'initial' as covariate, create 9 treatments
dat <- transform(dat, trt=factor(paste0(fumigant, dose)))
m1 <- aov(log(final) ~ block + trt + log(initial), data=dat)
anova(m1)
```

cochran.factorial *Factorial experiment in beans*

Description

Factorial experiment in beans

Usage

```
data("cochran.factorial")
```

Format

A data frame with 32 observations on the following 4 variables.

```
rep rep factor
block block factor
trt treatment factor, 16 levels
yield yield (pounds)
```

Details

Conducted by Rothamsted Experiment Station in 1936. The treatments are listed below.

d = dung: None, 10 tons/acre.

n = nitrochalk: None, 0.4 hundredweight nitrogen per acre.

p = Superphosphate: None, 0.6 hundredweight per acre

k = muriate of potash: None, 1 hundredweight K20 per acres.

The response variable is the yield of beans.

Source

Cochran, W.G. and Cox, G.M. (1957), *Experimental Designs*, 2nd ed., Wiley and Sons, New York, p. 160.

Examples

```
data(cochran.factorial)
dat <- cochran.factorial

# Split treatment into individual factors
dat <- transform(dat,
  d = -1 + 2 * grepl('d',trt),
  n = -1 + 2 * grepl('n',trt),
  p = -1 + 2 * grepl('p',trt),
  k = -1 + 2 * grepl('k',trt))
```

```

dat <- transform(dat, d=factor(d), n=factor(n), p=factor(p), k=factor(k))

# Cochran table 6.5.
m1 <- lm(yield ~ rep * block + (d+n+p+k)^3, data=dat)
anova(m1)

if(require(FrF2)){
  aliases(m1)
  MEPlot(m1, select=3:6,
         main="cochran.factorial - main effects plot")
}

```

cochran.latin

Latin square design in wheat

Description

Six wheat plots were sampled by six operators and shoot heights measured. The operators sampled plots in six ordered sequences. The dependent variate was the difference between measured height and true height of the plot.

Format

A data frame with 36 observations on the following 4 variables.

row row

col column

operator operator factor

diff difference between measured height and true height

Source

Cochran, W.G. and Cox, G.M. (1957), *Experimental Designs*, 2nd ed., Wiley and Sons, New York.

Examples

```

data(cochran.latin)
dat <- cochran.latin

if(require(desplot)){
  desplot(diff~col*row, dat,
         text=operator, cex=1, # aspect unknown
         main="cochran.latin")
}

```

```
dat <- transform(dat, rf=factor(row), cf=factor(col))
aov.dat <- aov(diff ~ operator + Error(rf*cf), dat)
summary(aov.dat)
model.tables(aov.dat, type="means")
```

`cochran.lattice`*Balanced lattice experiment in cotton*

Description

Balanced lattice experiment in cotton

Usage

```
data("cochran.lattice")
```

Format

A data frame with 80 observations on the following 5 variables.

`y` percent of affected flower buds

`rep` replicate

`row` row

`col` column

`trt` treatment factor

Details

The experiment is a balanced lattice square with 16 treatments in a 4x4 layout in each of 5 replicates. The treatments were applied to cotton plants. Each plot was ten rows wide by 70 feet long (about 1/18 of an acre). (Estimated plot width is 34.5 feet.) Data were collected from the middle 4 rows. The data are the percentages of squares showing attack by boll weevils. A 'square' is the name given to a young flower bud.

The plot orientation is not clear.

Source

William G. Cochran, Gertrude M. Cox. *Experimental Designs*, 2nd Edition. Page 490.

Originally from: F. M. Wadley (1946). Incomplete block designs in insect population problems. *J. Economic Entomology*, 38, 651–654.

References

Walter Federer. Combining Standard Block Analyses With Spatial Analyses Under a Random Effects Model. Cornell Univ Tech Report BU-1373-MA. <http://hdl.handle.net/1813/31971>

Examples

```

data(cochran.lattice)
dat <- cochran.lattice

if(require(desplot)){
  desplot(y~row*col|rep, dat,
          text=trt, # aspect unknown, should be 2 or .5
          main="cochran.lattice")
}

## Not run:
# Random rep,row,column model often used by Federer
require(lme4)
dat <- transform(dat, rowf=factor(row), colf=factor(col))
m1 <- lmer(y ~ trt + (1|rep) + (1|rep:row) + (1|rep:col), data=dat)
summary(m1)

## End(Not run)

```

cochran.wireworms

Wireworms controlled by fumigants in a latin square

Description

Wireworms controlled by fumigants in a latin square

Format

A data frame with 25 observations on the following 4 variables.

row row

col column

trt fumigant treatment, 5 levels

worms count of wireworms per plot

Details

Plots were approximately 22 cm by 13 cm. Layout of the experiment was a latin square. The number of wireworms in each plot was counted, following soil fumigation the previous year.

Source

W. G. Cochran (1938). Some difficulties in the statistical analysis of replicated experiments. *Empire Journal of Experimental Agriculture*, 6, 157–175.

References

W. Cochran (1940). The analysis of variance when experimental errors follow the Poisson or binomial laws. *The Annals of Mathematical Statistics*, 11, 335-347.

G W Snedecor and W G Cochran, 1980. *Statistical Methods*, Iowa State University Press. Page 288.

Examples

```
data(cochran.wireworms)
dat <- cochran.wireworms

if(require(desplot)){
  desplot(worms ~ col*row, data=dat,
          text=trt, cex=1, # aspect unknown
          main="cochran.wireworms")
}

# Trt K is effective, but not the others. Really, this says it all.
require(lattice)
bwplot(worms ~ trt, dat, main="cochran.wireworms", xlab="Treatment")

# Snedecor and Cochran do ANOVA on sqrt(x+1).
dat <- transform(dat, rowf=factor(row), colf=factor(col))
m1 <- aov(sqrt(worms+1) ~ rowf + colf + trt, data=dat)
anova(m1)

# Instead of transforming, use glm
m2 <- glm(worms ~ trt + rowf + colf, data=dat, family="poisson")
anova(m2)

# GLM with random blocking.
if(require(lme4)){
m3 <- glmer(worms ~ -1 +trt +(1|rowf) +(1|colf), data=dat, family="poisson")
summary(m3)
## Fixed effects:
##      Estimate Std. Error z value Pr(>|z|)
## trtK  0.1393     0.4275  0.326  0.745
## trtM  1.7814     0.2226  8.002 1.22e-15 ***
## trtN  1.9028     0.2142  8.881 < 2e-16 ***
## trtO  1.7147     0.2275  7.537 4.80e-14 ***

}
```

Description

Potato yields in single-drill plots

Usage

```
data("connolly.potato")
```

Format

A data frame with 80 observations on the following 6 variables.

rep block

gen variety

row row

col column

yield yield, kg/ha

matur maturity group

Details

Connolly et al use this data to illustrate how yield can be affected by competition from neighboring plots.

This data uses M1, M2, M3 for maturity, while Connolly et al use FE (first early), SE (second early) and M (maincrop).

The trial was 20 sections, each of which was an independent row of 20 drills. The data here are four reps of single-drill plots from sections 1, 6, 11, and 16.

The neighbor covariate for a plot is defined as the average of the plots to the left and right. For drills at the edge of the trial, the covariate was the average of the one neighboring plot yield and the section (i.e. rep) mean.

It would be interesting to fit a model that uses differences in maturity between a plot and its neighbor as the actual covariate. Anyone...?

Source

Connolly, T and Currie, ID and Bradshaw, JE and McNicol, JW. (1993). Inter-plot competition in yield trials of potatoes *Solanum tuberosum* L. with single-drill plots. *Annals of Applied Biology*, 123, 367-377.

<http://doi.org/10.1111/j.1744-7348.1993.tb04099.x>

Used with permission of Iain Currie.

Examples

```

data(connolly.potato)
dat <- connolly.potato

# Field plan
if(require(desplot)){
  desplot(yield~col*row, data=dat,
          out1=rep, # aspect unknown
          main="connolly.potato yields (reps not contiguous)")
}

# Later maturities are higher yielding
require(lattice)
bwplot(yield~matur, dat, main="connolly.potato yield by maturity")

# Observed raw means. Matches Connolly table 2.
mn <- aggregate(yield~gen, data=dat, FUN=mean)
mn[rev(order(mn$yield)),]

# Create a covariate which is the average of neighboring plot yields
if(require(reshape2)){
  mat <- acast(dat, row~col, value.var='yield')
  mat2 <- matrix(NA, nrow=4, ncol=20)
  mat2[,2:19] <- (mat[, 1:18] + mat[, 3:20])/2
  mat2[, 1] <- (mat[, 1] + apply(mat, 1, mean))/2
  mat2[, 20] <- (mat[, 20] + apply(mat, 1, mean))/2
  dat2 <- melt(mat2)
  colnames(dat2) <- c('row','col','cov')
  dat <- merge(dat, dat2)
  # xyplot(yield ~ cov, data=dat, type=c('p','r'))

# Connolly et al fit a model with avg neighbor yield as a covariate
m1 <- lm(yield ~ 0 + gen + rep + cov, data=dat)
coef(m1)['cov'] # = -.303 (Connolly obtained -.31)

# Block names and effects
bnm <- c("R1","R2","R3","R4")
beff <- c(0, coef(m1)[c('repR2','repR3','repR4')])
# Variety names and effects
vnm <- paste0("V", formatC(1:20, width=2, flag='0'))
veff <- coef(m1)[1:20]

# Adjust yield for variety and block effects
dat <- transform(dat, yadj = yield - beff[match(rep,bnm)]
                - veff[match(gen,vnm)])

# Similar to Connolly Fig 1. Point pattern doesn't quite match
xyplot(yadj~cov, data=dat, type=c('p','r'),
       main="connolly.potato",
       xlab="Avg yield of nearest neighbors",

```

```

      ylab="Yield, adjusted for variety and block effects")
    }

```

 cornelius.maize

Multi-environment trial of maize for 9 cultivars at 20 locations.

Description

Maize yields for 9 cultivars at 20 locations.

Usage

```
data("cornelius.maize")
```

Format

A data frame with 180 observations on the following 3 variables.

env environment factor, 20 levels

gen genotype/cultivar, 9 levels

yield yield, kg/ha

Details

Cell means (kg/hectare) for the CIMMYT EVT16B maize yield trial.

Source

P L Cornelius and J Crossa and M S Seyedsadr. (1996). Statistical Tests and Estimators of Multiplicative Models for Genotype-by-Environment Interaction. Book: *Genotype-by-Environment Interaction*. Pages 199-234.

References

Forkman, Johannes and Piepho, Hans-Peter. (2014). Parametric bootstrap methods for testing multiplicative terms in GGE and AMMI models. *Biometrics*, 70(3), 639-647. <http://doi.org/10.1111/biom.12162>

Examples

```

data(cornelius.maize)
dat <- cornelius.maize
# dotplot(gen~yield|env,dat) # We cannot compare genotype yields easily
# Subtract environment mean from each observation
if(require(reshape2)){
  mat <- acast(dat, gen~env)
  mat <- scale(mat, scale=FALSE)
}

```

```

dat2 <- melt(mat)
names(dat2) <- c('gen','env','yield')
require(lattice)
bwplot(yield ~ gen, dat2,
        main="cornelius.maize - environment centered yields")
}

## Not run:
# This reproduces the analysis of Forkman and Piepho.

test.pc <- function(Y0, type="AMMI", n.boot=10000, maxpc=6) {

  # Test the significance of Principal Components in GGE/AMMI

  # Singular value decomposition of centered/double-centered Y
  Y <- sweep(Y0, 1, rowMeans(Y0)) # subtract environment means
  if(type=="AMMI") {
    Y <- sweep(Y, 2, colMeans(Y0)) # subtract genotype means
    Y <- Y + mean(Y0)
  }
  lam <- svd(Y)$d

  # Observed value of test statistic.
  # t.obs[k] is the proportion of variance explained by the kth term out of
  # the k...M terms, e.g. t.obs[2] is lam[2]^2 / sum(lam[2:M]^2)
  t.obs <- { lam^2/rev(cumsum(rev(lam^2))) } [1:(M-1)]
  t.boot <- matrix(NA, nrow=n.boot, ncol=M-1)

  # Centering rows/columns reduces the rank by 1 in each direction.
  I <- if(type=="AMMI") nrow(Y0)-1 else nrow(Y0)
  J <- ncol(Y0)-1
  M <- min(I, J) # rank of Y, maximum number of components
  M <- min(M, maxpc) # Optional step: No more than 5 components

  for(K in 0:(M-2)){ # 'K' multiplicative components in the svd

    for(bb in 1:n.boot){
      E.b <- matrix(rnorm((I-K) * (J-K)), nrow = I-K, ncol = J-K)
      lam.b <- svd(E.b)$d
      t.boot[bb, K+1] <- lam.b[1]^2 / sum(lam.b^2)
    }

  }

  # P-value for each additional multiplicative term in the SVD.
  # P-value is the proportion of time bootstrap values exceed t.obs
  colMeans(t.boot > matrix(rep(t.obs, n.boot), nrow=n.boot, byrow=TRUE))
}

dat <- cornelius.maize

# Convert to matrix format
require(reshape2)

```

```
dat <- acast(dat, env~gen, value.var='yield')

## R> test.pc(dat,"AMMI")
## [1] 0.0000 0.1505 0.2659 0.0456 0.1086 # Forkman: .00 .156 .272 .046 .111

## R> test.pc(dat,"GGE")
## [1] 0.0000 0.2934 0.1513 0.0461 0.2817 # Forkman: .00 .296 .148 .047 .285

## End(Not run)
```

corsten.interaction *Multi-environment trial of corn*

Description

The data is the yield (kg/acre) of 20 genotypes of corn at 7 locations.

Format

A data frame with 140 observations on the following 3 variables.

gen genotype, 20 levels

loc location, 7 levels

yield yield, kg/acre

Details

The data is used by Corsten & Denis (1990) to illustrate two-way clustering by minimizing the interaction sum of squares.

In their paper, the labels on the location dendrogram have a slight typo. The order of the loc labels shown is 1 2 3 4 5 6 7. The correct order of the loc labels is 1 2 4 5 6 7 3.

Source

L C A Corsten and J B Denis, (1990). Structuring Interaction in Two-Way Tables By Clustering. *Biometrics*, 46, 207–215. Table 1. <http://doi.org/10.2307/2531644>

Used with permission of Jean-Baptiste Denis.

Examples

```
data(corsten.interaction)
dat <- corsten.interaction

if(require(reshape2)){
m1 <- melt(dat, measure.var='yield')
dmat <- acast(m1, loc~gen)
```

```
# Corsten (1990) uses this data to illustrate simultaneous row and
# column clustering based on interaction sums-of-squares.
# There is no (known) function in R to reproduce this analysis
# (please contact the package maintainer if this is not true).
# For comparison, the 'heatmap' function clusters the rows and
# columns _independently_ of each other.
heatmap(dmat, main="corsten.interaction")
}
```

| | |
|----------------|---|
| cox.stripsplit | <i>Strip-split plot of barley with fertilizer, calcium, and soil factors.</i> |
|----------------|---|

Description

Strip-split plot of barley with fertilizer, calcium, and soil factors.

Format

A data frame with 96 observations on the following 5 variables.

rep replicate, 4 levels
soil soil, 3 levels
fert fertilizer, 4 levels
calcium calcium, 2 levels
yield yield of winter barley

Details

Four different fertilizer treatments are laid out in vertical strips, which are then split into subplots with different levels of calcium. Soil type is stripped across the split-plot experiment, and the entire experiment is then replicated three times.

Sometimes called a split-block design.

Source

Comes from the notes of Gertrude Cox and A. Rotti.

References

SAS/STAT(R) 9.2 User's Guide, Second Edition. Example 23.5 Strip-Split Plot. <http://support.sas.com/documentation/cdl/en>

Examples

```

data(cox.stripsplit)
dat <- cox.stripsplit

# Raw means
aggregate(yield ~ calcium, data=dat, mean)
aggregate(yield ~ soil, data=dat, mean)
aggregate(yield ~ calcium, data=dat, mean)

if(require(HH)) {
  interaction2wt(yield ~ rep + soil + fert + calcium, dat,
                x.between=0, y.between=0,
                main="cox.stripsplit")
}

# Traditional AOV
m1 <- aov(yield~ fert*calcium*soil +
          Error(rep/(fert+soil+calcium:fert+soil:fert)),
          data=dat)
summary(m1)

# With balanced data, the following are all basically identical

if(require(lme4)){
  # The 'rep:soil:fert' term causes problems...so we drop it.
  m2 <- lmer(yield ~ fert*soil*calcium + (1|rep) + (1|rep:fert) +
            (1|rep:soil) + (1|rep:fert:calcium), data=dat)
}

# -----
## Not run:

# afex uses Kenward-Rogers approach for denominator d.f.
if(require(afex)){
  mixed(yield ~ fert*soil*calcium + (1|rep) + (1|rep:fert) +
        (1|rep:soil) + (1|rep:fert:calcium) + (1|rep:soil:fert), data=dat,
        control=lmerControl(check.nobs.vs.rankZ="ignore"))
  ##          Effect      stat ndf    ddf F.scaling p.value
  ## 1 (Intercept) 1350.8113   1  3.0009      1  0.0000
  ## 2          fert   3.5619   3  9.0000      1  0.0604
  ## 3          soil   3.4659   2  6.0000      1  0.0999
  ## 4        calcium   1.8835   1 12.0000      1  0.1950
  ## 5      fert:soil   1.2735   6 18.0000      1  0.3179
  ## 6    fert:calcium   4.4457   3 12.0000      1  0.0255
  ## 7    soil:calcium   0.2494   2 24.0000      1  0.7813
  ## 8  fert:soil:calcium  0.3504   6 24.0000      1  0.9027
}

## End(Not run)

```

cramer.cucumber

Cucumber yields and quantitative traits

Description

Cucumber yields and quantitative traits

Usage

```
data("cramer.cucumber")
```

Format

A data frame with 24 observations on the following 9 variables.

cycle cycle

rep replicate

plants plants per plot

flowers number of pistillate flowers

branches number of branches

leaves number of leaves

totalfruit total fruit number

culledfruit culled fruit number

earlyfruit early fruit number

Details

The data are used to illustrate path analysis of the correlations between phenotypic traits.

Source

Christopher S. Cramer, Todd C. Wehner, and Sandra B. Donaghy. 1999. Path Coefficient Analysis of Quantitative Traits. In: Handbook of Formulas and Software for Plant Geneticists and Breeders, page 89.

Used with permission of Christopher Cramer.

References

Cramer, C. S., T. C. Wehner, and S. B. Donaghy. 1999. PATHSAS: a SAS computer program for path coefficient analysis of quantitative data. *J. Hered.* 90, 260-262 <https://doi.org/10.1093/jhered/90.1.260>

Examples

```

data(cramer.cucumber)
dat <- cramer.cucumber

require(lattice)
splom(dat[3:9], main="cramer.cucumber - traits by cycle", group=dat$cycle,
      auto.key=list(columns=3))

## Not run:
# derived traits
dat <- transform(dat,
                 marketable = totalfruit-culledfruit,
                 branchesperplant = branches/plants,
                 nodesperbranch = leaves/(branches+plants),
                 femalenodes = flowers+totalfruit)
dat <- transform(dat,
                 perfenod = (femalenodes/leaves),
                 fruitset = totalfruit/flowers,
                 fruitperplant = totalfruit / plants,
                 marketableperplant = marketable/plants,
                 earlyperplant=earlyfruit/plants)

# just use cycle 1
dat1 <- subset(dat, cycle==1)

# define independent and dependent variables
indep <- c("branchesperplant", "nodesperbranch", "perfenod", "fruitset")
dep0 <- "fruitperplant"
dep <- c("marketable", "earlyperplant")

# standardize trait data for cycle 1
sdat <- data.frame(scale(dat1[1:8, c(indep,dep0,dep)]))

# slopes for dep0 ~ indep
X <- as.matrix(sdat[,indep])
Y <- as.matrix(sdat[,c(dep0)])
# estdep <- solve(t(X)
estdep <- solve(crossprod(X), crossprod(X,Y))
estdep
## branchesperplant 0.7160269
## nodesperbranch   0.3415537
## perfenod         0.2316693
## fruitset         0.2985557

# slopes for dep ~ dep0
X <- as.matrix(sdat[,dep0])
Y <- as.matrix(sdat[,c(dep)])
# estind2 <- solve(t(X)
estind2 <- solve(crossprod(X), crossprod(X,Y))
estind2
## marketable earlyperplant
##    0.97196    0.8828393

```

```

# correlation coefficients for indep variables
corrind=cor(sdat[,indep])
round(corrind,2)
##          branchesperplant nodesperbranch perfenod fruitset
## branchesperplant          1.00          0.52   -0.24    0.09
## nodesperbranch            0.52          1.00   -0.44    0.14
## perfenod                  -0.24         -0.44    1.00    0.04
## fruitset                   0.09          0.14    0.04    1.00

# Correlation coefficients for dependent variables
corrdep=cor(sdat[,c(dep0, dep)])
round(corrdep,2)
##          fruitperplant marketable earlyperplant
## fruitperplant          1.00          0.97          0.88
## marketable             0.97          1.00          0.96
## earlyperplant          0.88          0.96          1.00

result = corrind
result = result*matrix(estdep,ncol=4,nrow=4,byrow=TRUE)
round(result,2) # match SAS output columns 1-4
##          branchesperplant nodesperbranch perfenod fruitset
## branchesperplant          0.72          0.18   -0.06    0.03
## nodesperbranch            0.37          0.34   -0.10    0.04
## perfenod                  -0.17         -0.15    0.23    0.01
## fruitset                   0.07          0.05    0.01    0.30

resdep0 = rowSums(result)
resdep <- cbind(resdep0,resdep0)*matrix(estind2, nrow=4,ncol=2,byrow=TRUE)
colnames(resdep) <- dep
# slightly different from SAS output last 2 columns
round(cbind(fruitperplant=resdep0, round(resdep,2)),2)
##          fruitperplant marketable earlyperplant
## branchesperplant          0.87          0.84          0.76
## nodesperbranch            0.65          0.63          0.58
## perfenod                  -0.08         -0.08         -0.07
## fruitset                   0.42          0.41          0.37

## End(Not run)

```

crampton.pig

Weight gain in pigs for different treatments

Description

Weight gain in pigs for different treatments, with initial weight and feed eaten as covariates.

Usage

```
data("crampton.pig")
```

Format

A data frame with 50 observations on the following 5 variables.

```
treatment feed treatment
rep replicate
weight1 initial weight
feed feed eaten
weight2 final weight
```

Details

A study of the effect of initial weight and feed eaten on the weight gaining ability of pigs with different feed treatments.

The data are extracted from Ostle. It is not clear that 'replicate' is actually a blocking replicate as opposed to a repeated measurement. The original source document needs to be consulted.

Source

Crampton, EW and Hopkins, JW. (1934). The Use of the Method of Partial Regression in the Analysis of Comparative Feeding Trial Data, Part II. *The Journal of Nutrition*, 8, 113-123.

References

Bernard Ostle. *Statistics in Research*

Goulen. *Methods of Statistical Analysis, 1st ed.* Page 256-259.

Examples

```
data(crampton.pig)
dat <- crampton.pig

dat <- transform(dat, gain=weight2-weight1)
require(lattice)
# Trt 4 looks best
xyplot(gain ~ feed, dat, group=treatment, type=c('p','r'),
       auto.key=list(columns=5),
       xlab="Feed eaten", ylab="Weight gain", main="crampton.pig")

# Basic Anova without covariates
m1 <- lm(weight2 ~ treatment + rep, data=dat)
anova(m1)
# Add covariates
m2 <- lm(weight2 ~ treatment + rep + weight1 + feed, data=dat)
anova(m2)
# Remove treatment, test this nested model for significant treatments
m3 <- lm(weight2 ~ rep + weight1 + feed, data=dat)
anova(m2,m3) # p-value .07. F=2.34 matches Ostle
```

crossa.wheat

*Multi-environment trial of wheat for 18 genotypes at 25 locations***Description**

Wheat yields for 18 genotypes at 25 locations

Format

A data frame with 450 observations on the following 3 variables.

loc location

locgroup location group: Grp1-Grp2

gen genotype

gengroup genotype group: W1, W2, W3

yield grain yield, tons/ha

Details

Grain yield from the 8th Elite Selection Wheat Yield Trial to evaluate 18 bread wheat genotypes at 25 locations in 15 countries.

Cross et al. used this data to cluster loctions into 2 mega-environments and clustered genotypes into 3 wheat clusters.

Locations

| Code | Country | Location | Latitude (N) | Elevation (m) |
|------|--------------|-------------|--------------|---------------|
| AK | Algeria | El Khroub | 36 | 640 |
| AL | Algeria | Setif | 36 | 1,023 |
| BJ | Bangladesh | Joydebpur | 24 | 8 |
| CA | Cyprus | Athalassa | 35 | 142 |
| EG | Egypt | E1 Gemmeiza | 31 | 8 |
| ES | Egypt | Sakha | 31 | 6 |
| EB | Egypt | Beni-Suef | 29 | 28 |
| IL | India | Ludhiana | 31 | 247 |
| ID | India | Delhi | 29 | 228 |
| JM | Jordan | Madaba | 36 | 785 |
| KN | Kenya | Njoro | 0 | 2,165 |
| MG | Mexico | Guanajuato | 21 | 1,765 |
| MS | Mexico | Sonora | 27 | 38 |
| MM | Mexico | Michoacfin | 20 | 1,517 |
| NB | Nepal | Bhairahwa | 27 | 105 |
| PI | Pakistan | Islamabad | 34 | 683 |
| PA | Pakistan | Ayub | 32 | 213 |
| SR | Saudi Arabia | Riyadh | 24 | 600 |
| SG | Sudan | Gezira | 14 | 411 |
| SE | Spain | Encinar | 38 | 20 |

| | | | | |
|----|----------|------------|--------|-----|
| SJ | Spain | Jerez | 37 | 180 |
| SC | Spain | Cordoba | 38 | 110 |
| SS | Spain | Sevilla | 38 | 20 |
| TB | Tunisia | Beja | 37 | 150 |
| TC | Thailand | Chiang Mai | 18 820 | |

Source

Crossa, J and Fox, PN and Pfeiffer, WH and Rajaram, S and Gauch Jr, HG. (1991). AMMI adjustment for statistical analysis of an international wheat yield trial. *Theoretical and Applied Genetics*, 81, 27–37. <http://doi.org/10.1007/BF00226108>

Used with permission of Jose' Crossa.

References

Jean-Louis Laffont, Kevin Wright and Mohamed Hanafi (2013). Genotype + Genotype x Block of Environments (GGB) Biplots. *Crop Science*, 53, 2332-2341. <http://doi.org/10.2135/cropsci2013.03.0178>

Examples

```
data(crossa.wheat)
dat <- crossa.wheat

# AMMI biplot. Fig 3 of Crossa et al.
if(require(agricolae)){
  m1 <- with(dat, AMMI(E=loc, G=gen, R=1, Y=yield))
  b1 <- m1$biplot[,1:4]
  b1$PC1 <- -1 * b1$PC1 # Flip vertical
  plot(b1$yield, b1$PC1, cex=0.0,
       text(b1$yield, b1$PC1, cex=.5, labels=row.names(b1),col="brown"),
       main="crossa.wheat AMMI biplot", xlab="Average yield", ylab="PC1", frame=TRUE)
  mn <- mean(b1$yield)
  abline(h=0, v=mn, col='wheat')

  g1 <- subset(b1,type=="GEN")
  text(g1$yield, g1$PC1, rownames(g1), col="darkgreen", cex=.5)

  e1 <- subset(b1,type=="ENV")
  arrows(mn, 0,
        0.95*(e1$yield - mn) + mn, 0.95*e1$PC1,
        col= "brown", lwd=1.8,length=0.1)
}

# GGB example
library(agridat)
data(crossa.wheat)
dat2 <- crossa.wheat

# FIXME: The loc grouping is now in the data
```

```
# Define mega-environment groups of locations
dat2$eg <- ifelse(is.element(dat2$loc,
                           c("KN", "NB", "PA", "BJ", "IL", "TC",
                             "JM", "PI", "AS", "ID", "SC", "SS",
                             "SJ", "MS", "MG", "MM")), "Grp1", "Grp2")

library(gge)
# Specify env.group as column in data frame
m2 <- gge(yield~gen*loc, dat2, env.group=eg, scale=FALSE)
biplot(m2, main="crossa.wheat - GGB biplot")
```

crowder.seeds

Germination of Orobanche seeds for two genotypes and two treatments.

Description

Number of Orobanche seeds tested/germinated for two genotypes and two treatments.

Format

plate Factor for replication
 gen Factor for genotype with levels 073, 075
 extract Factor for extract from bean, cucumber
 germ Number of seeds that germinated
 n Total number of seeds tested

Details

Egyptian broomrape, *orobanche aegyptiaca* is a parasitic plant family. The plants have no chlorophyll and grow on the roots of other plants. The seeds remain dormant in soil until certain compounds from living plants stimulate germination.

Two genotypes were studied in the experiment, *O. aegyptiaca* 73 and *O. aegyptiaca* 75. The seeds were brushed with one of two extracts prepared from either a bean plant or cucumber plant.

The experimental design was a 2x2 factorial, each with 5 or 6 reps of plates.

Source

Crowder, M.J., 1978. Beta-binomial anova for proportions. *Appl. Statist.*, 27, 34-37. <http://doi.org/10.2307/2346223>

References

N. E. Breslow and D. G. Clayton. 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88:9-25. <http://doi.org/10.2307/2290687>
 Y. Lee and J. A. Nelder. 1996. Hierarchical generalized linear models with discussion. *J. R. Statist. Soc. B*, 58:619-678.

Examples

```

data(crowder.seeds)
dat <- crowder.seeds
m1.glm <- m1.glmm <- m1.bb <- m1.hglm <- NA

# ----- Graphic
require(lattice)
dotplot(germ/n~gen|extract, dat, main="crowder.seeds")

# ----- GLM.
# family=binomial() fixes dispersion at 1
# family=quasibinomial() estimates dispersion, had larger std errors
m1.glm <- glm(cbind(germ,n-germ) ~ gen*extract,
              data=dat,
              #family="binomial",
              family=quasibinomial()
             )
summary(m1.glm)

# --- GLMM. Assumes Gaussian random effects
require(MASS)
m1.glmm <- glmmPQL(cbind(germ, n-germ) ~ gen*extract, random= ~1|plate,
                  family=binomial(), data=dat)
summary(m1.glmm)

# ----- AODS3 package
# require(aods3)
# m1.bb <- aodml(cbind(germ, n-germ) ~ gen * extract, data=dat, family="bb")

# ----- HGML package. Beta-binomial with beta-distributed random effects
# require(hglm)
# m1.hglm <- hglm(fixed= germ/n ~ I(gen=="075")*extract, weights=n, data=dat,
#                random=~1|plate, family=binomial(), rand.family=Beta(),
#                fix.disp=1)

## Not run:
# Compare coefficients

round(summary(m1.glm)$coef,2)
##
## (Intercept)          -0.41      0.25  -1.64    0.12
## gen075              -0.15      0.30  -0.48    0.64
## extractcucumber     0.54      0.34   1.58    0.13
## gen075:extractcucumber 0.78      0.42   1.86    0.08

round(summary(m1.glmm)$tTable,2)

```

```

##                               Value Std.Error DF t-value p-value
## (Intercept)                   -0.44    0.25 17   -1.80   0.09
## gen075                         -0.10    0.31 17   -0.34   0.74
## extractcucumber                 0.52    0.34 17    1.56   0.14
## gen075:extractcucumber          0.80    0.42 17    1.88   0.08

## round(summary(m1.bb)$BCoef,2)
##                               Estimate Std. Error z value Pr(> |z|)
## (Intercept)                   -0.44    0.22  -2.04   0.04
## gen075                         -0.10    0.27  -0.36   0.72
## extractcucumber                 0.52    0.30   1.76   0.08
## gen075:extractcucumber          0.80    0.38   2.11   0.03

## round(summary(m1.hglm)$FixCoefMat,2)
##                               Estimate Std. Error t-value Pr(>|t|)
## (Intercept)                   -0.47    0.24  -1.92   0.08
## I(gen == "075")TRUE           -0.08    0.31  -0.25   0.81
## extractcucumber                 0.51    0.33   1.53   0.16
## I(gen == "075")TRUE:extractcucumber 0.83    0.43   1.92   0.08

## End(Not run)

## Not run:

# --- rjags version ---

# JAGS/BUGS. See http://mathstat.helsinki.fi/openbugs/Examples/Seeds.html
# Germination rate depends on p, which is a logit of a linear predictor
# based on genotype and extract, plus random deviation to intercept

# To match the output on the BUGS web page, use: dat$gen=="073".
# We use dat$gen=="075" to compare with the parameterization above.
jdat =list(germ = dat$germ, n = dat$n,
          root = as.numeric(dat$extract=="cucumber"),
          gen = as.numeric(dat$gen=="075"),
          nobs = nrow(dat))

jinit = list(int = 0, gen075 = 0, extcuke = 0, g75ecuke = 0, tau = 10)

# Use logical names (unlike BUGS documentation)
mod.bug =
"model {
  for(i in 1:nobs) {
    germ[i] ~ dbin(p[i], n[i])
    b[i] ~ dnorm(0.0, tau)
    logit(p[i]) <- int + gen075 * gen[i] + extcuke * root[i] +
      g75ecuke * gen[i] * root[i] + b[i]
  }
  int ~ dnorm(0.0, 1.0E-6)
  gen075 ~ dnorm(0.0, 1.0E-6)
  extcuke ~ dnorm(0.0, 1.0E-6)
  g75ecuke ~ dnorm(0.0, 1.0E-6)
  tau ~ dgamma(0.001, 0.001)
}

```

```

    sigma <- 1 / sqrt(tau)
  }"

require(rjags)
oo <- textConnection(mod.bug)
j1 <- jags.model(oo, data=jdat, inits=jinit, n.chains=1)
close(oo)

c1 <- coda.samples(j1, c("int", "gen075", "g75ecuke", "extcuke", "sigma"),
  n.iter=20000)
summary(c1) # Medians are very similar to estimates from hglm
# require(lucid)
# print(vcov(c1),3)
##           Mean      SD      2.5
## extcuke   0.543  0.331 -0.118   0.542   1.2
## g75ecuke  0.807  0.436 -0.0586  0.802   1.7
## gen075   -0.0715 0.309 -0.665  -0.0806  0.581
## int      -0.479  0.241 -0.984  -0.473  -0.0299
## sigma     0.289  0.142  0.0505  0.279   0.596

# Plot observed data with HPD intervals for germination probability
c2 <- coda.samples(j1, c("p"), n.iter=20000)
hpd <- HPDinterval(c2)[[1]]
med <- summary(c2, quantiles=.5)$quantiles
fit <- data.frame(med, hpd)

if(require(latticeExtra)){
  obs <- dotplot(1:21 ~ germ/n, dat, ylab="plate",
    col=as.numeric(dat$gen), pch=substring(dat$extract,1))
  obs + segplot(1:21 ~ lower + upper, data=fit, centers=med)
}

# -----
## --- R2jags version ---

require("agridat")
require("R2jags")
dat <- crowder.seeds

# To match the output on the BUGS web page, use: dat$gen=="073".
# We use dat$gen=="075" to compare with the parameterization above.
jdat =list(germ = dat$germ, n = dat$n,
  root = as.numeric(dat$extract=="cucumber"),
  gen = as.numeric(dat$gen=="075"),
  nobs = nrow(dat))

jinit = list(list(int = 0, gen075 = 0, extcuke = 0, g75ecuke = 0, tau = 10))

mod.bug = function() {
  for(i in 1:nobs) {
    germ[i] ~ dbin(p[i], n[i])
    b[i] ~ dnorm(0.0, tau)
    logit(p[i]) <- int + gen075 * gen[i] + extcuke * root[i] +

```

```

      g75ecuke * gen[i] * root[i] + b[i]
    }
    int ~ dnorm(0.0, 1.0E-6)
    gen075 ~ dnorm(0.0, 1.0E-6)
    extcuke ~ dnorm(0.0, 1.0E-6)
    g75ecuke ~ dnorm(0.0, 1.0E-6)
    tau ~ dgamma(0.001, 0.001)
    sigma <- 1 / sqrt(tau)
  }

parms <- c("int", "gen075", "g75ecuke", "extcuke", "sigma")

j1 <- jags(data=jdat, inits=jinit, parms, model.file=mod.bug,
           n.iter=20000, n.chains=1)
print(j1)
##          mu.vect sd.vect    2.5
## extcuke    0.519   0.325 -0.140  0.325  0.531  0.728  1.158
## g75ecuke   0.834   0.429 -0.019  0.552  0.821  1.101  1.710
## gen075    -0.096   0.305 -0.670 -0.295 -0.115  0.089  0.552
## int       -0.461   0.236 -0.965 -0.603 -0.455 -0.312  0.016
## sigma     0.255   0.148  0.033  0.140  0.240  0.352  0.572
## deviance 103.319   7.489 90.019 98.010 102.770 108.689 117.288

traceplot(as.mcmc(j1))
densityplot(as.mcmc(j1))
HPDinterval(as.mcmc(j1))

## End(Not run)

```

cullis.earlygen

Early generation variety trial in wheat

Description

Early generation variety trial in wheat

Format

A data frame with 670 observations on the following 5 variables.

gen genotype factor

row row

col column

entry entry (genotype) number

yield yield of each plot, kg/ha

weed weed score

Details

The data are from an unreplicated field experiment conducted at Tullibigeal, New South Wales, Australia in 1987-88. In each row, every 6th plot was the variety 'Kite'. Six other standard varieties were randomly interspersed over the trial. Each plot was 15m x 1.8m, "oriented with the longest side with rows".

The 'weed' variable is a visual score on a 0 to 10 scale, 0 = no weeds, 10 = 100 percent weeds.

The replicated check variety was numbered 526. A further 6 replicated commercially available varieties (numbered 527 to 532) were also randomly assigned to plots with between 3 to 5 plots of each. The aim of these trials is to identify and retain the top, say 20 percent of lines for further testing. Cullis et al. (1989) presented an analysis of early generation variety trials that included a one-dimensional spatial analysis. Below, a two-dimensional spatial analysis is presented.

Note: The 'row' and 'col' variables are as in the VSN link below (switched compared to the paper by Cullis et al.)

Field length: 67 plots * 1.8 m = 121 m

Field width: 10 rows * 15 m = 150 m

The orientation is not certain, but the alternative orientation would have a field roughly 20m x 1000m, which seems unlikely.

Source

Brian R. Cullis, Warwick J. Lill, John A. Fisher, Barbara J. Read and Alan C. Gleeson (1989). A New Procedure for the Analysis of Early Generation Variety Trials. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 38, 361-375. <http://doi.org/10.2307/2348066>

References

Unreplicated early generation variety trial in Wheat. <http://www.vsnl.co.uk/software/asreml/htmlhelp/asreml/xwheat.htm>

Examples

```
data(cullis.earlygen)
dat <- cullis.earlygen

# Show field layout of checks. Cullis Table 1.
dat$check <- ifelse(dat$entry < 8, dat$entry, NA)
if(require(desplot)){
  desplot(yield ~ col*row, dat,
          col="check", cex=0.5, flip=TRUE, aspect=121/150, # true aspect
          main="cullis.earlygen (yield)")

  grays <- colorRampPalette(c("white", "#252525"))
  desplot(weed ~ col*row, dat,
          at=0:6-0.5, col.regions=grays(7)[-1],
          flip=TRUE, aspect=121/150, # true aspect
          main="cullis.earlygen (weed)")
}
```

```

require(lattice)
bwplot(yield ~ as.character(weed), dat,
       horizontal=FALSE,
       xlab="Weed score", main="cullis.earlygen")

# -----

## Not run:
# asreml3
require(asreml)
# Start with the standard AR1xAR1 analysis
dat <- transform(dat, xf=factor(col), yf=factor(row))
dat <- dat[order(dat$xf, dat$yf),]
m2 <- asreml(yield ~ weed, data=dat, random= ~gen,
            rcov = ~ ar1(xf):ar1(yf))

# Variogram suggests a polynomial trend
m3 <- update(m2, fixed= yield~weed+pol(col,-1))

# Now add a nugget variance
m4 <- update(m3, random= ~ gen + units)

require(lucid)
vc(m4)
##           effect component std.error z.ratio constr
##      gen!gen.var  73770    10420      7.1   pos
##    units!units.var  30440     8074     3.8   pos
##      R!variance  54720    10630     5.1   pos
##      R!xf.cor     0.38     0.115     3.3 uncon
##      R!yf.cor     0.84     0.045    19   uncon

# Predictions from models m3 and m4 are non-estimable. Why?
# Use model m2 for predictions
predict(m2)$pred
##           gen predicted.value standard.error est.status
## 1   Banks           2723.534      93.14633  Estimable
## 2   Eno008           2981.057     162.85053  Estimable
## 3   Eno009           2978.009     161.56930  Estimable
## 4   Eno010           2821.399     153.96697  Estimable
## 5   Eno011           2991.610     161.53308  Estimable
## 6   Eno012           2771.148     162.21897  Estimable

# Compare AR1 with Moving Grid
require(mvngGrAd)
shape <- list(c(1),
             c(1),
             c(1:4),
             c(1:4))
# sketchGrid(10,10,20,20,shapeCross=shape, layers=1, excludeCenter=TRUE)
m5 <- movingGrid(rows=dat$row, columns=dat$col, obs=dat$yield,
                shapeCross=shape, layers=NULL)
dat$mg <- fitted(m5)
dat$ar1 <- fitted(m2)

```

```

head(dat[ , c('yield','ar1','mg')])
##   yield      ar1      mg
##  1  2652 2467.980 2531.998
## 11  3394 3071.681 3052.160
## 21  3148 2826.188 2807.031
## 31  3426 3026.985 3183.649
## 41  3555 3070.102 3195.910
## 51  3453 3006.352 3510.511
pairs(dat[ , c('yield','ar1','mg')])

## End(Not run)

# -----

## Not run:
## require(asreml4)
## # Start with the standard AR1xAR1 analysis
## dat <- transform(dat, xf=factor(col), yf=factor(row))
## dat <- dat[order(dat$xf, dat$yf),]
## m2 <- asreml(yield ~ weed, data=dat, random= ~gen,
##             resid = ~ ar1(xf):ar1(yf))

## # Variogram suggests a polynomial trend
## m3 <- update(m2, fixed= yield~weed+pol(col,-1))

## # Now add a nugget variance
## m4 <- update(m3, random= ~ gen + units)

## require(lucid)
## vc(m4)
## ##          effect component std.error z.ratio bound
## ##          gen 73780      10420      7.1   P 0
## ##          units 30440      8073      3.8   P 0.1
## ##          xf:yf(R) 54730     10630      5.1   P 0
## ##          xf:yf!xf!cor 0.38    0.115     3.3   U 0
## ##          xf:yf!yf!cor 0.84    0.045     19    U 0

## # Predictions from models m3 and m4 are non-estimable. Why?
## # Use model m2 for predictions
## predict(m2, classify="gen")$pvals
## ##          gen predicted.value std.error  status
## ##  1   Banks      2723.534  93.14719 Estimable
## ##  2   Eno008      2981.056 162.85241 Estimable
## ##  3   Eno009      2978.008 161.57129 Estimable
## ##  4   Eno010      2821.399 153.96943 Estimable
## ##  5   Eno011      2991.612 161.53507 Estimable

## # Compare AR1 with Moving Grid
## require(mvngGrAd)
## shape <- list(c(1),
##              c(1),

```

```

##           c(1:4),
##           c(1:4))
## # sketchGrid(10,10,20,20,shapeCross=shape, layers=1, excludeCenter=TRUE)
## m5 <- movingGrid(rows=dat$row, columns=dat$col, obs=dat$yield,
##                 shapeCross=shape, layers=NULL)
## dat$mg <- fitted(m5)
## dat$ar1 <- fitted(m2)
## head(dat[, c('yield','ar1','mg')])
## ##   yield      ar1      mg
## ## 1  2652 2467.980 2531.998
## ## 11 3394 3071.681 3052.160
## ## 21 3148 2826.188 2807.031
## ## 31 3426 3026.985 3183.649
## ## 41 3555 3070.102 3195.910
## ## 51 3453 3006.352 3510.511
## pairs(dat[, c('yield','ar1','mg')])

## End(Not run)

```

darwin.maize

Darwin's maize data of crossed/inbred plant heights

Description

Darwin's maize data of crossed/inbred plant heights.

Format

A data frame with 30 observations on the following 4 variables.

pot Pot factor, 4 levels

pair Pair factor, 12 levels

type Type factor, self-pollinated, cross-pollinated

height Height, in inches (measured to 1/8 inch)

Details

Charles Darwin, in 1876, reported data from an experiment that he had conducted on the heights of corn plants. The seeds came from the same parents, but some seeds were produced from self-fertilized parents and some seeds were produced from cross-fertilized parents. Pairs of seeds were planted in pots. Darwin hypothesized that cross-fertilization produced produced more robust and vigorous offspring.

Darwin wrote, "I long doubted whether it was worth while to give the measurements of each separate plant, but have decided to do so, in order that it may be seen that the superiority of the crossed plants over the self-fertilised, does not commonly depend on the presence of two or three extra fine plants on the one side, or of a few very poor plants on the other side. Although several observers

have insisted in general terms on the offspring from intercrossed varieties being superior to either parent-form, no precise measurements have been given;* and I have met with no observations on the effects of crossing and self-fertilising the individuals of the same variety. Moreover, experiments of this kind require so much time—mine having been continued during eleven years—that they are not likely soon to be repeated."

Darwin asked his cousin Francis Galton for help in understanding the data. Galton did not have modern statistical methods to approach the problem and said, "I doubt, after making many tests, whether it is possible to derive useful conclusions from these few observations. We ought to have at least 50 plants in each case, in order to be in a position to deduce fair results".

Later, R. A. Fisher used Darwin's data in a book about design of experiments and showed that a t-test exhibits a significant difference between the two groups.

Source

Darwin, C. R. 1876. *The effects of cross and self fertilisation in the vegetable kingdom*. London: John Murray. Page 16. http://darwin-online.org.uk/converted/published/1881_Worms_F1357/1876_CrossandSelfFertilisation

References

R. A. Fisher, (1935) *The Design of Experiments*, Oliver and Boyd. Page 30.

Examples

```
data(darwin.maize)
dat <- darwin.maize

# Compare self-pollination with cross-pollination
require(lattice)
bwplot(height~type, dat, main="darwin.maize")

## Not run:
require(reshape2)
dm <- melt(dat)
d2 <- dcast(dm, pot+pair~type)
d2$diff <- d2$cross-d2$self
t.test(d2$diff)
## One Sample t-test
## t = 2.148, df = 14, p-value = 0.0497
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 0.003899165 5.229434169

## End(Not run)
```

day.wheat.uniformity *Uniformity trial of wheat*

Description

Uniformity trial of wheat in 1903 in Missouri.

Usage

```
data("day.wheat.uniformity")
```

Format

A data frame with 3090 observations on the following 4 variables.

row row

col col

grain grain weight, grams

straw straw weight, grams

Details

These data are from the Shelbina field of the Missouri Agricultural Experiment Station. The field (plat) was about 1/4 acre in area and apparently uniform throughout. In the fall of 1912, wheat was drilled in rows 8 inches apart, each row 155 feet long. The wheat was harvested in June, in 5-foot segments. The gross weight and the grain weight was measured, the straw weight was calculated by subtraction.

Field length: 100 rows, 8 inches apart = 66.66 feet

Field width: 31 series * 5 feet = 155 feet

Source

James Westbay Day (1916). The relation of size, shape, and number of replications of plats to probable error in field experimentation. Dissertation, University of Missouri. Table 1, page 22. <https://hdl.handle.net/10355/56391>

References

James W. Day (1920). The relation of size, shape, and number of replications of plats to probable error in field experimentation. *Agronomy Journal*, 12, 100-105. <http://doi.org/10.2134/agronj1920.00021962001200030002x>

Examples

```

data(day.wheat.uniformity)
dat <- day.wheat.uniformity

if(require(desplot)){
  desplot(grain~col*row, data=dat,
          flip=TRUE, aspect=(100*8)/(155*12), # true aspect
          main="day.wheat.uniformity - grain yield")
}

# similar to Day table IV
require(lattice)
xyplot(grain~straw, data=dat, main="day.wheat.uniformity", type=c('p','r'))
# cor(dat$grain, dat$straw) # .9498 # Day calculated 0.9416

## Not run:

if(require(desplot)){
  desplot(straw~col*row, data=dat,
          flip=TRUE, aspect=(100*8)/(155*12), # true aspect
          main="day.wheat.uniformity - straw yield")
}

# Day fig 2
coldat <- aggregate(grain~col, dat, sum)
xyplot(grain ~ col, coldat, type='l', ylim=c(2500,6500))
dat$rowgroup <- round((dat$row +1)/3,0)
rowdat <- aggregate(grain~rowgroup, dat, sum)
xyplot(grain ~ rowgroup, rowdat, type='l', ylim=c(2500,6500))

## End(Not run)

```

denis.missing

Multi-environment trial with structured missing values

Description

Grain yield was measured on 5 genotypes in 26 environments. Missing values were non-random, but structured.

Format

env environment, 26 levels
gen genotype factor, 5 levels
yield yield

Source

Denis, J. B. and C P Baril, 1992, Sophisticated models with numerous missing values: The multiplicative interaction model as an example. *Biul. Oceny Odmian*, 24–25, 7–31.

Used with permission of Jean-Baptists Denis.

References

H P Piepho, (1999) Stability analysis using the SAS system, *Agron Journal*, 91, 154–160. <http://doi.org/10.2134/agronj1999.0>

Examples

```
data(denis.missing)
dat <- denis.missing

# view missingness structure
if(require(reshape2)){
  acast(dat, env~gen, value.var='yield')
}

require(lattice)
redblue <- colorRampPalette(c("firebrick", "lightgray", "#375997"))
levelplot(yield ~ gen*env, data=dat,
           col.regions=redblue,
           main="denis.missing - incidence heatmap")

# stability variance (Table 3 in Piepho)
require("nlme")
m1 <- lme(yield ~ -1 + gen, data=dat, random= ~ 1|env,
          weights = varIdent(form= ~ 1|gen),
          na.action=na.omit)
svar <- m1$sigma^2 * c(1, coef(m1$modelStruct$varStruct, unc = FALSE))^2
round(svar, 2)
##          G5   G3   G1   G2
## 39.25 22.95 54.36 12.17 23.77

# m2 <- asreml(yield ~ gen, random = ~ env + at(gen):env, data=dat)
```

denis.ryegrass

Multi-environment trial of perennial ryegrass in France

Description

Plant strength of perennial ryegrass in France for 21 genotypes at 7 locations.

Format

A data frame with 147 observations on the following 3 variables.

gen genotype, 21 levels

loc location, 7 levels

strength average plant strength * 100

Details

INRA conducted a breeding trial in western France with 21 genotypes at 7 locations. The observed data is 'strength' averaged over 7-10 plants per plot and three plots per location (after adjusting for blocking effects). Each plant was scored on a scale 0-9.

The original data had a value of 86.0 for genotype G1 at location L4—this was replaced by an additive estimated value of 361.2 as in Gower and Hand (1996).

Source

Jean-Baptiste Denis and John C. Gower, 1996. Asymptotic confidence regions for biadditive models: interpreting genotype-environment interaction, *Applied Statistics*, 45, 479-493. <http://doi.org/10.2307/2986069>

References

Gower, J.C. and Hand, D.J., 1996. *Biplots*. Chapman and Hall.

Examples

```
data(denis.ryegrass)
dat <- denis.ryegrass

# biplots (without ellipses) similar to Denis figure 1
if(require(gge)){
  m1 <- gge(strength ~ gen*loc, dat, scale=FALSE)
  biplot(m1, main="denis.ryegrass biplot")
}
```

depalluel.sheep

Latin square of four breeds of sheep with four diets

Description

Latin square of four breeds of sheep with four diets

Usage

```
data("depalluel.sheep")
```

Format

A data frame with 32 observations on the following 5 variables.

food diet

animal animal number

breed sheep breed

weight weight, pounds

date months after start

Details

This may be the earliest known Latin Square experiment.

Four sheep from each of four breeds were randomized to four feeds and four slaughter dates.

Sheep that eat roots will eat more than sheep eating corn, but each acre of land produces more roots than corn.

de Palluel said: In short, by adopting the use of roots, instead of corn, for the fattening of all sorts of cattle, the farmers in the neighborhood of the capital will not only gain great profit themselves, but will also very much benefit the public by supplying this great city with resources, and preventing the sudden rise of meat in her markets, which is often considerable.

Source

M. Crette de Palluel (1788). On the advantage and economy of feeding sheep in the house with roots. *Annals of Agriculture*, 14, 133-139. <https://books.google.com/books?id=LXIqAAAAYAAJ&pg=PA133>

References

None

Examples

```
data(depalluel.sheep)
dat <- depalluel.sheep

# Not the best view...weight gain is large in the first month, then slows down
# and the linear line hides this fact
require(lattice)
xyplot(weight ~ date|food, dat, group=animal, type='l', auto.key=list(columns=4),
        xlab="Months since start",
        main="depalluel.sheep")
```

 digby.jointregression *Multi-environment trial of wheat*

Description

Yield of 10 spring wheat varieties for 17 locations in 1976.

Format

A data frame with 134 observations on the following 3 variables.

gen genotype, 10 levels

env environment, 17 levels

yield yield (t/ha)

Details

Yield of 10 spring wheat varieties for 17 locations in 1976.

Used to illustrate modified joint regression.

Source

Digby, P.G.N. (1979). Modified joint regression analysis for incomplete variety x environment data. *Journal of Agricultural Science*, 93, 81-86. <http://doi.org/10.1017/S0021859600086159>

References

Hans-Pieter Piepho, 1997. Analyzing Genotype-Environment Data by Mixed-Models with Multiplicative Terms. *Biometrics*, 53, 761-766. <http://doi.org/10.2307/2533976>

RJOINT procedure in GenStat. <http://www.vsni.co.uk/software/genstat/htmlhelp/server/RJOINT.htm>

Examples

```
data(digby.jointregression)
dat <- digby.jointregression

# Simple gen means, ignoring unbalanced data.
# Matches Digby table 2, Unadjusted Mean
round(tapply(dat$yield, dat$gen, mean),3)

# Two-way model. Matches Digby table 2, Fitting Constants
m00 <- lm(yield ~ 0 + gen + env, dat)
round(coef(m00)[1:10]-2.756078+3.272,3) # Adjust intercept
# genG01 genG02 genG03 genG04 genG05 genG06 genG07 genG08 genG09 genG10
# 3.272 3.268 4.051 3.724 3.641 3.195 3.232 3.268 3.749 3.179

n.gen <- nlevels(dat$gen)
```

```

n.env <- nlevels(dat$env)

# Estimate theta (env eff)
m0 <- lm(yield ~ -1 + env + gen, dat)
thetas <- coef(m0)[1:n.env]
thetas <- thetas-mean(thetas) # center env effects
# Add env effects to the data
dat$theta <- thetas[match(paste("env",dat$env,sep=""), names(thetas))]

# Initialize beta (gen slopes) at 1
betas <- rep(1, n.gen)

done <- FALSE
while(!done){

  betas0 <- betas

  # M1: Fix thetas (env effects), estimate beta (gen slope)
  m1 <- lm(yield ~ -1 + gen + gen:theta, data=dat)
  betas <- coef(m1)[-c(1:n.gen)]
  dat$beta <- betas[match(paste("gen",dat$gen,":theta",sep=""), names(betas))]
  # print(betas)

  # M2: Fix betas (gen slopes), estimate theta (env slope)
  m2 <- lm(yield ~ env:beta + gen -1, data=dat)
  thetas <- coef(m2)[-c(1:n.gen)]
  thetas[is.na(thetas)] <- 0 # Change last coefficient from NA to 0
  dat$theta <- thetas[match(paste("env",dat$env,":beta",sep=""), names(thetas))]
  # print(thetas)

  # Check convergence
  chg <- sum(((betas-betas0)/betas0)^2)
  cat("Relative change in betas",chg,"\n")
  if(chg < .0001) done <- TRUE

}

require(lattice)
xyplot(yield ~ theta|gen, data=dat, xlab="theta (environment effect)",
       main="digby.jointregression - stability plot")

# Digby Table 2, modified joint regression

# Genotype sensitivities (slopes)
round(betas,3) # Match Digby table 2, Modified joint regression sensitivity
# genG01 genG02 genG03 genG04 genG05 genG06 genG07 genG08 genG09 genG10
# 0.953 0.739 1.082 1.024 1.142 0.877 1.089 0.914 1.196 0.947

# Env effects. Match Digby table 3, Modified joint reg
round(thetas,3)+1.164-.515 # Adjust intercept to match
# envE01 envE02 envE03 envE04 envE05 envE06 envE07 envE08 envE09 envE10
# -0.515 -0.578 -0.990 -1.186 1.811 1.696 -1.096 0.046 0.057 0.825
# envE11 envE12 envE13 envE14 envE15 envE16 envE17

```



```
# -0.576 1.568 -0.779 -0.692 0.836 -1.080 0.649

# How to use gnm package for this?
# require(gnm)
# m3 <- gnm(yield ~ gen + Mult(gen,env), data=dat) # slopes negated
```

diggle.cow

Bodyweight of cows in a 2-by-2 factorial experiment

Description

Bodyweight of cows in a 2-by-2 factorial experiment.

Format

A data frame with 598 observations on the following 5 variables.

animal Animal factor, 26 levels
iron Factor with levels Iron, NoIron
infect Factor levels Infected, NonInfected
weight Weight in (rounded to nearest 5) kilograms
day Days after birth

Details

Diggle et al., 1994, pp. 100-101, consider an experiment that studied how iron dosing (none/standard) and micro-organism (infected or non-infected) influence the weight of cows.

Twenty-eight cows were allocated in a 2-by-2 factorial design with these factors. Some calves were inoculated with tuberculosis at six weeks of age. At six months, some calves were maintained on supplemental iron diet for a further 27 months.

The weight of each animal was measured at 23 times, unequally spaced. One cow died during the study and data for another cow was removed.

Source

Diggle, P. J., Liang, K.-Y., & Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Page 100-101. Retrieved Oct 2011 from <http://www.maths.lancs.ac.uk/~diggle/lda/Datasets/>

References

Lepper, AWD and Lewis, VM, 1989. Effects of altered dietary iron intake in Mycobacterium paratuberculosis-infected dairy cattle: sequential observations on growth, iron and copper metabolism and development of paratuberculosis. *Research in veterinary science*, 46, 289–296.

Arunas P. Verbyla and Brian R. Cullis and Michael G. Kenward and Sue J. Welham, (1999), The analysis of designed experiments and longitudinal data by using smoothing splines. *Appl. Statist.*, 48, 269–311.

SAS/STAT(R) 9.2 User's Guide, Second Edition. <http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/v>

Examples

```

data(diggle.cow)
dat <- diggle.cow

# Figure 1 of Verbyla 1999
require(lattice)
if(require(latticeExtra)){
  useOuterStrips(xyplot(weight ~ day|iron*infect, dat, group=animal,
                        type='b', cex=.5,
                        main="diggle.cow"))
}

# Scaling
dat <- transform(dat, time = (day-122)/10)

# -----

## Not run:
# asreml3
require(asreml)

# Smooth for each animal. No treatment effects. Similar to SAS Output 38.6.9

m1 <- asreml(weight ~ 1 + lin(time) + animal + animal:lin(time), data=dat,
             random = ~ animal:spl(time))
p1 <- predict(m1, data=dat, classify="animal:time",
             predictpoints=list(time=seq(0,65.9, length=50)))
p1 <- p1$pred$pval
p1 <- merge(dat, p1, all=TRUE) # to get iron/infect merged in
foo1 <- xyplot(weight ~ day|iron*infect, dat, group=animal)
foo2 <- xyplot(predicted.value ~ day|iron*infect, p1, type='l', group=animal)
print(foo1+foo2)

## End(Not run)

# -----

## Not run:
## require(asreml4)
## require(latticeExtra)

## # Smooth for each animal. No treatment effects. Similar to SAS Output 38.6.9

## m1 <- asreml(weight ~ 1 + lin(time) + animal + animal:lin(time), data=dat,
##             random = ~ animal:spl(time))
## p1 <- predict(m1, data=dat, classify="animal:time",
##             design.points=list(time=seq(0,65.9, length=50)))
## p1 <- p1$pvals
## p1 <- merge(dat, p1, all=TRUE) # to get iron/infect merged in
## foo1 <- xyplot(weight ~ day|iron*infect, dat, group=animal)

```

```
## foo2 <- xyplot(predicted.value ~ day|iron*infect, p1, type='l', group=animal)
## print(foo1+foo2)

## End(Not run)
```

draper.safflower.uniformity

Uniformity trial of safflower

Description

Uniformity trial of safflower in Arizona in 1958.

Usage

```
data("draper.safflower.uniformity")
```

Format

A data frame with 640 observations on the following 4 variables.

expt experiment

row row

col column

yield yield per plot (gramsm)

Details

Experiments were conducted at the Agricultural Experiment Station Farm at Eloy, Arizona. The crop was harvested in July 1958.

The crop was planted in two rows 12 inches apart on vegetable beds 40 inches center to center.

In each test, the end ranges and one row of plots on one side were next to alleys, and those plots gave estimates of border effects.

Experiment E4 (four foot test).

Sandy streaks were present in the field. Average yield was 1487 lb/ac. A diagonal fertility gradient was in this field. Widening the plot was equally effective as lengthening the plot to reduce variability. The optimum plot size was 1 bed wide, 24 feet long. Considering economic costs, the optimum size was 1 bed, 12 feet long.

Field length: 18 ranges * 4 feet = 72 feet

Field width: 16 beds * 3.33 = 53 feet

Experiment E5 (five foot test)

Average yield 2517 lb/ac, typical for this crop. Combining plots lengthwise was more effective than widening the plots, in order to reduce variability. The optimum plot size was 1 bed wide, 25 feet long. Considering economic costs, the optimum size was 1 bed, 18 feet long.

Field length: 18 ranges * 5 feet = 90 feet.

Field width: 14 beds * 3.33 = 46.6 feet.

Source

Arlen D. Draper. (1959). Optimum plot size and shape for safflower yield tests. Dissertation. University of Arizona. <http://hdl.handle.net/10150/319371>

References

None

Examples

```

data("draper.safflower.uniformity")
dat4 <- subset(draper.safflower.uniformity, expt=="E4")
dat5 <- subset(draper.safflower.uniformity, expt=="E5")

if(require(desplot)){
  desplot(yield~col*row, dat4,
          flip=TRUE, tick=TRUE, aspect=72/53, # true aspect
          main="draper.safflower.uniformity (four foot)")
}

if(require(desplot)){
  desplot(yield~col*row, dat5,
          flip=TRUE, tick=TRUE, aspect=90/46, # true aspect
          main="draper.safflower.uniformity (five foot)")
}

## Not run:
require(agricolae)
require(reshape2)
# Remove border plots at either end and right side
dat4 <- subset(dat4, row>1 & row<20)
dat4 <- subset(dat4, col<17)
dat5 <- subset(dat5, row>1 & row<20)
dat5 <- subset(dat5, col<15)

s4 <- index.smith(acast(dat4, row~col, value.var='yield'),
                 main="draper.safflower.uniformity (four foot)",
                 col="red")$uni
s4 # match Draper table 2, p 22

## s5 <- index.smith(acast(dat5, row~col, value.var='yield'),
##                   main="draper.safflower.uniformity (five foot)",
##                   col="red")$uni
## s5 # match Draper table 1, p 21

## End(Not run)

```

durban.competition *Sugar beet yields with competition effects*

Description

Sugar beet yields with competition effects

Format

A data frame with 114 observations on the following 5 variables.

gen Genotype factor, 36 levels plus Border

col Column

block Row/Block

wheel Position relative to wheel tracks

yield Root yields, kg/plot

Details

This sugar-beet trial was conducted in 1979.

Single-row plots, 12 m long, 0.5 m between rows. Each block is made up of all 36 genotypes laid out side by side. Guard/border plots are at each end. Root yields were collected.

Wheel tracks are located between columns 1 and 2, and between columns 5 and 6, for each set of six plots. Each genotype was randomly allocated once to each pair of plots (1,6), (2,5), (3,4) across the three reps. Wheel effect were not significant in `_this_` trial.

Field length: 3 blocks * 12m + 2*0.5m spacing = 37m

Field width: 18m + 1m guard rows = 19m

Source

Durban, M., Currie, I. and R. Kempton, 2001. Adjusting for fertility and competition in variety trials. *J. of Agricultural Science*, 136, 129–140.

Retrieved from <http://www.ma.hw.ac.uk/~iain/research/JAgSciData/data/Trial1.dat>

Used with permission of Iain Currie.

Examples

```
data(durban.competition)
dat <- durban.competition

# Check that genotypes were balanced across wheel tracks.
with(dat, table(gen,wheel))

if(require(desplot)){
```

```

desplot(yield ~ col*block, dat,
        out1=block, text=gen, col=wheel, aspect=37/19, # true aspect
        main="durban.competition")
}

# Calculate residual after removing block/genotype effects
m1 <- lm(yield ~ gen + block, data=dat)
dat$res <- resid(m1)

## desplot(res ~ col*block, dat, out1=block, text=gen, col=wheel,
##         main="durban.competition - residuals")

# Calculate mean of neighboring plots
dat$comp <- NA
dat$comp[3:36] <- ( dat$yield[2:35] + dat$yield[4:37] ) / 2
dat$comp[41:74] <- ( dat$yield[40:73] + dat$yield[42:75] ) / 2
dat$comp[79:112] <- ( dat$yield[78:111] + dat$yield[80:113] ) / 2

# Demonstrate the competition effect
# Competitor plots have low/high yield -> residuals are negative/positive
require(lattice)
xyplot(res~comp, dat, type=c('p','r'), main="durban.competition",
       xlab="Average yield of neighboring plots", ylab="Residual")

```

durban.rowcol

Row column design of a spring barley trial with many varieties

Description

Row column design of a spring barley trial with many varieties

Format

A data frame with 544 observations on the following 5 variables.

row row

bed bed (column)

rep rep, 2 levels

gen genotype, 272 levels

yield yield, tonnes/ha

Details

Spring barley variety trial of 272 entries (260 new varieties, 12 control). Grown at the Scottish Crop Research Institute in 1998. Row-column design with 2 reps, 16 rows (north/south) by 34 beds (east/west). The land sloped downward from row 16 to row 1. Plot yields were converted to tonnes per hectare.

Plot dimensions are not given.

Source

Durban, Maria and Hackett, Christine and McNicol, James and Newton, Adrian and Thomas, William and Currie, Iain. 2003. The practical use of semiparametric models in field trials, *Journal of Agric Biological and Envir Stats*, 8, 48-66. <http://doi.org/10.1198/1085711031265>

Retrieved from: <ftp://ftp.bioss.sari.ac.uk/pub/maria>

Used with permission of Maria Durban.

Examples

```
data(durban.rowcol)
dat <- durban.rowcol

if(require(desplot)){
  desplot(yield~bed*row, dat,
          out1=rep, num=gen, # aspect unknown
          main="durban.rowcol")
}

# Durban 2003 Figure 1
m10 <- lm(yield~gen, data=dat)
dat$resid <- m10$resid
## require(lattice)
## xyplot(resid~row, dat, type=c('p','smooth'), main="durban.rowcol")
## xyplot(resid~bed, dat, type=c('p','smooth'), main="durban.rowcol")

# Figure 3
coplot(resid~bed|factor(row), data=dat, cex=.5,
       panel=function(x,y,...) panel.smooth(x,y,span=.75,...))
title("durban.rowcol")

# -----

## Not run:
# Figure 5 - field trend
# note, Durban used gam package like this
# m1lo <- gam(yield ~ gen + lo(row, span=10/16) + lo(bed, span=9/34), data=dat)
require(mgcv)
m1lo <- gam(yield ~ gen + s(row) + s(bed, k=5), data=dat)
new1 <- expand.grid(row=unique(dat$row),bed=unique(dat$bed))
new1 <- cbind(new1, gen="G001")
p1lo <- predict(m1lo, newdata=new1)
wireframe(p1lo~row+bed, new1, aspect=c(1,.5), main="Field trend") # Figure 5

## End(Not run)

# -----

## Not run:
# Figure 7 - variograms
```

```

# asreml3
require(asreml)
dat <- transform(dat, rowf=factor(row), bedf=factor(bed))
dat <- dat[order(dat$rowf, dat$bedf),]

m1a1 <- asreml(yield~gen + lin(rowf) + lin(bedf), data=dat,
              random=~spl(rowf) + spl(bedf) + units,
              family=asreml.gaussian(dispersion=1))
m1a2 <- asreml(yield~gen + lin(rowf) + lin(bedf), data=dat,
              random=~spl(rowf) + spl(bedf) + units, rcov=~ar1(rowf):ar1(bedf))
m1a3 <- asreml(yield~gen, data=dat, random=~units, rcov=~ar1(rowf):ar1(bedf))

v7a <- asreml.variogram(x=dat$bedf, y=dat$rowf, z=m1a3$residuals)
wireframe(gamma ~ x*y, v7a, aspect=c(1,.5)) # Fig 7a

v7b <- asreml.variogram(x=dat$bedf, y=dat$rowf, z=m1a2$residuals)
wireframe(gamma ~ x*y, v7b, aspect=c(1,.5)) # Fig 7b

v7c <- asreml.variogram(x=dat$bedf, y=dat$rowf, z=m1lo$residuals)
wireframe(gamma ~ x*y, v7c, aspect=c(1,.5)) # Fig 7c

## End(Not run)

# -----

## Not run:
## require(asreml4)
## dat <- transform(dat, rowf=factor(row), bedf=factor(bed))
## dat <- dat[order(dat$rowf, dat$bedf),]

## m1a1 <- asreml(yield~gen + lin(rowf) + lin(bedf), data=dat,
##              random=~spl(rowf) + spl(bedf) + units,
##              family=asr_gaussian(dispersion=1))
## m1a2 <- asreml(yield~gen + lin(rowf) + lin(bedf), data=dat,
##              random=~spl(rowf) + spl(bedf) + units,
##              resid = ~ar1(rowf):ar1(bedf))
## m1a2 <- update(m1a2)
## m1a3 <- asreml(yield~gen, data=dat, random=~units,
##              resid = ~ar1(rowf):ar1(bedf))

## # Figure 7
## v7a <- asr_varioGram(x=dat$bedf, y=dat$rowf, z=m1a3$residuals)
## wireframe(gamma ~ x*y, v7a, aspect=c(1,.5)) # Fig 7a

## v7b <- asr_varioGram(x=dat$bedf, y=dat$rowf, z=m1a2$residuals)
## wireframe(gamma ~ x*y, v7b, aspect=c(1,.5)) # Fig 7b

## v7c <- asr_varioGram(x=dat$bedf, y=dat$rowf, z=m1lo$residuals)
## wireframe(gamma ~ x*y, v7c, aspect=c(1,.5)) # Fig 7c

## End(Not run)

```

durban.splitplot *Split-plot barley variety trial with fungicide treatments*

Description

Split-plot barley variety trial with fungicide treatments.

Format

A data frame with 560 observations on the following 6 variables.

yield yield, tonnes/ha
block block, 4 levels
gen genotype, 70 levels
fung fungicide, 2 levels
row row
bed bed (column)

Details

Grown in 1995-1996 at the Scottish Crop Research Institute. Split-plot design with 4 blocks, 2 whole-plot fungicide treatments, and 70 barley varieties or variety mixes. Total area was 10 rows (north/south) by 56 beds (east/west).

Source

Durban, Maria and Hackett, Christine and McNicol, James and Newton, Adrian and Thomas, William and Currie, Iain. 2003. The practical use of semiparametric models in field trials, *Journal of Agric Biological and Envir Stats*, 8, 48-66. <http://doi.org/10.1198/1085711031265>.

Retrieved from: <ftp://ftp.bioss.sari.ac.uk/pub/maria>

Used with permission of Maria Durban.

Examples

```
data(durban.splitplot)
dat <- durban.splitplot

if(require(desplot)){
  desplot(yield~bed*row, dat,
          out1=block, out2=fung, num=gen, # aspect unknown
          main="durban.splitplot")
}

# Durban 2003, Figure 2
m20 <- lm(yield~gen*fung, data=dat)
```

```

dat$resid <- m20$resid
## require(lattice)
## xyplot(resid~row, dat, type=c('p','smooth'), main="durban.splitplot")
## xyplot(resid~bed, dat, type=c('p','smooth'), main="durban.splitplot")

# Figure 4 doesn't quite match due to different break points
coplot(resid~bed|row, data=dat, number=8, cex=.5,
        panel=function(x,y,...) panel.smooth(x,y,span=.4,...))
title("durban.splitplot")

# -----

## Not run:
# Figure 6 - field trend
# note, Durban used gam package like this
# m2lo <- gam(yield ~ gen*fung + lo(row, bed, span=.082), data=dat)
require(mgcv)
m2lo <- gam(yield ~ gen*fung + s(row, bed,k=45), data=dat)
new2 <- expand.grid(row=unique(dat$row), bed=unique(dat$bed))
new2 <- cbind(new2, gen="G01", fung="F1")
p2lo <- predict(m2lo, newdata=new2)
wireframe(p2lo~row+bed, new2, aspect=c(1,.5), main="Field trend")

## End(Not run)

# -----

## Not run:
# asreml3
require(asreml)
# Table 5, variance components. Table 6, F tests

dat <- transform(dat, rowf=factor(row), bedf=factor(bed))
dat <- dat[order(dat$rowf, dat$bedf),]
m2a2 <- asreml(yield ~ gen*fung, random=~block/fung+units, data=dat,
              rcov=~ar1v(rowf):ar1(bedf))
m2a2 <- update(m2a2)

require(lucid)
vc(m2a2)
##           effect component std.error z.ratio constr
##   block!block.var 0.0000001      NA      NA bound
## block:fung!block.var 0.01207   0.01513    0.8   pos
##   units!units.var 0.02463   0.002465   10    pos
##           R!variance 1      NA      NA   fix
##           R!rowf.cor 0.8836   0.03647   24   uncon
##           R!rowf.var 0.1262   0.04432   2.8   pos
##           R!bedf.cor 0.9202   0.02847   32   uncon

anova(m2a2)

## End(Not run)

```

```
# -----
## Not run:
## require(asreml4)
## # Table 5, variance components. Table 6, F tests

## dat <- transform(dat, rowf=factor(row), bedf=factor(bed))
## dat <- dat[order(dat$rowf, dat$bedf),]
## m2a2 <- asreml(yield ~ gen*fung, random=~block/fung+units, data=dat,
##               resid =~ar1v(rowf):ar1(bedf))
## m2a2 <- update(m2a2)

## require(lucid)
## vc(m2a2)
## ##          effect component std.error z.ratio bound
## ##          block      0         NA      NA      B  NA
## ##          block:fung 0.01206 0.01512   0.8      P  0
## ##          units     0.02463 0.002465 10       P  0
## ##          rowf:bedf(R) 1         NA      NA      F  0
## ##          rowf:bedf!rowf!cor 0.8836 0.03646 24       U  0
## ##          rowf:bedf!rowf!var 0.1261 0.04434 2.8      P  0
## ##          rowf:bedf!bedf!cor 0.9202 0.02846 32       U  0

## wald(m2a2)

## End(Not run)
```

eden.nonnormal

Height of barley plants in a study of non-normal data

Description

Height of barley plants in a study of non-normal data.

Usage

```
data("eden.nonnormal")
```

Format

A data frame with 256 observations on the following 3 variables.

pos position within block

block block (numeric)

height height of wheat plant

Details

This data was used in a very early example of a permutation test.

Eden & Yates used this data to consider the impact of non-normal data on the validity of a hypothesis test that assumes normality. They concluded that the skew data did not negatively affect the analysis of variance.

Grown at Rothamsted. Eight blocks of Yeoman II wheat. Sampling of the blocks was quarter-meter rows, four times in each row. Rows were selected at random. Position within the rows was partly controlled to make use of the whole length of the block. Plants at both ends of the sub-unit were measured. Shoot height is measured from ground level to the auricle of the last expanded leaf.

Source

T. Eden, F. Yates (1933). On the validity of Fisher's z test when applied to an actual example of non-normal data. *Journal of Agric Science*, 23, 6-17. <https://doi.org/10.1017/S0021859600052862>

References

Kenneth J. Berry, Paul W. Mielke, Jr., Janis E. Johnston *Permutation Statistical Methods: An Integrated Approach*.

Examples

```
data(eden.nonnormal)
dat <- eden.nonnormal
mean(dat$height) # 55.23 matches Eden table 1

# Eden figure 2
if(require(dplyr) & require(lattice)){
  # Blocks had different means, so subtract block mean from each datum
  dat <- group_by(dat, block)
  dat <- mutate(dat, blkmn=mean(height))
  dat <- transform(dat, dev=height-blkmn)

  histogram( ~ dev, data=dat, breaks=seq(from=-40, to=30, by=2.5),
             xlab="Deviations from block means",
             main="eden.nonnormal - heights skewed left")
}

## Not run:

# calculate skewness, permutation

library(dplyr)
library(lattice)
library(latticeExtra)

# Eden table 1
# anova(aov(height ~ factor(block), data=dat))
```

```

# Eden table 2,3. Note, this may be a different definition of skewness
# than is commonly used today (e.g. e1071::skewness).
skew <- function(x){
  n <- length(x)
  x <- x - mean(x)
  s1 = sum(x)
  s2 = sum(x^2)
  s3 = sum(x^3)
  k3=n/((n-1)*(n-2)) * s3 -3/n*s2*s1 + 2/n^2 * s1^3
  return(k3)
}
# Negative values indicate data are skewed left
dat <- group_by(dat, block)
summarize(dat, s1=sum(height),s2=sum(height^2), mean2=var(height), k3=skew(height))
##   block    s1      s2    mean2     k3
##   <int> <dbl> <dbl>   <dbl>   <dbl>
## 1     1 1682.0 95929.5 242.56048 -1268.5210
## 2     2 1858.0 111661.5 121.97984 -1751.9919
## 3     3 1809.5 108966.8 214.36064 -3172.5284
## 4     4 1912.0 121748.5 242.14516 -2548.2194
## 5     5 1722.0 99026.5 205.20565 -559.0629
## 6     6 1339.0 63077.0 227.36190 -801.2740
## 7     7 1963.0 123052.5 84.99093 -713.2595
## 8     8 1854.0 112366.0 159.67339 -1061.9919

# Another way to view skewness with qq plot. Panel 3 most skewed.
qqmath( ~ dev|factor(block), data=dat,
        as.table=TRUE,
        ylab="Deviations from block means",
        panel = function(x, ...) {
          panel.qqmathline(x, ...)
          panel.qqmath(x, ...)
        })

# Now, permutation test.
# Eden: "By a process of amalgamation the eight sets of 32 observations were
# reduced to eight sets of four and the data treated as a potential
# layout for a 32-plot trial".
dat2 <- transform(dat, grp = rep(1:4, each=8))
dat2 <- aggregate(height ~ grp+block, dat2, sum)
dat2$trt <- rep(letters[1:4], 8)
dat2$block <- factor(dat2$block)

# Treatments were assigned at random 1000 times
set.seed(54323)
fobs <- rep(NA, 1000)
for(i in 1:1000){
  # randomize treatments within each block
  # trick from http://stackoverflow.com/questions/25085537
  dat2$trt <- with(dat2, ave(trt, block, FUN = sample))
  fobs[i] <- anova(aov(height ~ block + trt, dat2))["trt", "F value"]
}

```

```

# F distribution with 3,21 deg freedom
# Similar to Eden's figure 4, but on a different horizontal scale
xval <- seq(from=0,to=max(fobs), length=50)
yval <- df(xval, df1 = 3, df2 = 21)
# Re-scale, 10 = max of histogram, 0.7 = max of density
histogram( ~ fobs, breaks=xval,
           xlab="F value",
           main="Observed (histogram) & theoretical (line) F values") +
  xyplot((10/.7)* yval ~ xval, type="l", lwd=2)

## End(Not run)

```

eden.potato

Potato yields in response to potash and nitrogen fertilizer

Description

Potato yields in response to potash and nitrogen fertilizer. Data from Fisher's 1929 paper *Studies in Crop Variation 6*. A different design was used each year.

Format

A data frame with 225 observations on the following 9 variables.

year year/type factor
yield yield, pounds per plot
block block
row row
col column
trt treatment factor
nitro nitrogen fertilizer, cwt/acre
potash potash fertilizer, cwt/acre
ptype potash type

Details

The data is of interest to show the gradual development of experimental designs in agriculture.

In 1925/1926 the potato variety was Kerr's Pink. In 1927 Arran Comrade.

In the 1925a/1926a qualitative experiments, the treatments are O=None, S=Sulfate, M=Muriate, P=Potash manure salts. The design was a Latin Square.

The 1925/1926b/1927 experiments were RCB designs with treatment codes defining the amount and type of fertilizer used. Note: the 't' treatment was not defined in the original paper.

Source

T Eden and R A Fisher, 1929. Studies in Crop Variation. VI. Experiments on the response of the potato to potash and nitrogen. *Journal of Agricultural Science*, 19: 201-213.

References

McCullagh, P. and Clifford, D., (2006). Evidence for conformal invariance of crop yields, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 462, 2119–2143.

Examples

```
data(eden.potato)
dat <- eden.potato

# 1925 qualitative
d5a <- subset(dat, year=='1925a')
if(require(desplot)){
  desplot(trt~col*row, d5a,
          text=yield, cex=1, shorten='no', # aspect unknown
          main="eden.potato: 1925 qualitative")
}
anova(m5a <- aov(yield~trt+factor(row)+factor(col), d5a)) # table 2

# 1926 qualitative
d6a <- subset(dat, year=='1926a')
if(require(desplot)){
  desplot(trt~col*row, d6a,
          text=yield, cex=1, shorten='no', # aspect unknown
          main="eden.potato: 1926 qualitative")
}
anova(m6a <- aov(yield~trt+factor(row)+factor(col), d6a)) # table 4

# 1925 quantitative
d5 <- subset(dat, year=='1925b')
if(require(desplot)){
  desplot(yield ~ col*row, d5,
          out1=block, text=trt, cex=1, # aspect unknown
          main="eden.potato: 1925 quantitative")
}
# Trt 't' not defined, seems to be the same as 'a'
require(lattice)
dotplot(trt~yield|block, d5,
        # aspect unknown
        main="eden.potato: 1925 quantitative")
anova(m5 <- aov(yield~trt+block, d5)) # table 6

# 1926 quantitative
d6 <- subset(dat, year=='1926b')
if(require(desplot)){
  desplot(yield ~ col*row, d6,
          out1=block, text=trt, cex=1, # aspect unknown
```

```

      main="eden.potato: 1926 quantitative")
}
anova(m6 <- aov(yield~trt+block, d6)) # table 7

# 1927 qualitative + quantitative
d7 <- droplevels(subset(dat, year==1927))
if(require(desplot)){
  desplot(yield ~ col*row, d7,
          out1=block, text=trt, cex=1, col=ptype, # aspect unknown
          main="eden.potato: 1927 qualitative + quantitative")
}

# Table 8. Anova, mean yield tons / acre
anova(m7 <- aov(yield~trt+block+ptype + ptype:potash, d7))
if(require(reshape2)){
  me7 <- melt(d7, measure.vars='yield')
  acast(me7, potash~nitro, fun=mean) * 40/2240 # English ton = 2240 lbs
  acast(me7, potash~ptype, fun=mean) * 40/2240
}

```

 engelstad.nitro

Multi-environment trial of corn with nitrogen fertilizer

Description

Corn yield response to nitrogen fertilizer for a single variety of corn at two locations over five years

Format

A data frame with 60 observations on the following 4 variables.

loc location, 2 levels

year year, 1962-1966

nitro nitrogen fertilizer kg/ha

yield yield, quintals/ha

Details

Corn yield response to nitrogen fertilizer for a single variety of corn at two locations in Tennessee over five years. The yield data is the mean of 9 replicates. The original paper fits quadratic curves to the data. Schabenberger and Pierce fit multiple models including linear plateau. The example below fits a quadratic plateau for one year/loc. In the original paper, the 1965 and 1966 data for the Knoxville location was not used as it appeared that the response due to nitrogen was minimal in 1965 and nonexistent in 1966. The economic optimum can be found by setting the tangent equal to the ratio of (fertilizer price)/(grain price).

Source

Engelstad, OP and Parks, WL. 1971. Variability in Optimum N Rates for Corn. *Agronomy Journal*, 63, 21–23.

References

Schabenberger, O. and Pierce, F.J., 2002. *Contemporary statistical models for the plant and soil sciences*, CRC. Page 254-259.

Examples

```
data(engelstad.nitro)
dat <- engelstad.nitro

if(require(latticeExtra)) {
  useOuterStrips(xyplot(yield ~ nitro | factor(year)*loc, dat,
    main="engelstad.nitro"))
}

# Fit a quadratic plateau model to one year/loc
j62 <- droplevels(subset(dat, loc=="Jackson" & year==1962))
# ymax is maximum yield, M is the change point, k affects curvature
m1 <- nls(yield ~ ymax*(nitro > M) +
  (ymax - (k/2) * (M-nitro)^2) * (nitro < M),
  data= j62,
  start=list(ymax=80, M=150, k=.01))

# Plot the raw data and model
newdat <- data.frame(nitro=seq(0,max(dat$nitro)))
p1 <- predict(m1, new=newdat)
plot(yield ~ nitro, j62)
lines(p1 ~ newdat$nitro, col="blue")
title("engelstad.nitro: quadratic plateau at Jackson 1962")

# Optimum nitro level ignoring prices = 225
coef(m1)['M']

# Optimum nitro level using $0.11 for N cost, $1.15 for grain price = 140
# Set the first derivative equal to N/corn price, k(M-nitro)=.11/1.15
coef(m1)['M']-(.11/1.15)/coef(m1)['k']
```

Description

Yield of 13 hybrids, grown in 10 locations across 2 years. Conducted in Yunnan, China.

Format

A data frame with 260 observations on the following 5 variables.

gen genotype
 maturity maturity, days
 year year
 loc location
 yield yield, Mg/ha

Details

Data are the mean of 3 reps.

These data were used to conduct a stability analysis of yield.

Source

Fan, X.M. and Kang, M.S. and Chen, H. and Zhang, Y. and Tan, J. and Xu, C. (2007). Yield stability of maize hybrids evaluated in multi-environment trials in Yunnan, China. *Agronomy Journal*, 99, 220-228. <https://doi.org/10.2134/agronj2006.0144>

Used with permission of Manjit Kang.

Examples

```
data(fan.stability)
dat <- fan.stability

dat$env <- factor(paste(dat$loc, dat$year, sep=""))
require(lattice)
dotplot(gen~yield|env, dat, main="fan.stability")

if(require(reshape2) & require(agricolae)){
  dm <- acast(dat, gen~env, value.var='yield')
  # Use 0.464 as pooled error from ANOVA. Calculate yield mean/stability.
  stability.par(dm, rep=3, MSerror=0.464) # Table 5 of Fan et al.
}
```

federer.diagcheck

Wheat experiment with diagonal checks

Description

Wheat experiment augmented with two check varieties in diagonal strips.

Format

A data frame with 180 observations on the following 4 variables.

```
row row
col column
gen genotype, 120 levels
yield yield
```

Details

This experiment was conducted by Matthew Reynolds, CIMMYT. There are 180 plots in the field, 60 for the diagonal checks (G121 and G122) and 120 for new varieties.

Federer used this data in multiple papers to illustrate the use of orthogonal polynomials to model field trends that are not related to the genetic effects.

Note: Federer and Wolfinger (2003) provide a SAS program for analysis of this data. However, when the SAS program is used to analyze this data, the results do not match the results given in Federer (1998) nor Federer and Wolfinger (2003). The differences are slight, which suggests a typographical error in the presentation of the data.

The R code below provides results that are consistent with the SAS code of Federer & Wolfinger (2003) when both are applied to this version of the data.

Plot dimensions are not given.

Source

Federer, Walter T. 1998. Recovery of interblock, intergradient, and intervariety information in incomplete block and lattice rectangle design experiments. *Biometrics*, 54, 471–481. <http://doi.org/10.2307/3109756>

References

Walter T Federer and Russell D Wolfinger, 2003. Augmented Row-Column Design and Trend Analysis, chapter 28 of *Handbook of Formulas and Software for Plant Geneticists and Breeders*, Haworth Press.

Examples

```
data(federer.diagcheck)
dat <- federer.diagcheck

# Show the layout as in Federer 1998.
dat$check <- ifelse(dat$gen == "G121" | dat$gen=="G122", "C", "N")
if(require(desplot)){
  desplot(yield ~ col*row, dat,
          text=gen, show.key=FALSE, # aspect unknown
          shorten='no', col=check, cex=.8, col.text=c("yellow","gray"),
          main="federer.diagcheck")
}
```

```

# Now reproduce the analysis of Federer 2003.

# Only to match SAS results
dat$row <- 16 - dat$row
dat <- dat[order(dat$col, dat$row), ]

# Add row / column polynomials to the data.
# The scaling factors sqrt() are arbitrary, but used to match SAS
nr <- length(unique(dat$row))
nc <- length(unique(dat$col))
rpoly <- poly(dat$row, degree=10) * sqrt(nc)
cpoly <- poly(dat$col, degree=10) * sqrt(nr)
dat <- transform(dat,
                 c1 = cpoly[,1], c2 = cpoly[,2], c3 = cpoly[,3],
                 c4 = cpoly[,4], c6 = cpoly[,6], c8 = cpoly[,8],
                 r1 = rpoly[,1], r2 = rpoly[,2], r3 = rpoly[,3],
                 r4 = rpoly[,4], r8 = rpoly[,8], r10 = rpoly[,10])
dat$trtn <- ifelse(dat$gen == "G121" | dat$gen=="G122", dat$gen, "G999")
dat$new <- ifelse(dat$gen == "G121" | dat$gen=="G122", "N", "Y")
dat <- transform(dat, trtn=factor(trtn), new=factor(new))

m1 <- lm(yield ~ c1 + c2 + c3 + c4 + c6 + c8
        + r1 + r2 + r4 + r8 + r10
        + c1:r1 + c2:r1 + c3:r1 + gen, data = dat)
# To get Type III SS use the following
if(require(car)) {
  Anova(m1, type=3) # Matches PROC GLM output
  ##           Sum Sq Df F value Pr(>F)
  ## (Intercept) 538948 1 159.5804 3.103e-16 ***
  ## c1          13781 1 4.0806 0.0494940 *
  ## c2          51102 1 15.1312 0.0003354 ***
  ## c3          45735 1 13.5419 0.0006332 ***
  ## c4          24670 1 7.3048 0.0097349 **
  ## c6          33002 1 9.7719 0.0031359 **
}

# lmer
if(require(lme4) & require(lucid)){

  dat$one <- factor(rep(1, nrow(dat)))

  # lmer with bobyqa (default)
  m2b <- lmer(yield ~ trtn + (0 + r1 + r2 + r4 + r8 + r10 +
    c1 + c2 + c3 + c4 + c6 + c8 + r1:c1 + r1:c2 + r1:c3 || one) +
    (1|new:gen),
    data = dat, control=lmerControl(check.nlev.gtr.1="ignore"))
  vc(m2b)
  ##      grp      var1 var2      vcov      sdcov
  ## new.gen (Intercept) <NA> 2869      53.57
  ## one      r1:c3 <NA> 5532      74.37
  ## one.1    r1:c2 <NA> 58230     241.3
  ## one.2    r1:c1 <NA> 128000     357.8

```

```

##   one.3      c8 <NA> 6456   80.35
##   one.4      c6 <NA> 1400   37.41
##   one.5      c4 <NA> 1792   42.33
##   one.6      c3 <NA> 2549   50.49
##   one.7      c2 <NA> 5942   77.08
##   one.8      c1 <NA>    0     0
##   one.9      r10 <NA> 1133   33.66
##   one.10     r8 <NA> 1355   36.81
##   one.11     r4 <NA> 2269   47.63
##   one.12     r2 <NA> 241.8  15.55
##   one.13     r1 <NA> 9200   95.92
## Residual    <NA> <NA> 4412   66.42

## lmer with Nelder_Mead gives 'wrong' results
## m2n <- lmer(yield ~ trtn + (0 + r1 + r2 + r4 + r8 + r10 +
##           c1 + c2 + c3 + c4 + c6 + c8 + r1:c1 + r1:c2 + r1:c3 || one) +
##           (1|new:gen)
##           , data = dat,
##           control=lmerControl(optimizer="Nelder_Mead",check.nlev.gtr.1="ignore"))
## vc(m2n)
##   groups      name variance  stddev
## new.gen (Intercept) 3228    56.82
## one     r1:c3      7688    87.68
## one.1   r1:c2     69750   264.1
## one.2   r1:c1    107400   327.8
## one.3   c8        6787    82.38
## one.4   c6        1636    40.45
## one.5   c4       12270   110.8
## one.6   c3        2686    51.83
## one.7   c2        7645    87.43
## one.8   c1         0     0.0351
## one.9   r10       1976    44.45
## one.10  r8        1241    35.23
## one.11  r4        2811    53.02
## one.12  r2        928.2   30.47
## one.13  r1       10360   101.8
## Residual      4127    64.24

}

# -----

## Not run:
# asreml3
require(asreml)
require(lucid)
m3 <- asreml(yield ~ -1 + trtn, data=dat,
            random = ~ r1 + r2 + r4 + r8 + r10 +
                    c1 + c2 + c3 + c4 + c6 + c8 + r1:c1 + r1:c2 + r1:c3 + new:gen)
coef(m3)
# REML cultivar means. Very similar to Federer table 2.
rev(sort(round(coef(m3)$fixed[3] + coef(m3)$random[137:256,],0)))
## gen_G060 gen_G021 gen_G011 gen_G099 gen_G002

```

```

##      974      949      945      944      942
## gen_G118 gen_G058 gen_G035 gen_G111 gen_G120
##      938      937      937      933      932
## gen_G046 gen_G061 gen_G082 gen_G038 gen_G090
##      932      931      927      927      926

vc(m3)
##          effect component std.error z.ratio constr
##      r1!r1.var   9201      13720   0.67   pos
##      r2!r2.var    241.7      1059    0.23   pos
##      r4!r4.var   2269      3915    0.58   pos
##      r8!r8.var   1355      2627    0.52   pos
##      r10!r10.var 1133      2312    0.49   pos
##      c1!c1.var     0.01         0     4.8   bound
##      c2!c2.var   5942      8969    0.66   pos
##      c3!c3.var   2549      4177    0.61   pos
##      c4!c4.var   1792      3106    0.58   pos
##      c6!c6.var   1400      2551    0.55   pos
##      c8!c8.var   6456      9702    0.67   pos
##      r1:c1!r1.var 128000     189700  0.67   pos
##      r1:c2!r1.var 58230      90820  0.64   pos
##      r1:c3!r1.var 5531      16550  0.33   pos
##      new:gen!new.var 2869      1367   2.1   pos
##      R!variance 4412         915    4.8   pos

## End(Not run)

# -----

## Not run:
## require(asreml4)
## require(lucid)
## m3 <- asreml(yield ~ -1 + trtn, data=dat,
##             random = ~ r1 + r2 + r4 + r8 + r10 +
##             c1 + c2 + c3 + c4 + c6 + c8 + r1:c1 + r1:c2 + r1:c3 + new:gen)
## coef(m3)
## # REML cultivar means. Very similar to Federer table 2.
## rev(sort(round(coef(m3)$fixed[3] + coef(m3)$random[137:256,],0)))
## ## gen_G060 gen_G021 gen_G011 gen_G099 gen_G002
## ##      974      949      945      944      942
## ## gen_G118 gen_G058 gen_G035 gen_G111 gen_G120
## ##      938      937      937      933      932
## ## gen_G046 gen_G061 gen_G082 gen_G038 gen_G090
## ##      932      931      927      927      926

## vc(m3)
## ##          effect component std.error z.ratio constr
## ##      r1!r1.var   9201      13720   0.67   pos
## ##      r2!r2.var    241.7      1059    0.23   pos
## ##      r4!r4.var   2269      3915    0.58   pos
## ##      r8!r8.var   1355      2627    0.52   pos
## ##      r10!r10.var 1133      2312    0.49   pos

```

```

## ##      c1!c1.var      0.01      0      4.8  bound
## ##      c2!c2.var    5942      8969    0.66   pos
## ##      c3!c3.var    2549      4177    0.61   pos
## ##      c4!c4.var    1792      3106    0.58   pos
## ##      c6!c6.var    1400      2551    0.55   pos
## ##      c8!c8.var    6456      9702    0.67   pos
## ##      r1:c1!r1.var 128000    189700  0.67   pos
## ##      r1:c2!r1.var  58230    90820  0.64   pos
## ##      r1:c3!r1.var  5531    16550  0.33   pos
## ##      new:gen!new.var 2869    1367   2.1   pos
## ##      R!variance  4412     915   4.8   pos

## End(Not run)

```

federer . tobacco *Height of tobacco plants exposed to radiation*

Description

Height of tobacco plants exposed to radiation

Format

A data frame with 56 observations on the following 4 variables.

row row

block block, numeric

dose radiation dose, roentgens

height height of 20 plants, cm

Details

An experiment conducted in 1951 and described in Federer (1954). The treatment involved exposing tobacco seeds to seven different doses of radiation. The seedlings were transplanted to the field in an RCB experiment with 7 treatments in 8 blocks. The physical layout of the experiment was in 8 rows and 7 columns.

Shortly after the plants were transplanted to the field it became apparent that an environmental gradient existed. The response variable was the total height (centimeters) of 20 plants.

Source

Walter T Federer and C S Schlottfeldt, 1954. The use of covariance to control gradients in experiments. *Biometrics*, 10, 282–290. <http://doi.org/10.2307/3001881>

References

- R. D. Cook and S. Weisberg (1999). *Applied Regression Including Computing and Graphics*.
- Walter T Federer and Russell D Wolfinger, 2003. PROC GLM and PROC MIXED Codes for Trend Analyses for Row-Column Designed Experiments, *Handbook of Formulas and Software for Plant Geneticists and Breeders*, Haworth Press.
- Paul N Hinz, (1987). Nearest-Neighbor Analysis in Practice, *Iowa State Journal of Research*, 62, 199–217.

Examples

```
data(federer.tobacco)
dat <- federer.tobacco

# RCB analysis. Treatment factor not significant.
dat <- transform(dat, dosef=factor(dose), rowf=factor(row),
                 blockf=factor(block))
m1 <- lm(height ~ blockf + dosef, data=dat)
anova(m1)

# RCB residuals show strong spatial trends
if(require(desplot)){
  dat$resid <- resid(m1)
  desplot(resid ~ row * block, data=dat,
          # aspect unknown
          main="federer.tobacco")
}

# Row-column analysis. Treatment now significant
m2 <- lm(height ~ rowf + blockf + dosef, data=dat)
anova(m2)
```

fisher.barley

Multi-environment trial of 5 barley varieties, 6 locations, 2 years

Description

Multi-environment trial of 5 barley varieties, 6 locations, 2 years

Usage

```
data("fisher.barley")
```


Format

A data frame with 60 observations on the following 4 variables.

```
yield yield, bu/ac
gen genotype/variety, 5 levels
env environment/location, 2 levels
year year, 1931/1932
```

Details

Trials of 5 varieties of barley were conducted at 6 stations in Minnesota during the years 1931-1932.

This is a subset of Immer's barley data. The yield values here are totals of 3 reps (Immer gave the average yield of 3 reps).

Source

Ronald Fisher (1935). *The Design of Experiments*.

References

George Fernandez (1991). Analysis of Genotype x Environment Interaction by Stability Estimates. *Hort Science*, 26, 947-950.

F. Yates & W. G. Cochran (1938). The Analysis of Groups of Experiments. *Journal of Agricultural Science*, 28, 556-580, table 1. <http://doi.org/10.1017/S0021859600050978>

G. K. Shukla, 1972. Some statistical aspects of partitioning of genotype-environmental components of variability. *Heredity*, 29, 237-245. Table 1. <http://doi.org/10.1038/hdy.1972.87>

Examples

```
data(fisher.barley)
dat <- fisher.barley

if(require(dplyr) & require(lattice)) {
  # Yates 1938 figure 1. Regression on env mean
  # Sum years within loc
  dat2 <- aggregate(yield ~ gen + env, data=dat, FUN=sum)
  # Avg within env
  emn <- aggregate(yield ~ env, data=dat2, FUN=mean)
  dat2$envmn <- emn$yield[match(dat2$env, emn$env)]
  xyplot(yield ~ envmn, dat2, group=gen, type=c('p','r'),
         main="fisher.barley - stability regression",
         xlab="Environment total", ylab="Variety mean",
         auto.key=list(columns=3))
}

# -----
## Not run:
```

```

# calculate stability according to the sum-of-squares approach used by
# Shukla (1972), eqn 11. match to Shukla, Table 4, M.S. column
# also matches fernandez, table 3, stabvar column
require(dplyr)
dat2 <- dat
dat2 <- group_by(dat2, gen,env)
dat2 <- summarize(dat2, yield=sum(yield)) # means across years
dat2 <- group_by(dat2, env)
dat2 <- mutate(dat2, envmn=mean(yield)) # env means
dat2 <- group_by(dat2, gen)
dat2 <- mutate(dat2, genmn=mean(yield)) # gen means
dat2 <- ungroup(dat2)
dat2 <- mutate(dat2, grandmn=mean(yield)) # grand mean
# correction factor overall
dat2 <- mutate(dat2, cf = sum((yield - genmn - envmn + grandmn)^2))
t=5; s=6 # t genotypes, s environments
dat2 <- group_by(dat2, gen)
dat2 <- mutate(dat2, ss=sum((yield-genmn-envmn+grandmn)^2))
# divide by 6 to scale down to plot-level
dat2 <- mutate(dat2, sig2i = 1/((s-1)*(t-1)*(t-2)) * (t*(t-1)*ss-cf)/6)
dat2[!duplicated(dat2$gen),c('gen','sig2i')]
##      <chr>      <dbl>
## 1 Manchuria  25.87912
## 2 Peatland   75.68001
## 3 Svansota   19.59984
## 4 Trebi      225.52866
## 5 Velvet     22.73051

## End(Not run)

# -----

## Not run:
# mixed model approach gives similar results (but not identical)
# asreml3
library(asreml)
dat2 <- dat2[order(dat2$gen),]

# G-side
m1g <- asreml(yield ~ gen, data=dat2,
             random = ~ env + at(gen):units,
             family=asreml.gaussian(dispersion=1.0))
m1g <- update(m1g)
summary(m1g)$varcomp[-1,1:2]/6
##
## at(gen, Manchuria):units!units.var   gamma   component
## at(gen, Peatland):units!units.var    70.4838297  70.4838297
## at(gen, Svansota):units!units.var    25.2558315  25.2558315
## at(gen, Trebi):units!units.var       231.6923935  231.6923935
## at(gen, Velvet):units!units.var      13.9189381  13.9189381
## R!variance                           0.1666667   0.1666667

# R-side estimates = G-side estimate + 0.1666 (resid variance)

```

```

m1r <- asreml(yield ~ gen, data=dat2,
              random = ~ env,
              rcov = ~ at(gen):units) # or diag(gen):units
m1r <- update(m1r)
summary(m1r)$varcomp[-1,1:2]/6
##                gamma component
## gen_Manchuria!variance  34.03643  34.03643
## gen_Peatland!variance   70.72723  70.72723
## gen_Svansota!variance   25.38494  25.38494
## gen_Trebi!variance     231.84662 231.84662
## gen_Velvet!variance     14.05591  14.05591

## End(Not run)

```

fisher.latin

Latin square experiment on mangolds

Description

Latin square experiment on mangolds. Used by R. A. Fisher.

Usage

```
data("fisher.latin")
```

Format

A data frame with 25 observations on the following 4 variables.

trt treatment factor, 5 levels

yield yield

row row

col column

Details

Yields are root weights. Data originally collected by Mercer and Hall as part of a uniformity trial.

This data is the same as the data from columns 1-5, rows 16-20, of the mercer.mangold.uniformity data in this package.

Unsurprisingly, there are no significant treatment differences.

Source

Mercer, WB and Hall, AD, 1911. The experimental error of field trials *The Journal of Agricultural Science*, 4, 107-132. Table 1. <http://doi.org/10.1017/S002185960000160X>

R. A. Fisher. *Statistical Methods for Research Workers*.

Examples

```
data(fisher.latin)
dat <- fisher.latin

# Standard latin-square analysis
m1 <- lm(yield ~ trt + factor(row) + factor(col), data=dat)
anova(m1)
```

foulley.calving

Calving difficulty by calf sex and age of dam

Description

Calving difficulty by calf sex and age of dam

Usage

```
data("foulley.calving")
```

Format

A data frame with 54 observations on the following 4 variables.

sex calf gender

age dam age factor, 9 levels

score score for birthing difficulty, S1 < S2 < S3

count count of births for each category

Details

These data are calving difficulty scores for purebred US Simmental cows.

The raw data show that the greatest calving difficulty is for young dams with male calves. Differences between male/female calves decreased with age of the dam.

The goodness of fit can be improved by using a scaling effect for age of dam.

Note: The paper by Foulley and Gianola has '21943' as the count for score 1, F, >8. This data uses '20943' so that the marginal totals from this data match the marginal totals given in the paper.

Source

JL Foulley, D Gianola (1996). Statistical Analysis of Ordered Categorical Data via a Structured Heteroskedastic Threshold Model. *Genet Sel Evol*, 28, 249–273. <http://doi.org/10.1051/gse:19960304>

Used with permission of Jean-Louis Foulley.

Examples

```

data(foulley.calving)
dat <- foulley.calving

## Plot
d2 <- transform(dat,
  age=ordered(age, levels=c("0.0-2.0", "2.0-2.5", "2.5-3.0",
    "3.0-3.5", "3.5-4.0",
    "4.0-4.5", "4.5-5.0", "5.0-8.0", "8.0+")),
  score=ordered(score, levels=c('S1', 'S2', 'S3')))
if(require(reshape2)){
d2 <- acast(dat, sex+age~score, value.var='count')
d2 <- prop.table(d2, margin=1)
require(lattice)
thm <- simpleTheme(col=c('skyblue', 'gray', 'pink'))
barchart(d2, par.settings=thm, main="foulley.calving",
  xlab="Frequency of calving difficulty", ylab="Calf gender and dam age",
  auto.key=list(columns=3, text=c("Easy", "Assited", "Difficult")))
}

## Ordinal multinomial model
## Not run:
require(ordinal)
m2 <- clm(score ~ sex*age, data=dat, weights=count, link='probit')
summary(m2)

## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## sexM           0.500605   0.015178  32.982 < 2e-16 ***
## age2.0-2.5     -0.237643   0.013846 -17.163 < 2e-16 ***
## age2.5-3.0     -0.681648   0.018894 -36.077 < 2e-16 ***
## age3.0-3.5     -0.957138   0.018322 -52.241 < 2e-16 ***
## age3.5-4.0     -1.082520   0.024356 -44.446 < 2e-16 ***
## age4.0-4.5     -1.146834   0.022496 -50.981 < 2e-16 ***
## age4.5-5.0     -1.175312   0.028257 -41.594 < 2e-16 ***
## age5.0-8.0     -1.280587   0.016948 -75.559 < 2e-16 ***
## age8.0+        -1.323749   0.024079 -54.974 < 2e-16 ***
## sexM:age2.0-2.5 0.003035   0.019333   0.157  0.87527
## sexM:age2.5-3.0 -0.076677   0.026106  -2.937  0.00331 **
## sexM:age3.0-3.5 -0.080657   0.024635  -3.274  0.00106 **
## sexM:age3.5-4.0 -0.135774   0.032927  -4.124  3.73e-05 ***
## sexM:age4.0-4.5 -0.124303   0.029819  -4.169  3.07e-05 ***
## sexM:age4.5-5.0 -0.198897   0.038309  -5.192  2.08e-07 ***
## sexM:age5.0-8.0 -0.135524   0.022804  -5.943  2.80e-09 ***
## sexM:age8.0+   -0.131033   0.031852  -4.114  3.89e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Threshold coefficients:
##           Estimate Std. Error z value
## S1|S2  0.82504    0.01083   76.15

```

```
## S2|S3  1.52017    0.01138  133.62

## Note 1.52017 - 0.82504 = 0.695 matches Foulley's '2-3' threshold estimate

predict(m2) # probability of each category
detach("package:ordinal") # to avoid VarCorr clash with lme4

## End(Not run)
```

| | |
|-----------|---|
| fox.wheat | <i>Multi-environment trial of wheat for 22 varieties at 14 sites in Australia</i> |
|-----------|---|

Description

Wheat yields of 22 varieties at 14 sites in Australia

Usage

```
data("fox.wheat")
```

Format

A data frame with 308 observations on the following 4 variables.

gen genotype/variety factor, 22 levels

site site factor, 14 levels

yield yield, tonnes/ha

state state in Australia

Details

The 1975 Interstate Wheat Variety trial in Australia used RCB design with 4 blocks, 22 varieties in 14 sites. Wagga is represented twice, by trials sown in May and June.

The 22 varieties were a highly selected and represent considerable genetic diversity with four different groups. (i) from the University of Sydney: Timson, Songlen, Gamenya. (ii) widely grown on Mallee soils: Heron and Halberd. (iii) late maturing varieties from Victoria: Pinnacle, KL-21, JL-157. (iv) with Mexican parentage: WW-15 and Oxley.

Source

Fox, P.N. and Rathjen, A.J. (1981). Relationships between sites used in the interstate wheat variety trials. *Australian Journal of Agricultural Research*, 32, 691-702.

Electronic version supplied by Jonathan Godfrey.

Examples

```

data(fox.wheat)
dat <- fox.wheat

# Means of varieties. Slight differences from Fox and Rathjen suggest
# they had more decimals of precision than shown.
tapply(dat$yield, dat$gen, mean)

# Calculate genotype means, merge into the data
genm <- tapply(dat$yield, dat$gen, mean)
dat$genm <- genm[match(dat$gen, names(genm))]

# Calculate slopes for each site. Matches Fox, Table 2, Col A.
m1 <- lm(yield~site+site:genm, data=dat)
sort(round(coef(m1)[15:28],2), dec=TRUE)

# Figure 1 of Fox
require(lattice)
xyplot(yield~genm|state, data=dat, type=c('p','r'), group=site,
       auto.key=list(columns=4),
       main="fox.wheat", xlab="Variety mean across all sites",
       ylab="Variety yield at each site within states")

```

```
garber.multi.uniformity
```

Uniformity trials of oat hay and wheat grain

Description

Uniformity trials of oat hay and wheat grain, at West Virginia Agricultural Experiment Station, 1923-1924.

Format

A data frame with 270 observations on the following 4 variables.

row row

col column

oats yield of oat hay

wheat yield of wheat grain

Details

Each plot was 68 feet x 21 feet. After discarding a 3.5 foot border on all sides, the harvested area was 61 feet x 14 feet. The plots were laid out in doubles with a 14-foot roadway between the plots.

Note: The orientation of the plots is a guess.

Field width: 6 plots * 68 feet + 14 ft/roadway * 2 = 436 feet

Field length: 45 plots * 21 feet/plot = 945 feet

In 1923 the field was planted to oats and harvested as hay. In 1924 the field was planted to wheat and harvested for grain.

Garber: "Plots 211 to 214, and 261 to 264, [note, these are rows 11-14, columns 5-6] inclusive, were eliminated from this study because of the fact that a few years ago a straw stack had stood on or in the vicinity...which undoubtedly accounts for the relatively high yields on plots 261 to 264, inclusive."

Note: The source document contained mean-subtracted yields. The oat yield in row 22, column 5 was given as +59.7. This is obviously incorrect, since the negative yields all end in '.7' and positive yields all ended in '.3'. We used -59.7 as the centered yield value and added the mean of 1883.7 to all centered yields to obtain absolute yields.

Source

Garber, RJ and McIlvaine, TC and Hoover, MM. 1926. A study of soil heterogeneity in experiment plots. *Jour. Agr. Res*, 33, 255-268. <http://naldc.nal.usda.gov/download/IND43967148/PDF>

Examples

```
data(garber.multi.uniformity)
dat <- garber.multi.uniformity

if(require(desplot)){
  desplot(oats ~ col*row, data=dat,
          flip=TRUE, tick=TRUE, aspect=945/436, # true aspect
          main="garber.multi.uniformity oats")
}

if(require(desplot)){
  desplot(wheat ~ col*row, data=dat,
          flip=TRUE, tick=TRUE, aspect=945/436, # true aspect
          main="garber.multi.uniformity wheat")
}

with(dat, cor(oats, wheat)) # = .37 matches Garber
```

`gartner.corn`*Yield monitor data from a corn field in Minnesota*

Description

Yield monitor data from a corn field in Minnesota

Usage

```
data("gartner.corn")
```

Format

A data frame with 4949 observations on the following 8 variables.

`long` longitude

`lat` latitude

`mass` grain mass flow per second, pounds

`time` GPS time, in seconds

`seconds` seconds elapsed for each datum

`dist` distance traveled for each datum, in inches

`moist` grain moisture, percent

`elev` elevation, feet

Details

The data was collected 5 Nov 2011 from a corn field south of Mankato, Minnesota, using a combine-mounted yield monitor. <https://www.google.com/maps/place/43.9237575,-93.9750632>

Each harvested swath was 12 rows wide = 360 inches.

Time 0 is 5 Nov 2011, 12:38:03 Central Time. Time 16359 = 4.54 hours later.

Yield is calculated as total dry weight (corrected to 15.5 percent moisture), divided by 56 pounds (to get bushels), divided by the harvested area. $\text{drygrain} = [\text{massflow} * \text{seconds} * (100 - \text{moisture}) / (100 - 15.5)] / 56$ harvested area = (distance * swath width) / 6272640 yield = drygrain / area

Source

Originally from University of Minnesota Precision Agriculture Center. <http://www.soils.umn.edu/academics/classes/soil4111>

Retrieved 27 Aug 2015 from <https://web.archive.org/web/20100717003256/http://www.soils.umn.edu/academics/classes/soil4111>

Used under Creative Commons BY-SA 3.0 license.

Examples

```

data(gartner.corn)
dat <- gartner.corn

# Calculate yield
dat <- transform(dat, yield=(mass*seconds*(100-moist)/(100-15.5)/56)/(dist*360/6272640))
# Delete yield outliers
dat <- subset(dat, yield >50)

# Colors for yield
medy <- median(dat$yield)
ncols <- 20
wwidth <- 150
brks <- seq(from = -wwidth/2, to=wwidth/2, length=ncols-1)
brks <- c(-250, brks, 250) # 250 is safe..we cleaned data outside (50,450)?
yldbks <- brks + medy
dat <- transform(dat, yldbin = as.numeric(cut(yield, breaks= yldbks)))

# Add polygons for soil map units
# Go to: http://websoilsurvey.nrcs.usda.gov/app/WebSoilSurvey.aspx
# Click: Lat and Long. 43.924, -93.975
# Click the little AOI rectangle icon. Drag around the field
# In the AOI Properties, enter the Name: Gartner
# Click the tab Soil Map to see map unit symbols, names
# Click: Download Soils Data. Click: Create Download Link.
# Download the zip file and find the soilmu_a_aoi files.

# Read shape files
if(require(rgdal)){
# shp <- readOGR("C:/x/rpack/agridat/inst/files", "gartner.corn")
shp <- readOGR(system.file(package="agridat", "files"), "gartner.corn")

# Plot the shapefiles first to set up the coordinate system
plot(shp, xlim=range(dat$long), ylim=range(dat$lat))
box() # Add the yield points
redblue <- colorRampPalette(c("firebrick", "lightgray", "#375997"))
with(dat, points(long,lat, main="yield heat map",
                 col=redblue(ncols)[yldbin], cex=.75, pch=16))
plot(shp, add=TRUE, lwd=2) # Overlay soil polygons on top
title("gartner.corn - yield heatmap with soil map unit symbols")
# Manual annotation of soil map units
text(x = c(-93.97641, -93.97787, -93.97550, -93.97693, -93.97654, -93.97480,
           -93.97375, -93.978284, -93.977617, -93.976715, -93.975929),
     y = c(43.92185, 43.92290, 43.92358, 43.92445, 43.92532, 43.92553,
           43.92568, 43.922163, 43.926427, 43.926993, 43.926631),
     lab=c("110", "319", "319", "230", "105C", "110", "211", "110", "211", "230", "105C"))

# Trim off the ends of the field & re-do image above
dat <- subset(dat, lat < 43.925850 & lat > 43.921178)

# Identify the soil type for each yield point

```

```

dat$ix <- over(SpatialPoints(dat[ , c('long','lat')]),
              SpatialPolygons(shp@polygons))
dat$mu <- shp@data[, "MUSYM"][dat$ix]
# Check the points are properly identified
# with(dat, points(long,lat, col=redblue(ncols)[ix], cex=.75, pch=16))

# Aggregate points by soil type and analyze
tapply(dat$yield, dat$mu, mean)
tapply(dat$yield, dat$mu, sd)
require(lattice)
densityplot(~yield|mu, dat, layout=c(1,5),
           main="gartner.corn - yield density plot")
}

## Not run:
# Draw a 3D surface. Clearly shows the low drainage area
library(rgl)
dat <- transform(dat, x=long-min(long), y=lat-min(lat), z=elev-min(elev))
clear3d()
points3d(dat$x, dat$y, dat$z/50000,
         col=redblue(ncols)[dat$yldbin])
axes3d()
title3d(xlab='x', ylab='y', zlab='elev')
rgl.close()

## End(Not run)

```

gathmann.bt

*Impact of Bt corn on non-target species***Description**

Impact of Bt corn on non-target species

Format

A data frame with 16 observations on the following 3 variables.

gen genotype/maize, Bt ISO

thysan thysan abundance

aranei aranei abundance

Details

The experiment involved comparing a Bt maize and a near-isogenic control variety.

Species abundances were measured for *Thysanoptera* (thrips) and *Araneida* (spiders) in 8 different plots.

Each response is probably a mean across repeated measurements.

Source

L. A. Hothorn, 2005. Evaluation of Bt-Maize Field Trials by a Proof of Safety. <http://www.seedtest.org/upload/cms/user/presen>
Used with permission of Achim Gathmann.

Examples

```

data(gathmann.bt)
dat <- gathmann.bt

# EDA suggests Bt vs ISO is significant for thysan, not for aranei
require(lattice)
if(require(reshape2)){
d2 <- melt(dat, id.var='gen')
bwplot(value ~ gen|variable, d2,
        main="gathmann.bt", ylab="Insect abundance",
        panel=function(x,y,...){
          panel.xyplot(jitter(as.numeric(x)),y,...)
          panel.bwplot(x,y,...)
        },
        scales=list(relation="free"))
}

## Not run:

# ----- Parametric CI. Thysan significant, aranei not significant.
require(equivalence)

th0 <- with(dat, tost(thysan[1:8], thysan[9:16], alpha=.05, paired=FALSE))
lapply(th0[c("estimate","tost.interval")], round, 2)
# 14.28-8.72=5.56, (2.51, 8.59) # match Gathmann p. 11

ar0 <- with(dat, tost(aranei[1:8], aranei[9:16], alpha=.05, epsilon=.4))
lapply(ar0[c("estimate","tost.interval")], round, 2)
# .57-.47=.10, (-0.19, 0.40) # match Gathmann p. 11

# ----- Non-parametric exact CI. Same result.
require(coin)

th1 <- wilcox_test(thysan ~ gen, data=dat, conf.int=TRUE, conf.level=0.90)
lapply(confint(th1), round, 2)
# 6.36, (2.8, 9.2) # Match Gathmann p. 11

ar1 <- wilcox_test(aranei ~ gen, data=dat, conf.int=TRUE, conf.level=0.90)
lapply(confint(ar1), round, 2)
# .05 (-.2, .4)

# ----- Log-transformed exact CI. Same result.
th2 <- wilcox_test(log(thysan) ~ gen, data=dat, alternative=c("two.sided"),

```

```

                                conf.int=TRUE, conf.level=0.9)
lapply(confint(th2), function(x) round(exp(x),2))
# 1.66, (1.38, 2.31) # Match Gathmann p 11

# ----- Log-transform doesn't work on aranei, but asinh(x/2) does
ar2 <- wilcox_test(asinh(aranei/2) ~ gen, data=dat,
                  alternative=c("two.sided"),
                  conf.int=TRUE, conf.level=0.9)
lapply(confint(ar2), function(x) round(sinh(x)*2,1))

## End(Not run)

```

gauch.soy

Multi-environment trial of soybeans in New York, 1977 to 1988

Description

New York soybean yields, 1977 to 1988, for 7 genotypes in 55 environments.

Format

A data frame with 1454 observations on the following 4 variables.

yield yield, kg/ha
 rep repeated measurement
 gen genotype, 7 levels
 env environment, 55 levels

Details

Soybean yields at 13 percent moisture for 7 genotypes in 55 environments with 4 replicates. Some environments had only 2 or 3 replicates. The experiment was an RCB design, but some plots were missing and there were many other soybean varieties in the experiment. The replications appear in random order and do *NOT* define blocks. Environment names are a combination of the first letter of the location name and the last two digits of the year. The location codes are C=Chazy, N=Canton, L=Lockport, G=Geneseo, R=Romulus, A=Aurora, I=Ithica, V=Valatie, D=Riverhead. Plots were 7.6 m long, four rows wide (middle two rows were harvested).

This data has been widely used (in various subsets) to promote the benefits of AMMI (Additive Main Effects Multiplicative Interactions) analyses.

The gen x env means of Table 1 (Zobel et al 1998) are least-squares means (personal communication).

Source

Retrieved Sep 2011 from http://www.microcomputerpower.com/matmodel/matmodelmatmodel_sample_.html
 Used with permission of Hugh Gauch.

References

Zobel, RW and Wright, MJ and Gauch Jr, HG. 1998. Statistical analysis of a yield trial, *Agronomy journal*, 80, 388-393. <https://doi.org/10.2134/agronj1988.00021962008000030002x>

Examples

```
data(gauch.soy)
dat <- gauch.soy

# AMMI biplot
if(require(agricolae)){
  # Figure 1 of Zobel et al 1988, means vs PC1 score
  dat2 <- droplevels(subset(dat, is.element(env, c("A77", "C77", "V77",
    "V78", "A79", "C79", "G79", "R79", "V79", "A80", "C80", "G80", "L80", "D80",
    "R80", "V80", "A81", "C81", "G81", "L81", "D81", "R81", "V81", "A82", "L82",
    "G82", "V82", "A83", "I83", "G83", "A84", "N84", "C84", "I84", "G84"))))

  m2 <- with(dat2, AMMI(env, gen, rep, yield))
  bip <- m2$biplot
  with(bip, plot(yield, PC1, type='n', main="gauch.soy -- AMMI biplot"))
  with(bip, text(yield, PC1, rownames(bip),
    col=ifelse(bip$type=="GEN", "darkgreen", "blue"),
    cex=ifelse(bip$type=="GEN", 1.5, .75)))
}
```

giles.wheat

Straw length and ear emergence for wheat genotypes.

Description

Straw length and ear emergence for wheat genotypes. Data are unbalanced with respect to experiment year and genotype.

Usage

```
data("giles.wheat")
```

Format

A data frame with 247 observations on the following 4 variables.

gen genotype. Note, this is numeric!

env environment

straw straw length

emergence ear emergence, Julian date

Details

Highly unbalanced data of straw length and ear emergence date for wheat genotypes.

The 'genotype' column is called 'Accession number' in original data. The genotypes were chosen to represent the range of variation in the trait.

The Julian date was found to be preferable to other methods (such as days from sowing).

Piepho (2003) fit a bilinear model to the straw emergence data. This is similar to Finlay-Wilkinson regression.

Source

R. Giles (1990). Utilization of unreplicated observations of agronomic characters in a wheat germplasm collection. In: *Wheat Genetic Resources. Meeting Diverse Needs*. Wiley, Chichester, U.K., pp.113-130.

References

Piepho, HP (2003). Model-based mean adjustment in quantitative germplasm evaluation data. *Genetic Resources and Crop Evolution*, 50, 281-290. <http://doi.org/10.1023/A:1023503900759>

Examples

```
data(giles.wheat)
dat <- giles.wheat
dat <- transform(dat, gen=factor(gen))
dat <- transform(dat, env=factor(env))
dat1 <- subset(dat, !is.na(straw))
dat2 <- subset(dat, !is.na(emergence))

# Traits are not related
# with(dat, plot(straw~emergence))

# Show unbalancedness of data
if(require(lattice) & require(reshape2)){
  redblue <- colorRampPalette(c("firebrick", "lightgray", "#375997"))
  levelplot(acast(dat, gen ~ env, value.var='emergence'),
            col.regions=redblue,
            xlab="genotype", ylab="year",
            main="giles.wheat - straw length")
}

# Analysis of straw length

if(require(lsmeans) & require(gnm)){

# Mean across years. Matches Piepho Table 7 'Simple'
m1 = lm(straw ~ gen, data=dat1)
lsmeans(m1, 'gen')

# Simple two-way model. NOT the bi-additive model of Piepho.
```

```

m2 = lm(straw ~ gen + env, data=dat1)
lsmeans(m2, 'gen')

# Bi-additive model. Matches Piepho Table 6, rows (c)
library(gnm)
m3 <- gnm(straw ~ env + Mult(gen,env), data=dat1)
cbind(adjusted=round(fitted(m3),0), dat1)

}

# Analysis of Ear emergence

if(require(lsmeans)){

# Simple two-way model.
m4 = lm(emergence ~ 1 + gen + env, data=dat2)
lsmeans(m4, c('gen','env')) # Matches Piepho Table 9. rpws (c)
lsmeans(m4, 'gen') # Match Piepho table 10, Least Squares column

}

```

gilmour.serpentine

Wheat yield in South Australia with serpentine row/col effects

Description

An RCB experiment of wheat in South Australia, with strong spatial variation and serpentine row/column effects.

Format

A data frame with 330 observations on the following 5 variables.

col column

row row

rep replicate factor, 3 levels

gen wheat variety, 108 levels

yield yield

Details

A randomized complete block experiment. There are 108 varieties in 3 reps. Plots are 6 meters long, 0.75 meters wide, trimmed to 4.2 meters lengths before harvest. Trimming was done by spraying the wheat with herbicide. The sprayer travelled in a serpentine pattern up and down columns. The trial was sown in a serpentine manner with a planter that seeds three rows at a time (Left, Middle, Right).

Field length 22 plots * .75 m = 16.5 m

Field width 15 columns * 6 m = 90 m

Source

Arthur R Gilmour and Brian R Cullis and Arunas P Verbyla, 1997. Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agric Biol Env Statistics*, 2, 269-293.

Used with permission of Arthur Gilmour, in turn with permission from Gil Hollamby.

References

N. W. Galwey. 2014. *Introduction to Mixed Modelling: Beyond Regression and Analysis of Variance*. Table 10.9

Examples

```
data(gilmour.serpentine)
dat <- gilmour.serpentine

if(require(desplot)){
  desplot(yield~ col*row, data=dat,
          num=gen, show.key=FALSE, out1=rep,
          aspect = 16.5/90, # true aspect
          main="gilmour.serpentine")
}

# Extreme field trend. Blocking insufficient--needs a spline/smoother
# xyplot(yield~col, data=dat, main="gilmour.serpentine")

# -----

## Not run:
# asreml3
require(asreml)
dat <- transform(dat, rowf=factor(row), colf=factor(10*(col-8)))
dat <- dat[order(dat$rowf, dat$colf), ] # Sort order needed by asreml

# RCb
m0 <- asreml(yield ~ gen, data=dat, random=~rep)

# Add AR1 x AR1
m1 <- asreml(yield ~ gen, data=dat, rcov = ~ar1(rowf):ar1(colf))

# Add spline
m2 <- asreml(yield ~ gen + col, data=dat,
            random= ~ spl(col) + colf,
            rcov = ~ar1(rowf):ar1(colf))

# Figure 4 shows serpentine spraying
p2 <- predict(m2, data=dat, classify="colf")$predictions$pvals
plot(p2$predicted, type='b', xlab="column number", ylab="BLUP")

# Define column code (due to serpentine spraying)
```

```

# Rhelp doesn't like double-percent modulus symbol, so compute by hand
dat <- transform(dat, colcode = factor(dat$col-floor((dat$col-1)/4)*4 -1))

m3 <- asreml(yield ~ gen + lin(colf) + colcode, data=dat,
             random= ~ colf + rowf + spl(colf),
             rcov = ~ar1(rowf):ar1(colf))

# Figure 6 shows serpentine row effects
p3 <- predict(m3, data=dat, classify="rowf")$predictions$pvals
plot(p3$predicted, type='l', xlab="row number", ylab="BLUP")
text(1:22, p3$predicted, c('L','L','M','R','R','M','L','L',
                          'M','R','R','M','L','L','M','R','R','M','L','L','M','R'))

# Define row code (due to serpentine planting). 1=middle, 2=left/right
dat <- transform(dat, rowcode = factor(row))
levels(dat$rowcode) <- c('2','2','1','2','2','1','2','2','1',
                       '2','2','1','2','2','1','2','2','1','2','2','1','2')

m6 <- asreml(yield ~ gen + lin(colf) + colcode + rowcode, data=dat,
             random= ~ colf + rowf + spl(col),
             rcov = ~ar1(rowf):ar1(colf))
plot(variogram(m6), xlim=c(0:17), ylim=c(0,11), zlim=c(0,4000),
     main="gilmour.serpentine")

## End(Not run)

# -----

## Not run:
## require(asreml4)
## dat <- transform(dat, rowf=factor(row), colf=factor(10*(col-8)))
## dat <- dat[order(dat$rowf, dat$colf), ] # Sort order needed by asreml

## # RCB
## m0 <- asreml(yield ~ gen, data=dat, random=~rep)

## # Add AR1 x AR1
## m1 <- asreml(yield ~ gen, data=dat, resid = ~ar1(rowf):ar1(colf))

## # Add spline
## m2 <- asreml(yield ~ gen + col, data=dat,
##             random= ~ spl(col) + colf,
##             resid = ~ar1(rowf):ar1(colf))

## # Figure 4 shows serpentine spraying
## p2 <- predict(m2, data=dat, classify="colf")$pvals
## plot(p2$predicted, type='b', xlab="column number", ylab="BLUP")

## # Define column code (due to serpentine spraying)
## # Rhelp doesn't like double-percent modulus symbol, so compute by hand
## dat <- transform(dat, colcode = factor(dat$col-floor((dat$col-1)/4)*4 -1))

## m3 <- asreml(yield ~ gen + lin(colf) + colcode, data=dat,

```

```

##          random= ~ colf + rowf + spl(colf),
##          resid = ~ar1(rowf):ar1(colf))

## # Figure 6 shows serpentine row effects
## p3 <- predict(m3, data=dat, classify="rowf")$pvals
## plot(p3$predicted, type='l', xlab="row number", ylab="BLUP")
## text(1:22, p3$predicted, c('L','L','M','R','R','M','L','L',
##                            'M','R','R','M','L','L','M','R','R','M','L','L','M','R'))

## # Define row code (due to serpentine planting). 1=middle, 2=left/right
## dat <- transform(dat, rowcode = factor(row))
## levels(dat$rowcode) <- c('2','2','1','2','2','1','2','2','1',
##                          '2','2','1','2','2','1','2','2','1','2','2','1','2')

## m6 <- asreml(yield ~ gen + lin(colf) + colcode +rowcode, data=dat,
##              random= ~ colf + rowf + spl(col),
##              resid = ~ar1(rowf):ar1(colf))
## plot(varioGram(m6), xlim=c(0:17), ylim=c(0,11), zlim=c(0,4000),
##       main="gilmour.serpentine")

## End(Not run)

```

gilmour.slatehall *Slate Hall Farm 1978*

Description

Yields for a trial at Slate Hall Farm in 1978.

Format

A data frame with 150 observations on the following 5 variables.

row row

col column

yield yield (grams/plot)

gen genotype factor, 25 levels

rep rep factor, 6 levels

Details

The trial was of spring wheat at Slate Hall Farm in 1978. The experiment was a balanced lattice with 25 varieties in 6 replicates. The 'rep' labels are arbitrary (no rep labels appeared in the source data). Each row within a rep is an incomplete block. The plot size was 1.5 meters by 4 meters.

Field length: 15 plots x 1.5 meters = 22.5 m

Field width: 10 plots x 4 m = 40 m

Source

Arthur R Gilmour and Brian R Cullis and Arunas P Verbyla (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological, and Environmental Statistics*, 2, 269-293. <http://doi.org/10.2307/1400446>

References

None.

Examples

```

data(gilmour.slatehall)
dat <- gilmour.slatehall

if(require(desplot)){
  desplot(yield ~ col * row, dat,
          aspect=22.5/40, num=gen, out1=rep, cex=1,
          main="gilmour.slatehall")
}

# -----

## Not run:
# Model 4 of Gilmour et al 1997
# asreml3
require(asreml)
dat <- transform(dat, xf=factor(col), yf=factor(row))
dat <- dat[order(dat$xf, dat$yf), ]
m4 <- asreml(yield ~ gen + lin(row), data=dat,
            random = ~ dev(row) + dev(col),
            rcov = ~ ar1(xf):ar1(yf))
coef(m4)$fixed[1] # linear row
# [1] 31.72252 # (sign switch due to row ordering)

require(lucid)
vc(m4)
##      effect component std.error z.ratio constr
## dev(row) 20290      10260      2      pos
## dev(col)  2519       1959      1.3     pos
## R!variance 23950      4616      5.2     pos
## R!xf.cor   0.439      0.113      3.9    uncon
## R!yf.cor   0.125      0.117      1.1    uncon

plot(variogram(m4))

## End(Not run)

# -----

## Not run:
# Model 4 of Gilmour et al 1997

```

```

## require(asreml4)
## dat <- transform(dat, xf=factor(col), yf=factor(row))
## dat <- dat[order(dat$xf, dat$yf), ]
## m4 <- asreml(yield ~ gen + lin(row), data=dat,
##             random = ~ dev(row) + dev(col),
##             resid = ~ ar1(xf):ar1(yf))
## coef(m4)$fixed[1] # linear row
## # [1] 31.72252 # (sign switch due to row ordering)

## require(lucid)
## vc(m4)
## ##          effect component std.error z.ratio bound
## ##      dev(col) 2519      1959      1.3    P    0
## ##      dev(row) 20290     10260      2      P    0
## ##      xf:yf(R) 23950      4616      5.2    P    0
## ## xf:yf!xf!cor  0.439    0.113      3.9    U    0
## ## xf:yf!yf!cor  0.125    0.117      1.1    U    0

## plot(varioGram(m4))

## End(Not run)

```

gomez.fractionalfactorial

Rice fractional factorial experiment 1/2 2⁶.

Description

Rice fractional factorial experiment 1/2 2⁶. Two reps with 2 blocks in each rep.

Format

A data frame with 64 observations on the following 6 variables.

yield grain yield in tons/ha

rep replicate, 2 levels

block block within rep, 2 levels

trt treatment, levels (1) to abcdef

col column position in the field

row row position in the field

Details

Grain yield from a 2⁶ fractional factorial experiment in blocks of 16 plots each, with two replications.

Gomez has some inconsistencies. One example:

Page 171: treatment (1) in rep 1, block 2 and rep 2, block 1.

Page 172: treatment (1) in Rep 1, block 1 and rep 2, block 1.

This data uses the layout shown on page 171.

Source

Gomez, K.A. and Gomez, A.A.. 1984, *Statistical Procedures for Agricultural Research*. Wiley-Interscience. Page 171-172.

Used with permission of Kwanchai Gomez.

Examples

```
data(gomez.fractionalfactorial)
dat <- gomez.fractionalfactorial

# trt abcdef has the highest yield
# Gomez, Figure 4.8
if(require(desplot)){
  desplot(yield~col*row, dat,
          # aspect unknown
          text=trt, shorten="none", show.key=FALSE, cex=1,
          main="gomez.fractionalfactorial - treatment & yield")
}

# Split treatment into individual factors
dat <- transform(dat,
                 a = -1 + 2 * grepl('a',trt),
                 b = -1 + 2 * grepl('b',trt),
                 c = -1 + 2 * grepl('c',trt),
                 d = -1 + 2 * grepl('d',trt),
                 e = -1 + 2 * grepl('e',trt),
                 f = -1 + 2 * grepl('f',trt))

# Gomez table 4.24, trt SS totalled together.
# Why didn't Gomez nest block within rep?
m0 <- lm(yield ~ rep * block + trt, dat)
anova(m0)

# Gomez table 4.24, trt SS split apart
m1 <- lm(yield ~ rep * block + (a+b+c+d+e+f)^3, dat)
anova(m1)

## Not run:
require(FrF2)
aliases(m1)
MEPlot(m1, select=3:8,
       main="gomez.fractionalfactorial - main effects plot")

## End(Not run)
```

gomez.groupsplrit *Group balanced split-plot design in rice*

Description

Group balanced split-plot design in rice

Format

A data frame with 270 observations on the following 7 variables.

col column

row row

rep replicate factor, 3 levels

fert fertilizer factor, 2 levels

gen genotype factor, 45 levels

group grouping (genotype) factor, 3 levels

yield yield of rice

Details

Genotype group S1 is less than 105 days growth duration, S2 is 105-115 days growth duration, S3 is more than 115 days.

Source

Gomez, K.A. and Gomez, A.A.. 1984, *Statistical Procedures for Agricultural Research*. Wiley-Interscience. Page 120.

Used with permission of Kwanchai Gomez.

Examples

```
data(gomez.groupsplrit)
dat <- gomez.groupsplrit

# Gomez figure 3.10. Obvious fert and group effects
if(require(desplot)){
  desplot(group ~ col*row, dat,
          out1=rep, col=fert, text=gen, # aspect unknown
          main="gomez.groupsplrit")
}

# Gomez table 3.19 (not partitioned by group)
m1 <- aov(yield ~ fert*group + gen:group + fert:gen:group +
          Error(rep/fert/group), data=dat)
summary(m1)
```

gomez.heteroskedastic *RCB in rice with heteroskedastic varieties*

Description

RCB in rice with heteroskedastic varieties, 35 genotypes, 3 reps

Usage

```
data("gomez.heteroskedastic")
```

Format

A data frame with 105 observations on the following 4 variables.

```
gen genotype
group group of genotypes
rep replicate
yield yield
```

Details

RCB design with three reps. Genotypes 1-15 are hybrids, 16-32 are parents, 33-35 are checks.

Source

Gomez, K.A. and Gomez, A.A.. 1984, *Statistical Procedures for Agricultural Research*. Wiley-Interscience. Page 310.

Used with permission of Kwanchai Gomez.

References

None.

Examples

```
data(gomez.heteroskedastic)
dat <- gomez.heteroskedastic

# Fix the outlier as reported by Gomez p. 311
dat[dat$gen=="G17" & dat$rep=="R2", "yield"] <- 7.58

if(require(lattice)){
  bwplot(gen ~ yield, dat, group=as.numeric(dat$group),
         ylab="genotype", main="gomez.heterogeneous")
}
```



```
# Match Gomez table 7.28
m1 <- lm(yield ~ rep + gen, data=dat)
anova(m1)
## Response: yield
##           Df Sum Sq Mean Sq F value    Pr(>F)
## rep         2  3.306  1.65304   5.6164 0.005528 **
## gen        34 40.020  1.17705   3.9992 5.806e-07 ***
## Residuals 68 20.014  0.29432
```

gomez.multilocsplitplot

Multi-environment trial of rice, split-plot design

Description

Grain yield was measured at 3 locations with 3 reps per location. Within each rep, the main plot was 6 nitrogen fertilizer treatments and the sub plot was 2 rice varieties.

Format

A data frame with 108 observations on the following 5 variables.

```
loc location, 3 levels
nitro nitrogen in kg/ha
rep replicate, 3 levels
gen genotype, 2 levels
yield yield, kg/ha
```

Source

Gomez, K.A. and Gomez, A.A.. 1984, *Statistical Procedures for Agricultural Research*. Wiley-Interscience. Page 339.

Used with permission of Kwanchai Gomez.

Examples

```
data(gomez.multilocsplitplot)
dat <- gomez.multilocsplitplot
dat$nf <- factor(dat$nitro)

# Gomez figure 8.3
require(lattice)
xyplot(yield~nitro, dat, group=loc, type=c('p','smooth'), auto.key=TRUE,
       main="gomez.multilocsplitplot")

# AOV
```

```

# Be careful to use the right stratum, 'nf' appears in both strata.
# Still not quite the same as Gomez table 8.21
t1 <- terms(yield ~ loc * nf * gen + Error(loc:rep:nf),
            "Error", keep.order=TRUE)
m1 <- aov(t1, data=dat)
summary(m1)

## Not run:
# F values are somewhat similar to Gomez Table 8.21
require(lme4)
m2 <- lmer(yield ~ loc*nf*gen + (1|loc/rep/nf), dat)
anova(m2)
## Analysis of Variance Table
##           Df   Sum Sq Mean Sq F value
## loc         2   117942    58971  0.1525
## nf          5 72841432 14568286 37.6777
## gen         1   7557570    7557570 19.5460
## loc:nf      10 10137188   1013719  2.6218
## loc:gen     2   4270469   2135235  5.5223
## nf:gen      5   1501767    300353  0.7768
## loc:nf:gen 10   1502273    150227  0.3885

## End(Not run)

```

gomez.nitrogen

Soil nitrogen at three times for eight fertilizer treatments

Description

Soil nitrogen at three times for eight fertilizer treatments

Format

A data frame with 96 observations on the following 4 variables.

trt nitrogen treatment factor

nitro soil nitrogen content, percent

rep replicate

stage growth stage, three periods

Details

Eight fertilizer treatments were tested.

Soil nitrogen content was measured at three times. P1 = 15 days post transplanting. P2 = 40 days post transplanting. P3 = panicle initiation.

Source

Gomez, K.A. and Gomez, A.A.. 1984, *Statistical Procedures for Agricultural Research*. Wiley-Interscience. Page 259.

Used with permission of Kwanchai Gomez.

References

R-help mailing list, 9 May 2013. Data provided by Cyril Lundrigan. Analysis method by Rich Heiberger.

Examples

```
data(gomez.nitrogen)
dat <- gomez.nitrogen

# Note the depletion of nitrogen over time (stage)
if(require(HH)){
  interaction2wt(nitro ~ rep/trt + trt*stage, data=dat,
                x.between=0, y.between=0,
                main="gomez.nitrogen")

  # Just the fertilizer profiles
  with(dat, interaction.plot(stage, trt, nitro,
                            col=1:4, lty=1:3, main="gomez.nitrogen",
                            xlab="Soil nitrogen at three times for each treatment"))
}

# Gomez table 6.16
m1 <- aov(nitro ~ Error(rep/trt) + trt*stage, data=dat)
summary(m1)

# Gomez table 6.18
# Treatment          1 2 3 4 5 6 7 8
cont <- cbind("T7 vs others" = c( 1, 1, 1, 1, 1, 1, -7, 1),
              "T8 vs others" = c( 1, 1, 1, 1, 1, 1, 0, -6),
              "T2,T5 vs others" = c(-1, 2, -1, -1, 2, -1, 0, 0),
              "T2 vs T5" = c( 0, 1, 0, 0, -1, 0, 0, 0))
contrasts(dat$trt) <- cont
contrasts(dat$trt)

m2 <- aov(nitro ~ Error(rep/trt) + trt*stage, data=dat)
summary(m2, expand.split=FALSE,
        split=list(trt=list(
                    "T7 vs others"=1,
                    "T8 vs others"=2,
                    "T2,T5 vs others"=3,
                    "T2 vs T5"=4,
                    rest=c(5,6,7)),
                  "trt:stage"=list(
                    "(T7 vs others):P"=c(1,8),
```

```
"(T8 vs others):P"=c(2,9),  
"(T2,T5 vs others):P"=c(3,10),  
"(T2 vs T5):P"=c(4,11),  
"rest:P"=c(5,6,7,12,13,14))  
)
```

gomez.nonnormal1 *Insecticide treatment effectiveness*

Description

Insecticide treatment effectiveness

Usage

```
data("gomez.nonnormal1")
```

Format

A data frame with 36 observations on the following 3 variables.

trt insecticidal treatment

rep replicate

larvae number of larvae

Details

Nine treatments (including the control, T9) were used on four replicates. The number of living insect larvae were recorded.

The data show signs of non-normality, and a log transform was used by Gomez.

Source

Gomez, K.A. and Gomez, A.A.. 1984, *Statistical Procedures for Agricultural Research*. Wiley-Interscience. Page 300.

Used with permission of Kwanchai Gomez.

References

None.

Examples

```

data(gomez.nonnormal1)
dat <- gomez.nonnormal1

# Gomez figure 7.3
## require(dplyr)
## dat2 <- dat %>% group_by(trt)
## dat2 <- summarize(dat2, mn=mean(larvae), rng=diff(range(larvae)))
## plot(rng ~ mn, data=dat2,
##       xlab="mean number of larvae", ylab="range of number of larvae",
##       main="gomez.nonnormal1")

# Because some of the original values are less than 10,
# the transform used is log10(x+1) instead of log10(x).
dat <- transform(dat, tlarvae=log10(larvae+1))

# QQ plots for raw/transformed data
if(require(reshape2) & require(lattice))
  qqmath( ~ value|variable, data=melt(dat),
          main="gomez.nonnormal1 - raw/transformed QQ plot",
          scales=list(relation="free"))

# Gomez table 7.16
m1 <- lm(tlarvae ~ rep + trt, data=dat)
anova(m1)
## Response: tlarvae
##           Df Sum Sq Mean Sq F value    Pr(>F)
## rep        3  0.9567  0.31889   3.6511 0.0267223 *
## trt        8  3.9823  0.49779   5.6995 0.0004092 ***
## Residuals 24  2.0961  0.08734

```

gomez.nonnormal2 *RCB trial of rice measuring white heads*

Description

RCB trial of rice measuring white heads

Usage

```
data("gomez.nonnormal2")
```

Format

A data frame with 42 observations on the following 3 variables.

gen genotype

```
rep replicate
white percentage of white heads
```

Details

The data are the percent of white heads from a rice variety trial of 14 varieties with 3 reps. Because many of the values are less than 10, the suggested data transformation is $\sqrt{x+5}$.

Source

Gomez, K.A. and Gomez, A.A.. 1984, *Statistical Procedures for Agricultural Research*. Wiley-Interscience. Page 300.

Used with permission of Kwanchai Gomez.

References

None.

Examples

```
data(gomez.nonnormal2)
dat <- gomez.nonnormal2

# Gomez suggested sqrt transform
dat <- transform(dat, twhite = sqrt(white+.5))

# QQ plots for raw/transformed data
if(require(reshape2) & require(lattice))
  qqmath( ~ value|variable, data=melt(dat),
          main="gomez.nonnormal2 - raw/transformed QQ plot",
          scales=list(relation="free"))

# Gomez anova table 7.21
m1 <- lm(twhite ~ rep + gen, data=dat)
anova(m1)
## Response: twhite2
##           Df Sum Sq Mean Sq F value    Pr(>F)
## rep         2  2.401  1.2004  1.9137  0.1678
## gen        13 48.011  3.6931  5.8877 6.366e-05 ***
## Residuals 26 16.309  0.6273
```

gomez.nonnormal3

RCB of 12 rice varieties with leafhopper survival

Description

RCB of 12 rice varieties with leafhopper survival

Usage

```
data("gomez.nonnormal3")
```

Format

A data frame with 36 observations on the following 3 variables.

gen genotype/variety of rice

rep replicate

hoppers percentage of surviving leafhoppers

Details

For each rice variety, 75 leafhoppers were caged and the percentage of surviving insects was determined.

Gomez suggest replacing 0 values by $1/(4*75)$ and replacing 100 by $1-1/(4*75)$ where 75 is the number of insects.

In effect, this means, for example, that $(1/4)$ th of an insect survived.

Because the data are percents, Gomez suggested using the arcsin transformation.

Source

Gomez, K.A. and Gomez, A.A.. 1984, *Statistical Procedures for Agricultural Research*. Wiley-Interscience. Page 307.

Used with permission of Kwanchai Gomez.

References

None.

Examples

```
data(gomez.nonnormal3)
dat <- gomez.nonnormal3

# First, replace 0, 100 values
dat$hoppers <- dat$hoppers
dat <- transform(dat, hoppers=ifelse(hoppers==0, 1/(4*75), hoppers))
dat <- transform(dat, hoppers=ifelse(hoppers==100, 100-1/(4*75), hoppers))
# Arcsin transformation of percentage p converted to degrees
# is arcsin(sqrt(p))/(pi/2)*90
dat <- transform(dat, hoppers=asin(sqrt(hoppers/100))/(pi/2)*90

# QQ plots for raw/transformed data
if(require(reshape2) & require(lattice))
  qqmath(~ value|variable, data=melt(dat),
         main="gomez.nonnormal3 - raw/transformed QQ plot",
         scales=list(relation="free"))
```

```

m1 <- lm(thoppers ~ gen, data=dat)
anova(m1) # Match Gomez table 7.25
## Response: thoppers
##           Df Sum Sq Mean Sq F value    Pr(>F)
## gen        11 16838.7 1530.79  16.502 1.316e-08 ***
## Residuals  24  2226.4   92.77

```

gomez.rice.uniformity *Uniformity trial of rice*

Description

Uniformity trial of rice in Philippines.

Format

A data frame with 648 observations on the following 3 variables.

row row

col column

yield grain yield, grams/m²

Details

An area 20 meters by 38 meters was planted to rice variety IR8. At harvest, a 1-meter border was removed around the field and discarded. Each square meter (1 meter by 1 meter) was harvested and weighed.

Field width: 18 plots x 1 m = 18 m

Field length: 38 plots x 1 m = 38 m

Source

Gomez, K.A. and Gomez, A.A.. 1984, *Statistical Procedures for Agricultural Research*. Wiley-Interscience. Page 481.

Used with permission of Kwanchai Gomez.

Examples

```

data(gomez.rice.uniformity)
dat <- gomez.rice.uniformity

if(require(desplot)){
  # Raw data plot
  desplot(yield ~ col*row, data=dat,
          aspect=38/18, # true aspect
          main="gomez.rice.uniformity")
}

```



```

}

if(require(desplot) & require(reshape2)){
  # 3x3 moving average. Gomez figure 12.1
  dmat <- melt(dat, id.var=c('col','row'))
  dmat <- acast(dmat, row~col)
  m0 <- dmat
  cx <- 2:17
  rx <- 2:35
  dmat3 <- (m0[rx+1,cx+1]+m0[rx+1,cx]+m0[rx+1,cx-1]+
            m0[rx,cx+1]+m0[rx,cx]+m0[rx,cx-1]+
            m0[rx-1,cx+1]+m0[rx-1,cx]+m0[rx-1,cx-1])/9
  dat3 <- melt(dmat3)
  desplot(value~Var2*Var1, dat3,
          aspect=38/18,
          at=c(576,637,695,753,811,870,927),
          main="gomez.rice.uniformity smoothed")
}

if(require(agricolae)){
  # Gomez table 12.4
  tab <- index.smith(dmat,
                    main="gomez.rice.uniformity",
                    col="red")$uniformity
  tab <- data.frame(tab)

  ## # Gomez figure 12.2
  ## op <- par(mar=c(5,4,4,4)+.1)
  ## m1 <- nls(Vx ~ 9041/Size^b, data=tab, start=list(b=1))
  ## plot(Vx ~ Size, tab, xlab="Plot size, m^2")
  ## lines(fitted(m1) ~ tab$Size, col='red')
  ## axis(4, at=tab$Vx, labels=tab$CV)
  ## mtext("CV", 4, line=2)
  ## par(op)
}

```

gomez.seedrate

Rice yield at six different densities

Description

Rice yield at six different densities

Format

A data frame with 24 observations on the following 3 variables.

rate kg seeds per hectare
rep rep (block), four levels
yield yield, kg/ha

Details

Rice yield at six different densities in an RCB design.

Source

Gomez, K.A. and Gomez, A.A. 1984, *Statistical Procedures for Agricultural Research*. Wiley-Interscience. Page 26.

Used with permission of Kwanchai Gomez.

Examples

```
data(gomez.seedrate)
dat <- gomez.seedrate

require(lattice)
xyplot(yield ~ rate, data=dat, group=rep, type='b',
       main="gomez.seedrate", auto.key=list(columns=4))

# Quadratic response. Use raw polynomials so we can compute optimum
m1 <- lm(yield ~ rep + poly(rate,2,raw=TRUE), dat)
-coef(m1)[5]/(2*coef(m1)[6]) # Optimum is at 29

# Plot the model predictions
if(require(latticeExtra)){
  newdat <- expand.grid(rep=levels(dat$rep), rate=seq(25,150))
  newdat$pred <- predict(m1, newdat)
  p1 <- aggregate(pred ~ rate, newdat, mean) # average reps
  xyplot(yield ~ rate, data=dat, group=rep, type='b',
        main="gomez.seedrate (with model predictions)", auto.key=list(columns=4)) +
  xyplot(pred ~ rate, p1, type='l', col='black', lwd=2)
}
```

gomez.splitplot.subsample

Height of rice in a split-plot experiment with subsamples

Description

Height of rice in a split-plot experiment with subsamples

Format

A data frame with 186 observations on the following 5 variables.

time time factor, T1-T4

manage management, M1-M6

```
rep rep/block, R1-R3
sample subsample, S1-S2
height plant height (cm)
```

Details

A split-plot experiment in three blocks. Whole-plot is 'management', sub-plot is 'time' of application, with two subsamples. The data are the heights, measured on two single-hill sampling units in each plot.

Source

Gomez, K.A. and Gomez, A.A.. 1984, *Statistical Procedures for Agricultural Research*. Wiley-Interscience. Page 481.

Used with permission of Kwanchai Gomez.

Examples

```
data(gomez.splitplot.subsample)
dat <- gomez.splitplot.subsample
if(require(HH)){
  interaction2wt(height ~ rep + time + manage, data=dat,
                x.between=0, y.between=0,
                main="gomez.splitplot.subsample - plant height")
}

# Management totals, Gomez table 6.8
tapply(dat$height, dat$manage, sum)

# Gomez table 6.11 analysis of variance
m1 <- aov(height ~ rep + manage + time + manage:time +
          Error(rep/manage/time), data=dat)
summary(m1)
## Error: rep
##      Df Sum Sq Mean Sq
## rep  2  2632    1316

## Error: rep:manage
##           Df Sum Sq Mean Sq F value Pr(>F)
## manage    7  1482  211.77  2.239 0.0944 .
## Residuals 14  1324   94.59
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Error: rep:manage:time
##           Df Sum Sq Mean Sq F value Pr(>F)
## time        3  820.8  273.61  7.945 0.000211 ***
## manage:time 21  475.3   22.63  0.657 0.851793
## Residuals  48 1653.1   34.44
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 96  167.4    1.744
```

```
gomez.splitsplit      Rice split-split plot experiment
```

Description

Grain yield of three varieties of rice grown in a split-split plot arrangement with 3 reps, nitrogen level as the main plot, management practice as the sub-plot, and rice variety as the sub-sub plot.

Format

A data frame with 135 observations on the following 7 variables.

```
rep  block, 3 levels
nitro nitrogen fertilizer, in kilograms/hectare
management plot management
gen  genotype/variety of rice
yield yield
col  column position in the field
row  row position in the field
```

Source

Gomez, K.A. and Gomez, A.A.. 1984, *Statistical Procedures for Agricultural Research*. Wiley-Interscience. Page 143.

Used with permission of Kwanchai Gomez.

Examples

```
data(gomez.splitsplit)
dat <- gomez.splitsplit
dat$nf <- factor(dat$nitro)

if(require(desplot)){
  desplot(nf ~ col*row, dat,
          # aspect unknown
          out1=rep, col=management, num=gen, cex=1,
          main="gomez.splitsplit")
  desplot(yield ~ col*row, dat,
          # aspect unknown
```

```

        out1=rep, main="gomez.splitsplit")
    }

    if(require(HH)){
    position(dat$nf) <- c(0,50,80,110,140)
    interaction2wt(yield~rep+nf+management+gen, data=dat,
                  main="gomez.splitsplit",
                  x.between=0, y.between=0,
                  relation=list(x="free", y="same"),
                  rot=c(90,0), xlab="",
                  par.strip.text.input=list(cex=.7))
    }

    # AOV. Gomez page 144-153
    m0 <- aov(yield~ nf * management * gen + Error(rep/nf/management),
             data=dat)
    summary(m0) # Similar to Gomez, p. 153.

```

gomez.stripplot

Rice strip-plot experiment

Description

A strip-plot experiment with three reps, variety as the horizontal strip and nitrogen fertilizer as the vertical strip.

Format

yield Grain yield in kg/ha
rep Rep
nitro Nitrogen fertilizer in kg/ha
gen Rice variety
col column
row row

Details

Note, this is a subset of the the 'gomez.stripsplitplot' data.

Source

Gomez, K.A. and Gomez, A.A.. 1984, *Statistical Procedures for Agricultural Research*. Wiley-Interscience. Page 110.

Used with permission of Kwanchai Gomez.

References

Jan Gertheiss (2014). ANOVA for Factors With Ordered Levels. *J Agric Biological Environmental Stat*, 19, 258-277.

Examples

```

data(gomez.stripplot)
dat <- gomez.stripplot

# Gomez figure 3.7
if(require(desplot)){
  desplot(gen ~ col*row, data=dat,
          # aspect unknown
          out1=rep, out2=nitro, num=nitro, cex=1,
          main="gomez.stripplot")
}

# Gertheiss figure 1
# library(lattice)
# dotplot(factor(nitro) ~ yield|gen, data=dat)

# Gomez table 3.12
tapply(dat$yield, dat$rep, sum)
tapply(dat$yield, dat$gen, sum)
tapply(dat$yield, dat$nitro, sum)

# Gomez table 3.15. Anova table for strip-plot
dat <- transform(dat, nf=factor(nitro))
m1 <- aov(yield ~ gen * nf + Error(rep + rep:gen + rep:nf), data=dat)
summary(m1)
## Error: rep
##           Df  Sum Sq Mean Sq F value Pr(>F)
## Residuals  2 9220962 4610481

## Error: rep:gen
##           Df  Sum Sq Mean Sq F value Pr(>F)
## gen         5 57100201 11420040  7.653 0.00337 **
## Residuals 10 14922619  1492262
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Error: rep:nf
##           Df  Sum Sq Mean Sq F value Pr(>F)
## nf          2 50676061 25338031 34.07 0.00307 **
## Residuals  4 2974908  743727
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Error: Within
##           Df  Sum Sq Mean Sq F value Pr(>F)
## gen:nf      10 23877979 2387798  5.801 0.000427 ***

```

```
## Residuals 20 8232917 411646

# More compact view
## library(agricolae)
## with(dat, strip.plot(rep, nf, gen, yield))

## Analysis of Variance Table
## Response: yield
##      Df  Sum Sq Mean Sq F value    Pr(>F)
## rep    2  9220962  4610481  11.2001 0.0005453 ***
## nf     2  50676061 25338031  34.0690 0.0030746 **
## Ea     4   2974908   743727   1.8067 0.1671590
## gen     5  57100201 11420040   7.6528 0.0033722 **
## Eb    10  14922619  1492262   3.6251 0.0068604 **
## gen:nf 10 23877979  2387798   5.8006 0.0004271 ***
## Ec     20  8232917   411646

# Mixed-model version
## library(lme4)
## m3 <- lmer(yield ~ gen * nf + (1|rep) + (1|rep:nf) + (1|rep:gen), data=dat)
## anova(m3)

## Analysis of Variance Table
##      Df  Sum Sq Mean Sq F value
## gen     5 15751300  3150260  7.6528
## nf     2 28048730 14024365 34.0690
## gen:nf 10 23877979  2387798  5.8006
```

gomez.stripsplitplot *Rice strip-split-plot experiment*

Description

A strip-split-plot experiment with three reps, genotype as the horizontal strip, nitrogen fertilizer as the vertical strip, and planting method as the subplot factor.

Format

yield grain yield in kg/ha
 planting planting factor, P1=broadcast, P2=transplanted
 rep rep, 3 levels
 nitro nitrogen fertilizer, kg/ha
 gen genotype, G1 to G6
 col column
 row row

Details

Note, this is a superset of the the 'gomez.stripplot' data.

Source

Gomez, K.A. and Gomez, A.A.. 1984, *Statistical Procedures for Agricultural Research*. Wiley-Interscience. Page 155.

Used with permission of Kwanchai Gomez.

Examples

```
data(gomez.stripsplitplot)
dat <- gomez.stripsplitplot

# Layout
if(require(desplot)){
  desplot(gen ~ col*row, dat,
          out1=rep, col=nitro, text=planting, cex=1,
          main="gomez.stripsplitplot")
}

# Gomez table 4.19, ANOVA of strip-split-plot design
dat <- transform(dat, nf=factor(nitro))
m1 <- aov(yield ~ nf * gen * planting +
          Error(rep + rep:nf + rep:gen + rep:nf:gen), data=dat)
summary(m1)

## Not run:
# There is a noticeable linear trend along the y coordinate which may be
# an artifact that blocking will remove, or may need to be modeled.
# Note the outside values in the high-nitro boxplot.
require("HH")
  interaction2wt(yield ~ nitro + gen + planting + row, dat,
                x.between=0, y.between=0,
                x.relation="free")

## End(Not run)
```

gomez.wetdry

Rice yield in wet & dry seasons with nitrogen fertilizer treatments

Description

Rice yield in wet & dry seasons with nitrogen fertilizer treatments

Format

A data frame with 96 observations on the following 4 variables.

```
season season = wet/dry
nitrogen nitrogen fertilizer kg/ha
rep replicate
yield grain yield, t/ha
```

Details

Five nitrogen fertilizer treatments were tested in 2 seasons using 3 reps.

Source

Gomez, K.A. and Gomez, A.A.. 1984, *Statistical Procedures for Agricultural Research*. Wiley-Interscience. Page 318.

Used with permission of Kwanchai Gomez.

References

Rong-Cai Yang, Patricia Juskiw. (2011). Analysis of covariance in agronomy and crop research. *Canadian Journal of Plant Science*, 91:621-641. <http://doi.org/10.4141/cjps2010-032>

Examples

```
data(gomez.wetdry)
dat <- gomez.wetdry

require(lattice)
foo1 <- xyplot(yield ~ nitrogen|season, data=dat,
              group=rep,type='l',auto.key=list(columns=3),
              ylab="yield in each season",
              main="gomez.wetdry raw data & model")

# Yang & Juskiw fit a quadratic model with linear and quadratic
# contrasts using non-equal intervals of nitrogen levels.
# This example below omits the tedious contrasts

if(require(latticeExtra) & require(lme4)){
  m1 <- lmer(yield ~ season*poly(nitrogen, 2) + (1|season:rep), data=dat)
  pdat <- expand.grid(season=c('dry','wet'),
                    nitrogen=seq(from=0,to=150,by=5))
  pdat$pred <- predict(m1, newdata=pdat, re.form= ~ 0)
  foo1 +
    xyplot(pred ~ nitrogen|season, data=pdat, type='l',lwd=2,col="black")
} else {
  foo1
}
```

```

# m2 <-lmer(yield ~ poly(nitrogen, 2) + (1|season:rep), data=dat)
# anova(m1,m2)
## m2: yield ~ poly(nitrogen, 2) + (1 | season:rep)
## m1: yield ~ season * poly(nitrogen, 2) + (1 | season:rep)
##      Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
## m2  5  86.418  93.424 -38.209   76.418
## m1  8  64.216  75.425 -24.108   48.216 28.202    3 3.295e-06 ***

```

gotway.hessianfly *Hessian fly damage to wheat varieties*

Description

Hessian fly damage to wheat varieties

Format

block block factor, 4 levels
genotype factor, 16 wheat varieties
lat latitude, numeric
long longitude, numeric
y number of damaged plants
n number of total plants

Details

The response is binomial.

Each plot was square.

Source

C. A. Gotway and W. W. Stroup. A Generalized Linear Model Approach to Spatial Data Analysis and Prediction *Journal of Agricultural, Biological, and Environmental Statistics*, 2, 157-178.

<http://doi.org/10.2307/1400401>

References

The GLIMMIX procedure. <http://www.ats.ucla.edu/stat/SAS/glimmix.pdf>

Examples

```

## FIXME Could the spaMM package be useful here...????

data(gotway.hessianfly)
dat <- gotway.hessianfly

dat$prop <- dat$y / dat$n
if(require(desplot)){
  desplot(prop~long*lat, dat,
           aspect=1, # true aspect
           out1=block, text=gen, cex=1, shorten='no',
           main="gotway.hessianfly")
}

# -----

# Block random. See Glimmix manual, output 1.18.
# Note: (Different parameterization)
## Not run:
## require(lme4)
## l2 <- glmer(cbind(y, n-y) ~ gen + (1|block), data=dat, family=binomial,
##            control=glmerControl(check.nlev.gtr.1="ignore"))
## coef(l2)

## End(Not run)

# -----

## Not run:
# asreml3
require(asreml)
a2 <- asreml(prop ~ gen, data=dat, random=~block,
             family=asreml.binomial(),
             weights=n)

coef(a2)

## End(Not run)

# -----

## Not run:
## require(asreml4)
## a2 <- asreml(prop ~ gen, data=dat, random=~block,
##              family=asr_binomial(),
##              weights=n)
## coef(a2)

## End(Not run)

```

goulden.barley.uniformity

Uniformity trial of barley

Description

Uniformity trial of barley

Format

A data frame with 400 observations on the following 3 variables.

row row

col column

yield yield, grams per plot

Details

Yield in grams for 400 square-yard barley plots.

Field length: 20 plots x 1 yd = 20 yd

Field width: 20 plots x 1 yd = 20 yd

Source

C. H. Goulden, (1939). *Methods of statistical analysis*, 1st ed. Page 18. <https://archive.org/stream/methodsofstatist031744mb>

References

Erwin LeClerg, Warren Leonard, Andrew Clark (1962). *Field Plot Technique*, 2nd ed. Page 27.

Alternatively, page 39 in 1939 edition: <https://archive.org/stream/fieldplottechniq00leon>

Examples

```
data(goulden.barley.uniformity)
dat <- goulden.barley.uniformity

if(require(desplot)){
  desplot(yield ~ col*row, data=dat,
          aspect=20/20, # true aspect
          main="goulden.barley.uniformity")
}

# See LeClerg, Leonard, Clark
hist(dat$yield, main="goulden.barley.uniformity",
      breaks=c(21,40,59,78,97,116,135,154,173,192,211,230,249)+.5)
```

`goulden.eggs`*Sample of egg weights on 24 consecutive days*

Description

Sample of egg weights on 24 consecutive days

Usage

```
data("goulden.eggs")
```

Format

A data frame with 240 observations on the following 2 variables.

```
day day
weight weight
```

Details

Data are the weights of 10 eggs taken at random on each day for 24 days. Day 1 was Dec 10, and Day 24 was Jan 2.

The control chart for standard deviations shows 4 values beyond the upper limits. The data reveals a single, unusually large egg on each of these days. These are almost surely double-yolk eggs.

Source

Cyrus H. Goulden (1952). *Methods of Statistical Analysis*, 2nd ed. Page 425. <http://krishikosh.egranth.ac.in/handle/1/203411>

References

None.

Examples

```
data(goulden.eggs)
dat <- goulden.eggs

if(require(qicharts)){
  # Figure 19-4 of Goulden. (Goulden uses 1/n when calculating std dev)
  op <- par(mfrow=c(2,1))
  qic(weight, x = day, data = dat, chart = 'xbar',
       main = 'goulden.eggs - Xbar chart',
       xlab = 'Date', ylab = 'Avg egg weight' )
  qic(weight, x = day, data = dat, chart = 's',
       main = 'goulden.eggs - S chart',
       xlab = 'Date', ylab = 'Std dev egg weight' )
  par(op)
}
```

`goulden.latin`*Latin square experiment for testing fungicide*

Description

Latin square experiment for testing fungicide

Usage

```
data("goulden.latin")
```

Format

A data frame with 25 observations on the following 4 variables.

`trt` treatment factor, 5 levels

`yield` yield

`row` row

`col` column

Details

Five treatments were tested to control stem rust in wheat. Treatment codes and descriptions: A = Dusted before rains. B = Dusted after rains. C = Dusted once each week. D = Drifting, once each week. E = Not dusted.

Source

Cyrus H. Goulden (1952). *Methods of Statistical Analysis*, 2nd ed. Page 216. <http://krishikosh.egranth.ac.in/handle/1/203411>

Examples

```
data(goulden.latin)
dat <- goulden.latin

if(require(desplot)){
  desplot(yield ~ col*row, dat,
          text=trt, cex=1, # aspect unknown
          main="goulden.latin")
}

# Matches Goulden.
m1 <- lm(yield~ trt + factor(row) + factor(col), data=dat)
anova(m1)
```

goulden.splitsplit *Split-split plot experiment of wheat*

Description

Split-split plot experiment of wheat

Usage

```
data("goulden.splitsplit")
```

Format

A data frame with 160 observations on the following 9 variables.

row row

col column

yield yield

inoc inoculate

trt treatment number

gen genotype

dry dry/wet dust application

dust dust treatment

block block

Details

An interesting split-split plot experiment in which the sub-plot treatments have a 2*5 factorial structure.

An experiment was conducted in 1932 on the experimental field of the Dominion Rust Research Laboratory. The study was designed to determine the effect on the incidence of root rot, of variety of wheat, kinds of dust for seed treatment, method of application of the dust, and efficacy of soil inoculation with the root-rot organism.

The field had 4 blocks.

Each block has 2 whole plots for the genotypes.

Each whole-plot had 10 sub-plots for the 5 different kinds of dust and 2 methods of application.

Each sub-plot had 2 sub-sub-plots, one for inoculated soil and the other one for uninoculated soil.

Source

C. H. Goulden, (1939). *Methods of statistical analysis*, 1st ed. Page 18. <https://archive.org/stream/methodsofstatist031744mb>

References

None

Examples

```

data(goulden.splitsplit)
dat <- goulden.splitsplit

if(require(desplot)) {

  ## Experiment design. Goulden p. 152-153
  ## desplot(gen ~ col*row, data=dat,
  ##       out1=block, out2=trt, text=dust, col=inoc, cex=1,
  ##       main="goulden.splitsplit")

  desplot(yield ~ col*row, data=dat,
          out1=block, out2=gen, col=inoc, num=trt, cex=1,
          main="goulden.splitsplit")

}

# Match Goulden table 40
m1 <- aov(yield ~ gen
          + dust + dry + dust:dry + gen:dust + gen:dry + gen:dust:dry
          + inoc + inoc:gen + inoc:dust + inoc:dry
          + inoc:dust:dry + inoc:gen:dust + inoc:gen:dry
          + Error(block/(gen+gen:dust:dry+gen:inoc:dry)), data=dat)
summary(m1)

```

graybill.heteroskedastic

Multi-environment trial of wheat varieties with heteroskedastic yields

Description

Wheat varieties with heteroskedastic yields

Format

A data frame with 52 observations on the following 3 variables.

env environment, 13 levels

gen genotype, 4 levels

yield yield

Details

Yield of 4 varieties of wheat at 13 locations in Oklahoma, USA.

The data was used to explore variability between varieties.

Source

F. A. Graybill, 1954. Variance heterogeneity in a randomized block design, *Biometrics*, 10, 516-520.

References

Hans-Pieter Piepho, 1994. Missing observations in the analysis of stability. *Heredity*, 72, 141–145. <http://doi.org/10.1038/hdy.1994.20>

Examples

```
data(graybill.heteroskedastic)
dat <- graybill.heteroskedastic

# Genotypes are obviously not homoscedastic
boxplot(yield ~ gen, dat, main="graybill.heteroskedastic")

# Shukla stability variance of each genotype, same as Grubbs' estimate
# Matches Piepho 1994 page 143.
# Do not do this! Nowadays, use mixed models instead.
if(require("reshape2")){
  datm <- acast(dat, gen~env)
  w <- datm
  w <- sweep(w, 1, rowMeans(datm))
  w <- sweep(w, 2, colMeans(datm))
  w <- w + mean(datm)
  w <- rowSums(w^2)
  k=4; n=13
  sig2 <- k*w/((k-2)*(n-1)) - sum(w)/((k-1)*(k-2)*(n-1))
  ## sig2
  ##      G1      G2      G3      G4
  ## 145.98 -14.14  75.15  18.25

var.shukla <- function(x,N){
  # Estimate variance of shukla stability statistics
  # Piepho 1994 equation (5)
  K <- length(x) # num genotypes
  S <- outer(x,x)
  S1 <- diag(S)
  S2 <- rowSums(S) - S1
  S[!upper.tri(S)] <- 0 # Make S upper triangular
  # The ith element of S3 is the sum of the upper triangular elements of S,
  # excluding the ith row and ith column
  S3 <- sum(S) - rowSums(S) - colSums(S)
  var.si2 <- 2*S1/(N-1) + 4/( (N-1)*(K-1)^2 ) * ( S2 + S3/(K-2)^2 )
}
```

```

    return(var.si2)
  }

  # Set negative estimates to zero
  sig2[sig2<0] <- 0

  # Variance of shukla stat. Match Piepho 1994, table 5, example 1
  var.shukla(sig2,13)
  ##      G1      G2      G3      G4
  ## 4069.3296 138.9424 1423.0797 306.5270

}

```

gregory.cotton

Factorial experiment of cotton in Sudan.

Description

Factorial experiment of cotton in Sudan.

Usage

```
data("gregory.cotton")
```

Format

A data frame with 144 observations on the following 6 variables.

yield a numeric vector
 year year
 nitrogen nitrogen level
 date sowing date
 water irrigation amount
 spacing spacing between plants

Details

Experiment conducted in Sudan at the Gezira Research Farm in 1929-1930 and 1930-1931. The effects on yield of four factors was studied in all possible combinations.

Sowing dates in 1929: D1 = Jul 24, D2 = Aug 11, D3 = Sep 2, D4 = Sep 25.

Spacing: S1 = 25 cm between holes, S2 = 50 cm, S3 = 75 cm. The usual spacing is 50-70 cm.

Irrigation: I1 = Light, I2 = Medium, I3 = Heavy.

Nitrogen: N0 = None/Control, N1 = 600 rotls/feddan.

In each year there were $4 \times 3 \times 2 \times 2 = 72$ treatments, each replicated four times. The means are given here.

Gregory (1932) has two interesting graphics: 1. radial bar plot 2. photographs of 3D model of treatment means.

Source

Gregory, FG and Crowther, F and Lambert, AR (1932). The interrelation of factors controlling the production of cotton under irrigation in the Sudan. *The Journal of Agricultural Science*, 22, 617-638. Table 1, 10. <https://doi.org/10.1017/S0021859600054137>

References

Paterson, D. Statistical Technique in Agricultural Research, p. 211.

Examples

```
data(gregory.cotton)
dat <- gregory.cotton

if(require(dplyr)){
  # Main effect means, Gregory table 2
  dat
  dat
  dat
  dat
}

# Figure 2 of Gregory. Not recommended, but an interesting exercise.
# http://stackoverflow.com/questions/13887365
if(require(ggplot2)){
  d1 <- subset(dat, year=="Y1")
  d1 <- transform(d1, grp=factor(paste(date,nitrogen,water,spacing)))
  d1 <- d1[order(d1$grp),] # for angles
  # Rotate labels on the left half 180 deg. First 18, last 18 labels
  d1$ang <- 90+seq(from=(360/nrow(d1))/1.5, to=(1.5*(360/nrow(d1)))-360,
                  length.out=nrow(d1))+80
  d1$ang[1:18] <- d1$ang[1:18] + 180
  d1$ang[55:72] <- d1$ang[55:72] + 180
  # Labels on left half to right-adjusted
  d1$hjust <- 0
  d1$hjust[1:18] <- d1$hjust[55:72] <- 1

  gg <- ggplot(d1, aes(x=grp,y=yield,fill=factor(spacing))) +
    geom_col() +
    guides(fill=FALSE) + # no legend for 'spacing'
    coord_polar(start=-pi/2) + # default is to start at top
    labs(title="gregory.cotton 1929",x="",y="",label="") +
    # The bar columns are centered on 1:72, subtract 0.5 to add radial axes
    geom_vline(xintercept = seq(1, 72, by=3)-0.5, color="gray", size=.25) +
    geom_vline(xintercept = seq(1, 72, by=18)-0.5, size=1) +
```

```

geom_vline(xintercept = seq(1, 72, by=9)-0.5, size=.5) +
geom_hline(yintercept=c(1,2,3)) +
geom_text(data=d1, aes(x=grp, y=max(yield), label=grp, angle=ang, hjust=hjust),
          size=2) +
theme(panel.background=element_blank(),
      axis.title=element_blank(),
      panel.grid=element_blank(),
      axis.text.x=element_blank(),
      axis.text.y=element_blank(),
      axis.ticks=element_blank() )
print(gg)
}

```

gumpertz.pepper

*Phytophthora disease incidence in a pepper field***Description**

Phytophthora disease incidence in a pepper field

Format

A data frame with 800 observations on the following 6 variables.

field field factor, 2 levels

row x ordinate

quadrat y ordinate

disease presence (Y) or absence (N) of disease

water soil moisture percent

leaf leaf assay count

Details

Each field is 20 rows by 20 quadrates, with 2 to 3 bell pepper plants per plot. If any plant was wilted, dead, or had lesions, the Phytophthora disease was considered to be present in the plot. The soil pathogen load was assayed as the number of leaf disks colonized out of five. In field 2, the pattern of disease presence appears to follow soil water content. In field 1, no obvious trends were present.

Gumpertz et al. model the presence of disease using soil moisture and leaf assay as covariates, and using disease presence of neighboring plots as covariates in an autologistic model.

Source

Marcia L. Gumpertz; Jonathan M. Graham; Jean B. Ristaino (1997). Autologistic Model of Spatial Pattern of Phytophthora Epidemic in Bell Pepper: Effects of Soil Variables on Disease Presence. *Journal of Agricultural, Biological, and Environmental Statistics*, Vol. 2, No. 2., pp. 131-156.

Used with permission of Marcia Gumpertz. Research funded by USDA.

Examples

```

data(gumpertz.pepper)
dat <- gumpertz.pepper

# Gumpertz deletes two outliers
dat[ dat$field == "F1" & dat$row==20 & dat$quadrat==10, 'water'] <- NA
dat[ dat$field == "F2" & dat$row==5 & dat$quadrat==4, 'water'] <- NA

# Horizontal flip
dat <- transform(dat, row=21-row)

# Disease presence. Gumpertz fig 1a, 2a.
if(require(desplot)){
  grays <- colorRampPalette(c("#d9d9d9", "#252525"))
  desplot(disease ~ row*quadrat|field, data=dat,
          col.regions=c('white', 'black'), aspect=1, # uncertain aspect
          main="gumpertz.pepper disease presence", )

# Soil water. Gumpertz fig 1b, 2b
  desplot(water ~ row*quadrat|field, data=dat,
          col.regions=grays(5), aspect=1, # uncertain aspect
          at=c(5,7.5,10,12.5,15,18),
          main="gumpertz.pepper soil moisture")

# Leaf assay. Gumpertz fig 1c, 2c
  desplot(leaf ~ row*quadrat|field, data=dat,
          col.regions=grays(6),
          at=c(0,1,2,3,4,5,6)-.5, aspect=1, # uncertain aspect
          main="gumpertz.pepper leaf assay", )
}

# Use the inner 16x16 grid of plots in field 2
dat2 <- droplevels(subset(dat, field=="F2" & !is.na(water) &
                          row > 2 & row < 19 & quadrat > 2 & quadrat < 19))

m21 <- glm(disease ~ water + leaf, data=dat2, family=binomial)
coef(m21) # These match Gumpertz et al table 4, model 1
## (Intercept)      water      leaf
## -9.1019623    0.7059993    0.4603931
dat2$res21 <- resid(m21)
if(FALSE & require(desplot)){
  desplot(res21 ~ row*quadrat, data=dat2,
          main="gumpertz.pepper field 2, model 1 residuals")
# Still shows obvious trends. Gumpertz et al add spatial covariates for
# neighboring plots, but with only minor improvement in misclassification
}

```

hanks.sprinkler

Wheat yields in a line-source sprinkler experiment

Description

Three wheat varieties planted in 3 blocks, with a line sprinkler crossing all whole plots.

Format

A data frame with 108 observations on the following 7 variables.

block block

row row

subplot column

gen genotype, 3 levels

yield yield (tons/ha)

irr irrigation level, 1..6

dir direction from sprinkler, N/S

Details

A line-source sprinkler is placed through the middle of the experiment (between subplots 6 and 7). Subplots closest to the sprinkler receive the most irrigation. Subplots far from the sprinkler (near the edges) have the lowest yields.

One data value was modified from the original (following the example of other authors).

Source

Hanks, R.J., Sisson, D.V., Hurst, R.L., and Hubbard K.G. (1980). Statistical Analysis of Results from Irrigation Experiments Using the Line-Source Sprinkler System. *Soil Science Society of America Journal*, 44, 886-888. <http://doi.org/10.2136/sssaj1980.03615995004400040048x>

References

Johnson, D. E., Chaudhuri, U. N., and Kanemasu, E. T. (1983). Statistical Analysis of Line-Source Sprinkler Irrigation Experiments and Other Nonrandomized Experiments Using Multivariate Methods. *Soil Science Society American Journal*, 47, 309-312.

Stroup, W. W. (1989). Use of Mixed Model Procedure to Analyze Spatially Correlated Data: An Example Applied to a Line-Source Sprinkler Irrigation Experiment. *Applications of Mixed Models in Agriculture and Related Disciplines, Southern Cooperative Series Bulletin No. 343*, 104-122.

SAS Stat User's Guide. http://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_mixed

Examples

```

data(hanks.sprinkler)
dat <- hanks.sprinkler

# The line sprinkler is vertical between subplots 6 & 7
if(require(desplot)){
  desplot(yield~subplot*row, dat,
          out1=block, out2=irr, cex=1, # aspect unknown
          num=gen, main="hanks.sprinkler")
}

require(lattice)
xyplot(yield~subplot|block, dat, type=c('b'), group=gen,
       layout=c(1,3), auto.key=TRUE,
       main="hanks.sprinkler",
       panel=function(x,y,...){
         panel.xyplot(x,y,...)
         panel.abline(v=6.5, col='wheat')
       })

# -----

## This is the model from the SAS documentation
## proc mixed;
## class block gen dir irr;
## model yield = gen|dir|irr@2;
## random block block*dir block*irr;
## repeated / type=toep(4) sub=block*gen r;

# -----

## Not run:
# asreml3
require(asreml)

dat <- transform(dat, subf=factor(subplot),
                 irrf=factor(irr))
dat <- dat[order(dat$block, dat$gen, dat$subplot),]

m1 <- asreml(yield ~ gen + dir + irrf + gen:dir + gen:irrf + dir:irrf,
            data=dat,
            random= ~ block + block:dir + block:irrf,
            rcov= ~ block:gen:corb(subf, k=3))

require(lucid)
vc(m1)
##           effect component std.error z.ratio constr
##   block!block.var  0.2194    0.2393    0.92   pos
##   block:dir!block.var 0.01768  0.03154  0.56   pos
##   block:irrf!block.var 0.03539  0.03617  0.98   pos
##           R!variance  0.285     0.05086  5.6    pos

```

```

##          R!cor1  0.02802    0.1143    0.25  uncon
##          R!cor2  0.005095   0.1278    0.04  uncon
##          R!cor3 -0.3246     0.0905   -3.6   uncon

## # convert asreml correlations to SAS covariances
## round(.2850 * c(1, .02802, .005095, -.3246),4) # res var * (cor1, cor2, cor3)
## [1] 0.2850 0.0080 0.0015 -0.0925

## End(Not run)

# -----

## Not run:
## require(asreml4)

## dat <- transform(dat, subf=factor(subplot),
##                  irrf=factor(irrf))
## dat <- dat[order(dat$block, dat$gen, dat$subplot),]

## # FIXME asreml4
## m1 <- asreml(yield ~ gen + dir + irrf + gen:dir + gen:irrf + dir:irrf,
##             data=dat,
##             random= ~ block + block:dir + block:irrf,
##             resid = ~ block:gen:corb(subf, b=3))

## require(lucid)
## vc(m1)
## ##          effect component std.error z.ratio constr
## ##          block!block.var  0.2194    0.2393    0.92   pos
## ##          block:dir!block.var  0.01768  0.03154    0.56   pos
## ##          block:irrf!block.var  0.03539  0.03617    0.98   pos
## ##          R!variance  0.285    0.05086    5.6   pos
## ##          R!cor1  0.02802    0.1143    0.25  uncon
## ##          R!cor2  0.005095   0.1278    0.04  uncon
## ##          R!cor3 -0.3246     0.0905   -3.6   uncon

## ## # convert asreml correlations to SAS covariances
## ## round(.2850 * c(1, .02802, .005095, -.3246),4) # res var * (cor1, cor2, cor3)
## ## [1] 0.2850 0.0080 0.0015 -0.0925

## detach(package:asreml4)

## End(Not run)

```

hanover.whitepine

Mating crosses of white pine trees

Description

Mating crosses of white pine trees

Usage

```
data("hanover.whitepine")
```

Format

A data frame with 112 observations on the following 4 variables.

```
rep replicate
female female parent
male male parent
length epicotyl length, cm
```

Details

Four male (pollen parent) White Pine trees were mated to seven female trees and 2654 progeny were grown in four replications, one plot per mating in each replication. Parent trees were sourced from Idaho, USA. The data are plot means of epicotyl length.

Becker (1984) used these data to demonstrate the calculation of heritability.

Source

Hanover, James W and Barnes, Burton V. (1962). Heritability of height growth in year-old western white pine. *Proc Forest Genet Workshop*. 22, 71–76.

Data prepared electronically by Kevin Wright from:

Walter A. Becker (1984). *Manual of Quantitative Genetics*, 4th ed. Page 83.

References

None.

Examples

```
data(hanover.whitepine)
dat <- hanover.whitepine

require(lattice)
# Relatively high male-female interaction in growth comared
# to additive gene action. Response is more consistent within
# male progeny than female progeny.
# with(dat, interaction.plot(female, male, length))
# with(dat, interaction.plot(male, female, length))
bwplot(length ~ male|female, data=dat,
        main="hanover.whitepine - length for male:female crosses",
        xlab="Male parent", ylab="Epicotyl length")

# Progeny sums match Becker p 83
sum(dat$length) # 380.58
aggregate(length ~ female + male, data=dat, FUN=sum)
```

```

# Sum of squares matches Becker p 85
m1 <- aov(length ~ rep + male + female + male:female, data=dat)
anova(m1)

## Not run:
# Variance components match Becker p. 85
require(lme4)
require(lucid)
m2 <- lmer(length ~ (1|rep) + (1|male) + (1|female) + (1|male:female), data=dat)
#as.data.frame(lme4::VarCorr(m2))
vc(m2)
##          grp          var1 var2   vcov  sdcor
## male:female (Intercept) <NA> 0.1369 0.3699
##          female (Intercept) <NA> 0.02094 0.1447
##            male (Intercept) <NA> 0.1204 0.3469
##              rep (Intercept) <NA> 0.01453 0.1205
##      Residual          <NA> <NA> 0.2004 0.4477

# Becker used this value for variability between individuals, within plot
s2w <- 1.109

# Calculating heritability for individual trees
s2m <- .120
s2f <- .0209
s2mf <- .137
vp <- s2m + s2f + s2mf + s2w # variability of phenotypes = 1.3869
4*s2m / vp # heritability male 0.346
4*s2f / vp # heritability female 0.06
2*(s2m+s2f)/vp # heritability male+female .203
# As shown in the boxplot, heritability is stronger through the
# males than through the females.

## End(Not run)

```

harris.multi.uniformity

Uniformity trials with multiple crops, 15 years on the same land

Description

Uniformity trials with multiple crops, at Huntley Field Station, Montana, 1911-1925.

Format

A data frame with 1058 observations on the following 5 variables.

series series (field coordinate)

plot plot number (field coordinate)

year year, 1911-1925

crop crop factor

yield yield, numeric

Details

Field width: 2 plots * 317 ft + 5 feet alley = 639 feet

Field length: 23 plots * 23.3 feet = 536 feet

All yields are given in pound per plot, except sugar beets, which are given in tons per acre. Harris (1928) shows a map of the location.

Harris 1920: In the spring of 1911 this field was laid out into 46 plots, each measuring 23.5 by 317 feet and containing 0.17 acre, arranged in two parallel series of 23 plots each. The two series of plots were separated merely by a temporary irrigation ditch. In 1911 it was planted to sugar beets, and in the spring of 1912 it was seeded to alfalfa, and one cutting was harvested that year. This stand remained on the ground during 1913 and 1914, when the entire field was fall-plowed. In 1913 three cuttings were made, but the third cutting was lost in a heavy wind which scattered and mixed the crop before weighings from the various plots could be made. The first cutting, designated as alfalfa I, was made on plots one-half the original size. The second cutting was harvested from plots one-quarter the original size. The first and second cuttings in 1914 were weighed for plots one-quarter the original size—that is, 0.0425-acre plots—while the third cutting was recorded for plots one-third the original size. These furnish the data for alfalfa I, II, and III for 1914. Total yields for the first and second cuttings in 1913 and 1914 and for the first, second, and third cuttings in 1914 are also considered. In 1915 and 1916 ear corn was grown. In 1917 the fields were planted to oats, and records were made of grain, straw, and total yield. In 1918 silage corn was grown. In 1919 the land produced a crop of barley.

Harris 1928: The southeast corner of Series II, the east series, is about 80 feet from the main canal, and the southwest corner of Series III is about 50 feet from Ouster Coulee. The main project canal carries normally during the irrigation season about 400 second-feet of water. The water surface in the canal is about 4 feet above the high corner of the field. It is evident from surface conditions, as well as from borings made between the canal and the field, that there is extensive seepage from the canal into the subsoil of the field. The volume of this seepage has been larger in recent years than it was in the earlier years of the cropping experiments, probably because the canal bank has been worn away by internal erosion, exposing a stratum of sandy subsoil that underlies the canal and part of the field.

Whereas in the earlier crops Series II was better for alfalfa, Series III was better for alfalfa in the later period. The writers feel inclined to suggest that in the earlier experiments the height of the water table had no harmful effect upon a deep-rooted crop such as alfalfa. It is quite possible that during drier periods the higher water table actually favored alfalfa growth on Series II. The higher water tables of recent years have probably had a deleterious influence, which has been especially marked on Series II, where the water apparently comes nearer to the surface than in Series III.

Source

Harris, J Arthur and Scofield, CS. (1920). Permanence of differences in the plots of an experimental field. *Jour. Agr. Res*, 20, 335-356. <http://naldc.nal.usda.gov/catalog/IND43966236>

Harris, J Arthur and Scofield, CS. (1928). Further studies on the permanence of differences in the plots of an experimental field. *Jour. Agr. Res*, 36, 15-40. <http://naldc.nal.usda.gov/catalog/IND43967538>

Examples

```

data(harris.multi.uniformity)
dat <- harris.multi.uniformity
# Combine year/crop into 'harvest'
dat <- transform(dat, harv = factor(paste0(year, ".", crop)))
# Convert 1911 from tons to pounds
dat$yield[dat$year==1911] <- 340 * dat$yield[dat$year==1911]

# Average yields. Harris 1928, table 2.
aggregate(yield~harv, dat, mean)

# Corrgram
if(require(reshape2) & require(corrgram)){
  mat <- acast(dat, series+plot~harv, value.var='yield')
  corrgram(mat, main="harris.multi.uniformity - correlation of crop yields")
}

# Compare to Harris 1928, table 4. More positive than negative correlations.
# densityplot(as.vector(cor(mat)), xlab="correlations",
#             main="harris.multi.uniformity")

## Not run:

# Standardize yields for each year
mats <- scale(mat)
# Melt and re-name columns so we can make field maps. Obvious spatial
# patterns that persist over years
d2 <- melt(mats)
names(d2) <- c('ord', 'harv', 'yield')
d2$series <- as.numeric(substring(d2$ord,1,1))
d2$plot <- as.numeric(substring(d2$ord,3))

# Series 2 is on the east side, so switch 2 and 3 for correct plotting
d2$xord <- 5 - dat$series
# Note that for alfalfa, higher-yielding plots in 1912-1914 were
# lower-yielding in 1922-1923.
if(require(desplot)){
  desplot(yield ~ xord*plot|harv, d2,
          aspect=536/639, flip=TRUE, # true aspect
          main="harris.multi.uniformity")

# Crude fertility map by averaging across years shows probable
# sub-surface water effects
agg <- aggregate(yield ~ xord + plot, data=d2, mean)
desplot(yield ~ xord + plot, agg,
        aspect=536/639, # true aspect
        main="harris.multi.uniformity fertility")
}

```

```
## End(Not run) # dontrun
```

```
harris.wateruse      Water use by horticultural trees
```

Description

Water use by horticultural trees

Format

A data frame with 1040 observations on the following 6 variables.

```
species species factor, 2 levels
age age factor, 2 levels
tree tree factor, 40 (non-consecutive) levels
day day, numeric
water water use, numeric
```

Details

Ten trees in each of four groups (two species, by two ages) were assessed for water usage, approximately every five days.

Missing values are included for the benefit of `asreml`, which needs a 'balanced' data set due to the kronecker-like syntax of the R matrix.

Source

Schabenberger, Oliver and Francis J. Pierce. 2002. *Contemporary Statistical Models for the Plant and Soil Sciences*. CRC Press. Page 512.

Used with permission of Roger Harris at Virginia Polytechnic.

Examples

```
data(harris.wateruse)
dat <- harris.wateruse

# Compare to Schabenberger & Pierce, fig 7.23
if(require(latticeExtra)){
  useOuterStrips(xyplot(water ~ day|species*age,dat, as.table=TRUE,
                       group=tree, type=c('p','smooth'),
                       main="harris.wateruse 2 species, 2 ages (10 trees each)"))
}

# Note that measurements on day 268 are all below the trend line and
```

```

# thus considered outliers. Delete them.
dat <- subset(dat, day!=268)

# Schabenberger figure 7.24
xyplot(water ~ day|tree,dat, subset=age=="A2" & species=="S2",
       as.table=TRUE, type=c('p','smooth'),
       ylab="Water use profiles of individual trees",
       main="harris.wateruse (Age 2, Species 2)")

# Rescale day for nicer output, and convergence issues, add quadratic term
dat <- transform(dat, ti=day/100)
dat <- transform(dat, ti2=ti*ti)

# Start with a subgroup: age 2, species 2
d22 <- droplevels(subset(dat, age=="A2" & species=="S2"))

# ----- Model 1, for subgroup A2,S2

# First, a fixed quadratic that is common to all trees, plus
# a random quadratic deviation for each tree.

## Schabenberger, Output 7.26
## proc mixed;
##   class tree;
##   model water = ti ti*ti / s;
##   random intercept ti ti*ti/subject=tree;

require(nlme)
## We use pdDiag() to get uncorrelated random effects
m1n <- lme(water ~ 1 + ti + ti2, data=d22, na.action=na.omit,
          random = list(tree=pdDiag(~1+ti+ti2)))

## Not run:
# Various other models with lme4 & asreml

require(lucid)
vc(m1n)
##      effect variance   stddev
## (Intercept)  0.2691 0.5188
##          ti      0    0.0000144
##          ti2    0    0.0000039
##      Residual  0.1472 0.3837

require(lme4)
m1l <- lmer(water ~ 1 + ti + ti2 + (1|tree) +
           (0+ti|tree) + (0+ti2|tree), data=d22)

vc(m1l)
##      grp      var1 var2   vcov   sdcov
## tree (Intercept) <NA> 0.2691 0.5188
## tree.1          ti <NA> 0      0

```

```

## tree.2          ti2 <NA> 0      0
## Residual        <NA> <NA> 0.1472 0.3837

# Once the overall quadratic trend has been removed, there is not
# too much evidence for consecutive observations being correlated
d22r <- subset(d22, !is.na(water))
d22r$res <- resid(m1n)
xyplot(res ~ day|tree,d22r,
        as.table=TRUE, type=c('p','smooth'),
        ylab="residual",
        main="harris.wateruse - Residuals of individual trees")
op <- par(mfrow=c(4,3))
tapply(d22r$res, d22r$tree, acf)
par(op)

# ----- Model 2, add correlation of consecutive measurements

## Schabenberger (page 516) adds correlation.
## Note how the fixed quadratic model is on the "ti = day/100" scale
## and the correlated observations are on the "day" scale. The
## only impact this has on the fitted model is to increase the
## correlation parameter by a factor of 100, which was likely
## done to get better convergence.

## proc mixed data=age2sp2;
## class tree;
## model water = ti ti*ti / s ;
## random intercept /subject=tree s;
## repeated /subject=tree type=sp(exp)(day);

## Same as SAS, use ti for quadratic, day for correlation
m2l <- lme(water ~ 1 + ti + ti2, data=d22,
          random = ~ 1|tree,
          cor = corExp(form=~ day|tree),
          na.action=na.omit)
m2l # Match output 7.27. Same fixef, ranef, variances, exp corr

vc(m2l)
##      effect variance stddev
## (Intercept)  0.2656 0.5154
## Residual    0.1541 0.3926

# ---

## Now use asreml. When I tried rcov=~tree:exp(ti),
## the estimated parameter value was on the 'boundary', i.e. 0.
## Changing rcov to the 'day' scale produced a sensible estimate
## that matched SAS.
## Note: SAS and asreml use different parameterizations for the correlation
## SAS uses exp(-d/phi) and asreml uses phi^d.
## SAS reports 3.79, asreml reports 0.77, and exp(-1/3.7945) = 0.7683274
## Note: normally a quadratic would be included as 'pol(day,2)'

```

```

require(asreml)
d22 <- d22[order(d22$tree, d22$day),]
m2a <- asreml(water ~ 1 + ti + ti2,
             data=d22,
             random = ~ tree,
             rcov=~tree:exp(day))

vc(m2a)
##          effect component std.error z.ratio constr
## tree!tree.var    0.2656  0.1301     2      pos
## R!variance      0.1541  0.01611    9.6      pos
## R!day.pow       0.7683  0.04191    18      uncon

# ----- Model 3. Full model for all species/ages. Schabenberger p. 518

## /* Continuous AR(1) autocorrelations included */
## proc mixed data=wateruse;
## class age species tree;
## model water = age*species age*species*ti age*species*ti*ti / noint s;
## random intercept ti / subject=age*species*tree s;
## repeated / subject=age*species*tree type=sp(exp)(day);

m3l <- lme(water ~ 0 + age:species + age:species:ti + age:species:ti2,
          data=dat, na.action=na.omit,
          random = list(tree=pdDiag(~1+ti)),
          cor = corExp(form=~ day|tree),
          )

m3l # Match Schabenberger output 7.27. Same fixef, ranef, variances, exp corr

vc(m3l)
##          effect variance stddev
## (Intercept)  0.1549  0.3936
## ti          0.02785  0.1669
## Residual    0.16    0.4

# --- asreml

dat <- dat[order(dat$tree,dat$day),]
m3a <- asreml(water ~ 0 + age:species + age:species:ti + age:species:ti2,
             data=dat,
             random = ~ age:species:tree + age:species:tree:ti,
             rcov = ~ tree:exp(day)
             )

vc(m3a) # Note: day.pow = .8091 = exp(-1/4.7217)
##          effect component std.error z.ratio constr
## age:species:tree!age.var    0.1549  0.07192     2.2      pos
## age:species:tree:ti!age.var  0.02785  0.01343     2.1      pos
## R!variance                   0.16    0.008917    18      pos

```



```
##          R!day.pow  0.8091  0.01581  51  uncon
```

```
## End(Not run)
```

harrison.priors *Ranges of analytes in soybean from other authors*

Description

Ranges of analytes in soybean from other authors

Format

A data frame with 80 observations on the following 5 variables.

source Source document

substance Analyte substance

min minimum amount (numeric)

max maximum analyte amount (numeric)

number number of substances

Details

Harrison et al. show how to construct an informative Bayesian prior from previously-published ranges of concentration for several analytes.

The units for daidzein, genistein, and glycitein are micrograms per gram.

The raffinose and stachyose units were converted to a common 'percent' scale.

The author names in the 'source' variable are shortened forms of the citations in the supplemental information of Harrison et al.

Source

Jay M. Harrison, Matthew L. Breeze, Kristina H. Berman, George G. Harrigan. 2013. Bayesian statistical approaches to compositional analyses of transgenic crops 2. Application and validation of informative prior distributions. *Regulatory Toxicology and Pharmacology*, 65, 251-258. <https://doi.org/10.1016/j.yrtph.2012.12.002>

Data retrieved from the Supplemental Information of this source.

References

Jay M. Harrison, Derek Culp, George G. Harrigan. 2013. Bayesian MCMC analyses for regulatory assessments of safety in food composition *Proceedings of the 24th Conference on Applied Statistics in Agriculture (2012)*.

Examples

```

data(harrison.priors)
dat <- harrison.priors

d1 <- subset(dat, substance=="daidzein")

# Stack the data to 'tall' format and calculate empirical cdf
d1t <- with(d1, data.frame(xx = c(min, max), yy=c(1/(number+1), number/(number+1))))

# Harrison 2012 Example 4: Common prior distribution
# Harrison uses the minimum and maximum levels of daidzein from previous
# studies as the first and last order statistics of a lognormal
# distribution, and finds the best-fit lognormal distribution.

m0 <- mean(log(d1t$xx)) # 6.37
s0 <- sd(log(d1t$xx)) # .833
mod <- nls(yy ~ plnorm(xx, meanlog, sdlog), data=d1t,
          start=list(meanlog=m0, sdlog=s0))
coef(mod) # Matches Harrison 2012
## meanlog sdlog
## 6.4187829 0.6081558

plot(yy~xx, data=d1t, xlim=c(0,2000), ylim=c(0,1),
     main="harrison.priors - Common prior", xlab="daidzein level", ylab="CDF")
mlog <- coef(mod)[1] # 6.4
slog <- coef(mod)[2] # .61
xvals <- seq(0, 2000, length=100)
lines(xvals, plnorm(xvals, meanlog=mlog, sdlog=slog))

d1a <- d1
d1a$source <- as.character(d1a$source)
d1a[19,'source'] <- "(All)" # Add a blank row for the densitystrip
d1
if(require(latticeExtra)){
  # Plot the range for each source, a density curve (with arbitrary
  # vertical scale) for the common prior distribution, and a density
  # strip by stacking the individual bands and using transparency
  segplot(factor(source) ~ min+max, d1a,
         main="harrison.priors", xlab="daidzein level", ylab="source") +
  xyplot(5000*dlnorm(xvals, mlog, slog)~xvals, type='l') +
  segplot(factor(rep(1,18)) ~ min+max, d1, 4, level=d1$number,
         col.regions="gray20", alpha=.1)
}

```

Description

Average daily gain of 65 steers for 3 lines, 9 sires.

Usage

```
data("harvey.lsmmeans")
```

Format

A data frame with 65 observations on the following 7 variables.

line line of the dam

sire sire

damage age class of the dam

calf calf number

weanage calf age at weaning

weight calf weight at start of feeding

adg average daily gain

Details

The average daily gain 'adg' for each of 65 Hereford steers.

The calf age at weaning and initial weight at the beginning of the test feeding is also given.

The steers were fed for the same length of time in the feed lot.

It is assumed that each calf has a unique dam and there are no twins or repeat matings.

Harvey (1960) is one of the earliest papers presenting least squares means (lsmmeans).

Source

Harvey, Walter R. (1960). Least-squares Analysis of Data with Unequal Subclass Numbers. Technical Report ARS No 20-8. USDA, Agricultural Research Service. Page 101-102.

Reprinted as ARS H-4, 1975. <https://archive.org/details/least-squares-anal04harv>

References

Also appears in the 'dmm' package as 'harv101.df' See that package vignette for a complete analysis of the data.

Examples

```
data(harvey.lsmmeans)
dat = harvey.lsmmeans
```

```
require(lattice)
dotplot(adg ~ sire|line,dat,
        main="harvey.lsmmeans", xlab="sire", ylab="average daily gain")
```

```
## Not run:
# Model suggested by Harvey on page 103
m0 <- lm(adg ~ 1 + line + sire + damage + line:damage + weanage + weight, data=dat)
# Due to contrast settings, it can be hard to compare model coefficients to Harvey,
# but note the slopes of the continuous covariates match Harvey p. 107, where his
# b is weanage, d is weight
# coef(m0)
#      weanage      weight
# -0.008154879  0.001970446

# A quick attempt to reproduce table 4 of Harvey, p. 109. Not right.
# require(lsmmeans)
# lsmmeans(m0,c('line','sire','damage'))

## End(Not run)
```

 harville.lamb

Birth weight of lambs from different lines/sires

Description

Birth weight of lambs from different lines/sires

Usage

```
data("harville.lamb")
```

Format

A data frame with 62 observations on the following 4 variables.

```
line  genotype line number
sire   sire number
damage dam age, class 1,2,3
weight lamb birth weight
```

Details

Weight at birth of 62 lambs. There were 5 distinct lines.

Some sires had multiple lambs. Each dam had one lamb.

The age of the dam is a category: 1 (1-2 years), 2 (2-3 years) or 3 (over 3 years).

Note: Jiang, gives the data in table 1.2, but there is a small error. Jiang has a weight 9.0 for sire 31, line 3, age 3. The correct value is 9.5.

Source

David A. Harville and Alan P. Fenech (1985). Confidence Intervals for a Variance Ratio, or for Heritability, in an Unbalanced Mixed Linear Model. *Biometrics*, 41, 137-152. <http://doi.org/10.2307/2530650>

References

Jiming Jiang, *Linear and Generalized Linear Mixed Models and Their Applications*. Table 1.2.

Andre I. Khuri, *Linear Model Methodology*. Table 11.5. Page 368. <https://books.google.com/books?id=UfDvCAAQBAJ&>

Daniel Gianola, Keith Hammond. *Advances in Statistical Methods for Genetic Improvement of Livestock*. Table 8.1, page 165.

Examples

```
data(harville.lamb)
dat <- harville.lamb
dat <- transform(dat, line=factor(line), sire=factor(sire), damage=factor(damage))

library(lattice)
bwplot(weight ~ line, dat,
        main="harville.lamb",
        xlab="line", ylab="birth weights")

## Not run:
if(require(lme4) & require(lucid)){

  m1 <- lmer(weight ~ -1 + line + damage + (1|sire), data=dat)
  summary(m1)
  vc(m1) # Khuri reports variances 0.5171, 2.9616
  ##      grp      var1 var2  vcov  sdcor
  ##   sire (Intercept) <NA> 0.5171 0.7191
  ## Residual      <NA> <NA> 2.962  1.721

}

## End(Not run)
```

 hayman.tobacco

Diallel cross of Aztec tobacco

Description

Diallel cross of Aztec tobacco in 2 reps

Format

year year
 block block factor, 2 levels
 male male parent, 8 levels
 female female parent
 day mean flowering time (days)

Details

Data was collected in 1951 (Hayman 1954a) and 1952 (Hayman 1954b).

In each year there were 8 varieties of Aztec tobacco, *Nicotiana rustica* L..

Each cross/self was represented by 10 progeny, in two plots of 5 plants each. The data are the mean flowering time per plot.

Note, the 1951 data as published in Hayman (1954a) Table 5 contain "10 times the mean flowering time". The data here have been divided by 10 so as to be comparable with the 1952 data.

Hayman (1954b) says "Table 2 lists...three characters from a diallel cross of *Nicotiana rustica* varieties which was repeated for three years." This seems to indicate that the varieties are the same in 1951 and 1952. Calculating the GCA effects separately for 1951 and 1952 and then comparing these estimates shows that they are highly correlated.

Source

B. I. Hayman (1954a). The Analysis of Variance of Diallel Tables. *Biometrics*, 10, 235-244. Table 5, page 241. <http://doi.org/10.2307/3001877>

Hayman, B.I. (1954b). The theory and analysis of diallel crosses. *Genetics*, 39, 789-809. Table 3, page 805. <http://www.genetics.org/content/39/6/789.full.pdf>

References

For 1951 data

Mohring, Melchinger, Piepho. (2011). REML-Based Diallel Analysis. *Crop Science*, 51, 470-478.

For 1952 data

C. Clark Cockerham and B. S. Weir. (1977). Quadratic analyses of reciprocal crosses. *Biometrics* 33, 187-203. Appendix C.

Examples

```
# 1951 data. Fit the first REML model of Mohring 2011 Supplement.
```

```
data(hayman.tobacco)
dat1 <- subset(hayman.tobacco, year==1951)
```

```
# Make a factor 'comb' in which G1xG2 is the same cross as G2xG1
dat1 <- transform(dat1, comb =
  ifelse(as.character(male) < as.character(female),
```

```

                                paste0(male,female), paste0(female,male)))
# 'dr' is the direction of the cross, 0 for self
dat1$dr <- 1
dat1 <- transform(dat1,
                  dr = ifelse(as.character(male) < as.character(female), -1, dr))
dat1 <- transform(dat1,
                  dr = ifelse(as.character(male) == as.character(female), 0, dr))

# ASREML code for Mixed Griffing.
# Mohring Table 2, column 2 (after dividing by 10^2) gives variances:
# GCA 12.77, SCA 11.09, RSCA .65, Error 4.23.
# Mohring Supplement ASREML code part1 model is:
# y ~ mu r !r mother and(father) combination combination.dr
# Note that the levels of 'male' and 'female' are the same, so the
# and(female) term tells asreml to use the same levels (or, equivalently,
# fix the correlation of the male/female levels to be 1.
# The block effect is minimal and therefore ignored.
## m1 <- asreml(day~1, data=dat1,
##             random = ~ male + and(female) + comb + comb:dr)
## require(lucid)
## vc(m1)
##           effect component std.error z.ratio      con
##   male!male.var   12.77     7.502    1.7 Positive
##   comb!comb.var   11.11     3.353    3.3 Positive
## comb:dr!comb.var  0.6603    0.4926    1.3 Positive
##           R!variance  4.185     0.7449    5.6 Positive

# -----

# 1952 data. Reproduce table 3 and figure 2 of Hayman 1954b.

dat2 <- subset(hayman.tobacco, year==1952)

# Does flowering date follow a gamma distn? Maybe.
require(lattice)
densityplot(~day, data=dat2, main="hayman.tobacco",
            xlab="flowering date")

d1 <- subset(dat2, block=='B1')
d2 <- subset(dat2, block=='B2')

if(require(reshape2)){
m1 <- acast(d1, male~female, value.var='day')
m2 <- acast(d2, male~female, value.var='day')

mn1 <- (m1+t(m1))/2
mn2 <- (m2+t(m2))/2

# Variance and covariance of 'rth' offspring
vr1 <- apply(mn1, 1, var)
vr2 <- apply(mn2, 1, var)
wr1 <- apply(mn1, 1, cov, diag(mn1))

```

```

wr2 <- apply(mn2, 1, cov, diag(mn2))
# Remove row names to prevent a mild warning
rownames(mn1) <- rownames(mn2) <- NULL
summ <- data.frame(rbind(mn1,mn2))
summ$block <- rep(c('B1','B2'), each=8)
summ$vr <- c(vr1,vr2)
summ$wr <- c(wr1,wr2)
summ$male <- rep(1:8,2) # Vr and Wr match Hayman table 3

with(summ, plot(wr~vr, type='n', main="hayman.tobacco"))
with(summ, text(vr, wr, male)) # Match Hayman figure 2
abline(0,1,col="gray")
# Hayman notes that 1 and 3 do not lie along the line, so modifies them
# and re-analyzes.
}

```

hazell.vegetables *Gross profit for 4 vegetable crops in 6 years*

Description

Gross profit for 4 vegetable crops in 6 years

Usage

```
data("hazell.vegetables")
```

Format

A data frame with 6 observations on the following 5 variables.

year year factor, 6 levels
 carrot Carrot profit, dollars/acre
 celery Celery profit, dollars/acre
 cucumber Cucumber profit, dollars/acre
 pepper Pepper profit, dollars/acre

Details

The values in the table are gross profits (loss) in dollars per acre. The criteria in the example below are (1) total acres < 200, (2) total labor < 10000, (3) crop rotation.

The example shows how to use linear programming to maximize expected profit.

Source

P.B.R. Hazell, (1971). A linear alternative to quadratic and semivariance programming for farm planning under uncertainty. *Am. J. Agric. Econ.*, 53, 53-62. <http://doi.org/10.2307/3180297>

References

Carlos Romero, Tahir Rehman. (2003). *Multiple Criteria Analysis for Agricultural Decisions*. Elsevier.

Examples

```

data("hazell.vegetables")
dat <- hazell.vegetables

if(FALSE & require(lattice)) {
  xyplot(carrot+celery+cucumber+pepper ~ year,dat,
         ylab="yearly profit by crop",
         type='b', auto.key=list(columns=4),
         panel.hline=0)
}

# optimal strategy for planting crops (calculated below)
dat2 <- apply(dat[,-1], 1, function(x) x*c(0, 27.5, 100, 72.5))/1000
colnames(dat2) <- rownames(dat)
barplot(dat2, legend.text=c(" 0 carrot", "27.5 celery", " 100 cucumber", "72.5 pepper"),
        xlim=c(0,7), ylim=c(-5,120),
        col=c('orange','green','forestgreen','red'),
        xlab="year", ylab="Gross profit, $1000",
        main="hazell.vegetables - retrospective profit from optimal strategy",
        args.legend=list(title="acres, crop"))

## Not run:

require(linprog)
# colMeans(dat[ , -1])
# 252.8333 442.6667 283.8333 515.8333

# cvec = avg across-years profit per acre for each crop
cvec <- c(253, 443, 284, 516)

# Maximize c'x for Ax=b
A <- rbind(c(1,1,1,1), c(25,36,27,87), c(-1,1,-1,1))
colnames(A) <- names(cvec) <- c("carrot","celery","cucumber","pepper")
rownames(A) <- c('land','labor','rotation')

# bvec criteria = (1) total acres < 200, (2) total labor < 10000,
# (3) crop rotation.

bvec <- c(200,10000,0)
const.dir <- c("<=", "<=", "<=")

m1 <- solveLP(cvec, bvec, A, maximum=TRUE, const.dir=const.dir, lpSolve=TRUE)
# m1$solution # optimal number of acres for each crop
# carrot celery cucumber pepper
# 0.00000 27.45098 100.00000 72.54902

```

```

# Average income for this plan
## sum(cvec * m1$solution)
## [1] 77996.08

# Year-to-year income for this plan
## as.matrix(dat[,-1])
##           [,1]
## [1,] 80492.16
## [2,] 80431.37
## [3,] 81884.31
## [4,] 106868.63
## [5,] 37558.82
## [6,] 80513.73

# optimum allocation that minimizes year-to-year income variability.
# brute-force search

# For generality, assume we have unequal probabilities for each year.
probs <- c(.15, .20, .20, .15, .15, .15)
# Randomly allocate crops to 200 acres, 100,000 times
#set.seed(1)
mat <- matrix(runif(4*100000), ncol=4)
mat <- 200*sweep(mat, 1, rowSums(mat), "/")
# each row is one strategy, showing profit for each of the six years
# profit <- mat
profit <- tcrossprod(mat, as.matrix(dat[,-1])) # Each row is profit, columns are years
# calculate weighted variance using year probabilities
wtvar <- apply(profit, 1, function(x) cov.wt(as.data.frame(x), wt=probs)$cov)
# five best planting allocations that minimizes the weighted variance
ix <- order(wtvar)[1:5]
mat[ix,]
## carrot celery cucumber pepper
##           [,1]      [,2]      [,3]      [,4]
## [1,] 71.26439 28.09259 85.04644 15.59657
## [2,] 72.04428 27.53299 84.29760 16.12512
## [3,] 72.16332 27.35147 84.16669 16.31853
## [4,] 72.14622 29.24590 84.12452 14.48335
## [5,] 68.95226 27.39246 88.61828 15.03700

## End(Not run)

```

heady.fertilizer

Yield of corn, alfalfa, clover with two fertilizers

Description

Yield of corn, alfalfa, clover with two fertilizers

Usage

```
data("heady.fertilizer")
```

Format

A data frame with 81 observations on the following 3 variables.

crop crop

rep replicate (not block)

P phosphorous, pounds/acre

K potassium, pounds/acre

N nitrogen, pounds/acre

yield yield

Details

Heady et al. fit two-variable semi-polynomial response surfaces for each crop.

Clover and alfalfa yields are in tons/acre. The clover and alfalfa experiments were grown in 1952.

Corn yields are given as bu/acre. The corn experiments were grown in 1952 and 1953. The same test plots were used in 1953 and in 1952, but no fertilizer was applied in 1953—any response in yield is due to residual fertilizer from 1952.

All experiments used an incomplete factorial design. Not all treatment combinations were present.

Source

Earl O. Heady, John T. Pesek, William G. Brown. (1955). Crop Response Surfaces and Economic Optima in Fertilizer Use. Agricultural Experiment Station, Iowa State College. Research bulletin 424. Pages 330-332. https://lib.dr.iastate.edu/cgi/viewcontent.cgi?filename=12&article=1032&context=ag_researchbulletins

References

Pesek, John and Heady, Earl O. 1956. A two nutrient-response function with determination of economic optima for the rate and grade of fertilizer for alfalfa. *Soil Science Society of America Journal*, 20, 240-246. <https://doi.org/10.2136/sssaj1956.03615995002000020025x>

Examples

```
data(heady.fertilizer)
dat <- heady.fertilizer

require(lattice)
xyplot(yield ~ P|crop, data=dat, scales=list(relation="free"),
       groups=factor(paste(dat$N,dat$K)), auto.key=list(columns=5),
       main="heady.fertilizer", xlab="Phosphorous")

# Corn. Matches Heady, p. 292
d1 <- subset(dat, crop=="corn")
```

```

m1 <- lm(yield ~ N + P + sqrt(N) + sqrt(P) + sqrt(N*P), data=d1)
summary(m1)

# Alfalfa. Matches Heady, p. 292. Also Pesek equation 3, p. 241
d2 <- subset(dat, crop=="alfalfa")
m2 <- lm(yield ~ K + P + sqrt(K) + sqrt(P) + sqrt(K*P), data=d2)
summary(m2)
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.8735521  0.1222501  15.326 < 2e-16 ***
## K           -0.0013943  0.0007371  -1.891 0.061237 .
## P           -0.0050195  0.0007371  -6.810 5.74e-10 ***
## sqrt(K)      0.0617458  0.0160142   3.856 0.000196 ***
## sqrt(P)      0.1735383  0.0160142  10.837 < 2e-16 ***
## sqrt(K * P) -0.0014402  0.0007109  -2.026 0.045237 *

# Clover. Matches Heady, p. 292.
d3 <- subset(dat, crop=="clover")
m3 <- lm(yield ~ P + sqrt(K) + sqrt(P) + sqrt(K*P), data=d3)
summary(m3)

# Corn with residual fertilizer. Matches Heady eq 56, p. 322.
d4 <- subset(dat, crop=="corn2")
m4 <- lm(yield ~ N + P + sqrt(N) + sqrt(P) + sqrt(N*P), data=d4)
summary(m4)

## Not run:
require(rgl)
with(d1, plot3d(N,P,yield))
with(d2, plot3d(K,P,yield))
with(d3, plot3d(K,P,yield))
with(d4, plot3d(N,P,yield)) # Mostly linear in both N and P
rgl.close()

## End(Not run)

```

henderson.milkfat *Milk fat yields for a single cow*

Description

Average daily fat yields (kg/day) from milk from a single cow for each of 35 weeks.

Format

A data frame with 35 observations on the following 2 variables.

week week, numeric

yield yield, kg/day

Source

Charles McCulloch. Workshop on Generalized Linear Mixed Models.
Used with permission of Charles McCulloch and Harold Henderson.

References

None.

Examples

```
data(henderson.milkfat)
dat <- henderson.milkfat

plot(yield~week, data=dat, cex = 0.8, ylim=c(0,.9),
     main="henderson.milkfat", xlab = "Week",
     ylab = "Fat yield (kg/day)")

# Yield ~ a * t^b * exp(g*t) # where t is time
m1 <- nls(yield ~ alpha * week^beta * exp(gamma * week),
         data=dat,
         start=list(alpha=.1, beta=.1, gamma=.1))

# Or, take logs and fit a linear model
# log(yield) ~ log(alpha) + beta*log(t) + gamma*t
m2 <- lm(log(yield) ~ 1 + log(week) + week, dat)

# Or, use glm and a link to do the transform
m3 <- glm(yield ~ 1 + log(week) + week, quasi(link = "log"), dat)

# Note: m2 has E[log(y)] = log(alpha) + beta*log(t) + gamma*t
# and m3 has log(E[y]) = log(alpha) + beta*log(t) + gamma*t

# Generalized additive models
require("mgcv")
m4 <- gam(log(yield) ~ s(week), gaussian, dat)
m5 <- gam(yield ~ s(week), quasi(link = "log"), dat)

# Model predictions
pdat <- data.frame(week = seq(1, 35, by = 0.1))
pdat <- transform(pdat, p1 = predict(m1, pdat),
                 p2 = exp(predict(m2, pdat)), # back transform
                 p3 = predict(m3, pdat, type="resp"), # response scale
                 p4 = exp(predict(m4, pdat)),
                 p5 = predict(m5, pdat, type="response"))

# Compare fits
with(pdat, {
  lines(week, p1)
```

```

lines(week, p2, col = "red", lty="dotted")
lines(week, p3, col = "red", lty="dashed")
lines(week, p4, col = "blue", lty = "dashed")
lines(week, p5, col = "blue")
})

legend("topright",
  c("obs", "lm, log-transformed", "glm, log-link",
    "gam, log-transformed", "gam, log-link"),
  lty = c("solid", "dotted", "dashed", "dashed", "solid"),
  col = c("black", "red", "red", "blue", "blue"),
  cex = 0.8, bty = "n")

```

hernandez.nitrogen *Multi-environment trial of corn with nitrogen fertilizer at 5 sites.*

Description

Corn response to nitrogen fertilizer at 5 sites.

Format

A data frame with 136 observations on the following 5 variables.

site site factor, 5 levels

loc location name

rep rep, 4 levels

nitro nitrogen, kg/ha

yield yield, Mg/ha

Details

Experiment was conducted in 2006 at 5 sites in Minnesota.

Source

Hernandez, J.A. and Mulla, D.J. 2008. Estimating uncertainty of economically optimum fertilizer rates, *Agronomy Journal*, 100, 1221-1229. <https://doi.org/10.2134/agronj2007.0273>

Electronic data kindly supplied by Jose Hernandez.

Examples

```

data(hernandez.nitrogen)
dat <- hernandez.nitrogen
cprice <- 118.1 # $118.1/Mg or $3/bu
nprice <- 0.6615 # $0.66/kg N or $0.30/lb N

# Hernandez optimized yield with a constraint on the ratio of the prices.
# Simpler to just calculate the income and optimize that.
dat <- transform(dat, inc = yield * cprice - nitro * nprice)
require(lattice)
xyplot(inc ~ nitro|site, dat, groups=rep, auto.key=list(columns=4),
       xlab="nitrogen", ylab="income", main="hernandez.nitrogen")

# Site 5 only
dat1 <- subset(dat, site=='S5')

# When we optimize on income, a simple quadratic model works just fine,
# and matches the results of the nls model below.
# Note, 'poly(nitro)' gives weird coeffs
lm1 <- lm(inc ~ 1 + nitro + I(nitro^2), data=dat1)
c1 <- coef(lm1)
-c1[2] / (2*c1[3])
##      nitro
## 191.7198   # Optimum nitrogen is 192 for site 5

## Not run:
# Use the delta method to get a conf int
require("car")
del1 <- deltaMethod(lm1, "-b1/(2*b2)", parameterNames= paste("b", 0:2, sep=""))
# Simple Wald-type conf int for optimum
del1$Est + c(-1,1) * del1$SE * qt(1-.1/2, nrow(dat1)-length(coef(lm1)))
## 118.9329 264.5067

# Nonlinear regression
# Reparameterize  $b_0 + b_1x + b_2x^2$  using  $th_2 = -b_1/2b_2$  so that  $th_2$  is optimum
nls1 <- nls(inc ~ th11 - (2*th2*th12)*nitro + th12*nitro^2,
          data = dat1, start = list(th11 = 5, th2 = 150, th12 = -0.1),)
summary(nls1)
# Wald conf int
wald <- function(object, alpha=0.1){
  nobs <- length(resid(object))
  npar <- length(coef(object))
  est <- coef(object)
  stderr <- summary(object)$parameters[,2]
  tval <- qt(1-alpha/2, nobs-npar)
  ci <- cbind(est - tval * stderr, est + tval * stderr)
  colnames(ci) <- paste(round(100*c(alpha/2, 1-alpha/2), 1), "pct", sep= "")
  return(ci)
}

```

```

round(wald(nls1),2)
##          5
## th11 936.44 1081.93
## th2  118.93  264.51  # th2 is the optimum
## th12 -0.03  -0.01

# Likelihood conf int
require(MASS)
round(confint(nls1, "th2", level = 0.9),2)
##          5
## 147.96 401.65

# Bootstrap conf int
require(boot)
dat1$fit <- fitted(nls1)
bootfun <- function(rs, i) { # bootstrap the residuals
  dat1$y <- dat1$fit + rs[i]
  coef(nls(y ~ th11- (2*th2*th12)*nitro + th12*nitro^2, dat1, start = coef(nls1)))
}
res1 <- scale(resid(nls1), scale = FALSE) # remove the mean. Why? It is close to 0.
set.seed(1) # Sometime the bootstrap fails, but this seed works
boot1 <- boot(res1, bootfun, R = 500)
boot.ci(boot1, index = 2, type = c("perc"), conf = 0.9)
## Level      Percentile
## 90

## End(Not run)

```

hessling.argentina *Relation between wheat yield and weather in Argentina*

Description

Relation between wheat yield and weather in Argentina

Format

A data frame with 30 observations on the following 15 variables.

yield average yield, kg/ha

year year

p05 precipitation (mm) in May

p06 precip in June

p07 precip in July

p08 precip in August
 p09 precip in September
 p10 precip in October
 p11 precip in November
 p12 precip in December
 t06 june temperature deviation from normal, deg Celsius
 t07 july temp deviation
 t08 august temp deviation
 t09 september temp deviation
 t10 october temp deviation
 t11 november temp deviation

Details

In Argentina wheat is typically sown May to August. Harvest begins in November or December.

Source

N. A. Hessling, 1922. Relations between the weather and the yield of wheat in the Argentine republic, *Monthly Weather Review*, 50, 302-308. [http://doi.org/10.1175/1520-0493\(1922\)50<302:RBTWAT>2.0.CO;2](http://doi.org/10.1175/1520-0493(1922)50<302:RBTWAT>2.0.CO;2)

Examples

```
data(hessling.argentina)
dat <- hessling.argentina

# Fig 1 of Hessling. Use avg Aug-Nov temp to predict yield
dat <- transform(dat, avetmp=(t08+t09+t10+t11)/4) # Avg temp
m0 <- lm(yield ~ avetmp, dat)
plot(yield~year, dat, ylim=c(100,1500), type='l',
main="hessling.argentina: observed (black) and predicted yield (blue)")
lines(fitted(m0)~year, dat, col="blue")

# A modern, PLS approach
if(require(pls)){
  yld <- dat[, "yield", drop=FALSE]
  yld <- as.matrix(sweep(yld, 2, colMeans(yld)))
  cov <- dat[,c("p06", "p07", "p08", "p09", "p10", "p11", "t08", "t09", "t10", "t11")]
  cov <- as.matrix(scale(cov))
  m2 <- pls(yld~cov)

  # biplot(m2, which="x", var.axes=TRUE, main="hessling.argentina")
}

if(require(corrgram)){
  corrgram(dat, main="hessling.argentina - correlations of yield and covariates")
}
```

hildebrand.systems *Multi-environment trial of maize for four cropping systems*

Description

Maize yields for four cropping systems at 14 on-farm trials.

Format

A data frame with 56 observations on the following 4 variables.

village village, 2 levels

farm farm, 14 levels

system cropping system

yield yield, t/ha

Details

Yields from 14 on-farm trials in Phalombe Project region of south-eastern Malawi. The farms were located near two different villages.

On each farm, four different cropping systems were tested. The systems were: LM = Local Maize, LMF = Local Maize with Fertilizer, CCA = Improved Composite, CCAF = Improved Composite with Fertilizer.

Source

P. E. Hildebrand, 1984. Modified Stability Analysis of Farmer Managed, On-Farm Trials. *Agronomy Journal*, 76, 271–274. <https://doi.org/10.2134/agronj1984.00021962007600020023x>

References

H. P. Piepho, 1998. Methods for Comparing the Yield Stability of Cropping Systems. *Journal of Agronomy and Crop Science*, 180, 193–213. <http://doi.org/10.1111/j.1439-037X.1998.tb00526.x>

Examples

```
data(hildebrand.systems)
dat <- hildebrand.systems

# Piepho 1998 Fig 1
require(lattice)
dotplot(yield ~ system, dat, groups=village, auto.key=TRUE,
        main="hildebrand.systems", xlab="cropping system by village")

# Plot of risk of 'failure' of System 2 vs System 1
s11 = .30; s22 <- .92; s12 = .34
```

```

mu1 = 1.35; mu2 = 2.70
lambda <- seq(from=0, to=5, length=20)
system1 <- pnorm((lambda-mu1)/sqrt(s11))
system2 <- pnorm((lambda-mu2)/sqrt(s22))

# A simpler view
plot(lambda, system1, type="l", xlim=c(0,5), ylim=c(0,1),
      xlab="Yield level", ylab="Prob(yield < level)",
      main="hildebrand.systems - risk of failure for each system")
lines(lambda, system2, col="red")

# Prob of system 1 outperforming system 2. Table 8
pnorm((mu1-mu2)/sqrt(s11+s22-2*s12))
# .0331

# -----

## Not run:
# asreml3
require(asreml)
# Environmental variance model, unstructured correlations

dat <- dat[order(dat$system, dat$farm),]
m1 <- asreml(yield ~ system, data=dat, rcov = ~us(system):farm)

# Means, table 5
p1 <- predict(m1, data=dat, classify="system")$predictions$pvals
## system pred.value std.error est.stat
## CCA 1.164 0.2816 Estimable
## CCAF 2.657 0.3747 Estimable
## LM 1.35 0.1463 Estimable
## LMF 2.7 0.2561 Estimable

# Variances, table 5
require(lucid)
vc(m1)[c(2,4,7,11),]
## effect component std.error z.ratio constr
## R!system.CCA:CCA 1.11 0.4354 2.5 pos
## R!system.CCAF:CCAF 1.966 0.771 2.5 pos
## R!system.LM:LM 0.2996 0.1175 2.5 pos
## R!system.LMF:LMF 0.9185 0.3603 2.5 pos

# Stability variance model
m2 <- asreml(yield ~ system, data=dat,
            random = ~ farm,
            rcov = ~ at(system):units)
p2 <- predict(m2, data=dat, classify="system")$predictions$pvals

# Variances, table 6
vc(m2)
## effect component std.error z.ratio constr
## farm!farm.var 0.2996 0.1175 2.5 pos
## system_CCA!variance 0.4136 0.1622 2.5 pos

```

```

## system_CCAF!variance 1.267      0.4969    2.5    pos
##   system_LM!variance 0.0000002    NA      NA    bound
##   system_LMF!variance 0.5304      0.208    2.5    pos

## End(Not run)

# -----

## Not run:
## require(asreml4)
## # Environmental variance model, unstructured correlations

## dat <- dat[order(dat$system, dat$farm),]
## m1 <- asreml(yield ~ system, data=dat,
##             resid = ~us(system):farm)

## # Means, table 5
## p1 <- predict(m1, data=dat, classify="system")$pvals
## ##   system pred.value std.error est.stat
## ##   CCA      1.164    0.2816 Estimable
## ##   CCAF     2.657    0.3747 Estimable
## ##   LM       1.35     0.1463 Estimable
## ##   LMF     2.7      0.2561 Estimable

## # Variances, table 5
## require(lucid)
## vc(m1)[c(2,4,7,11),]
## ##           effect component std.error z.ratio constr
## ##   R!system.CCA:CCA    1.11    0.4354    2.5    pos
## ##   R!system.CCAF:CCAF  1.966    0.771    2.5    pos
## ##   R!system.LM:LM     0.2996    0.1175    2.5    pos
## ##   R!system.LMF:LMF   0.9185    0.3603    2.5    pos

## # Stability variance model
## m2 <- asreml(yield ~ system, data=dat,
##             random = ~ farm,
##             resid = ~ dsum( ~ units|system))
## m2 <- update(m2)
## p2 <- predict(m2, data=dat, classify="system")$pvals

## # Variances, table 6
## vc(m2)
## ##           effect component std.error z.ratio constr
## ##   farm!farm.var 0.2996    0.1175    2.5    pos
## ##   system_CCA!variance 0.4136    0.1622    2.5    pos
## ##   system_CCAF!variance 1.267    0.4969    2.5    pos
## ##   system_LM!variance 0.0000002    NA      NA    bound
## ##   system_LMF!variance 0.5304    0.208    2.5    pos

## End(Not run)

```

holland.arthropods *Counts of arthropods in a grid-sampled wheat field*

Description

Counts of arthropods in a grid-sampled wheat field

Usage

```
data("holland.arthropods")
```

Format

A data frame with 63 observations on the following 8 variables.

row row

col column

n.brevicollis species counts

linyphiidae species counts

collembola species counts

carabidae species counts

lycosidae species counts

weedcover percent weed cover

Details

Arthropods were sampled at 30m x 30m grid in a wheat field near Wimborne, Dorset, UK on 6 dates in Jun/Jul 1996. Arthropod counts were aggregated across the 6 dates.

Holland et al. used SADIE (Spatial Analysis by Distance Indices) to look for spatial patterns. Significant patterns were found for *N. brevicollis*, Carabidae, Lycosidae. The Lycosidae counts were also significantly associated with weed cover.

Source

Holland J. M., Perry J. N., Winder, L. (1999). The within-field spatial and temporal distribution of arthropods within winter wheat. *Bulletin of Entomological Research*, 89: 499-513. Figure 3 (large grid in 1996). <http://doi.org/10.1017/S0007485399000656>

Used with permission of John Holland.

Examples

```

data(holland.arthropods)
dat <- holland.arthropods

# use log count to make it possible to have same scale for insects
if(require(reshape2) & require(lattice)){
  grays <- colorRampPalette(c("#d9d9d9", "#252525"))
  dat2 <- melt(dat, id.var=c('row', 'col'))
  contourplot(log(value) ~ col*row|variable, dat2,
              col.regions=grays(7), region=TRUE,
              main="holland.arthropods - log counts in winter wheat")
}

## Not run:
# individual species
require(lattice)
grays <- colorRampPalette(c("#d9d9d9", "#252525"))
contourplot(linyphiidae ~ col*row, dat, at=c(0,40,80,120,160,200), region=TRUE,
            col.regions=grays(5),
            main="holland.arthropods - linyphiidae counts in winter wheat")
contourplot(n.brevicollis ~ col*row, dat, region=TRUE)
contourplot(linyphiidae~ col*row, dat, region=TRUE)
contourplot(collembola ~ col*row, dat, region=TRUE)
contourplot(carabidae ~ col*row, dat, region=TRUE)
contourplot(lycosidae ~ col*row, dat, region=TRUE)
contourplot(weedcover ~ col*row, dat, region=TRUE)

## End(Not run)

```

holshouser.splitstrip *Split strip plot on soybeans*

Description

Split strip plot on soybeans

Format

A data frame with 160 observations on the following 8 variables.

block block factor, 4 levels

plot plot number

cultivar cultivar factor, 4 levels

spacing row spacing

```

pop population (thousand per acre)
yield yield
row row
col column

```

Details

Within each block, cultivars were whole plots. Withing whole plots, spacing was applied in strips vertically, and population was applied in strips horizontally.

Source

Schabenberger, Oliver and Francis J. Pierce. 2002. *Contemporary Statistical Models for the Plant and Soil Sciences*. CRC Press, Boca Raton, FL. Page 493.

Used with permission of David Holshouser at Virginia Polytechnic.

Examples

```

data(holshouser.splitstrip)
dat <- holshouser.splitstrip
dat$spacing <- factor(dat$spacing)
dat$pop <- factor(dat$pop)

# Experiment layout and field trends
if(require(desplot)){
  desplot(yield ~ col*row, data=dat,
          out1=block, # unknown aspect
          main="holshouser.splitstrip")
  desplot(spacing ~ col*row, data=dat,
          out1=block, out2=cultivar, # unknown aspect
          col=cultivar, text=pop, cex=.8, shorten='none', col.regions=c('wheat','white'),
          main="holshouser.splitstrip experiment design")
}

# Overall main effects and interactions
if(require(HH)){
  interaction2wt(yield~cultivar*spacing*pop, dat,
                x.between=0, y.between=0,
                main="holshouser.splitstrip")
}

## Schabenberger's SAS model, page 497
## proc mixed data=splitstripplot;
##   class block cultivar pop spacing;
##   model yield = cultivar spacing spacing*cultivar pop pop*cultivar
##             spacing*pop spacing*pop*cultivar / ddfm=satterth;
##   random block block*cultivar block*cultivar*spacing block*cultivar*pop;
## run;

```

```
## Not run:
## Now lme4. This design has five error terms--four are explicitly given.
require(lme4)
require(lucid)
m1 <- lmer(yield ~ cultivar * spacing * pop +
           (1|block) + (1|block:cultivar) + (1|block:cultivar:spacing) +
           (1|block:cultivar:pop), data=dat)
vc(m1) ## Variances match Schabenberger, page 498.
##           grp           var1 var2  vcov  sdcov
##   block:cultivar:pop (Intercept) <NA> 2.421  1.556
##   block:cultivar:spacing (Intercept) <NA> 1.244  1.116
##           block:cultivar (Intercept) <NA> 0.4523 0.6725
##           block (Intercept) <NA> 3.037  1.743
##           Residual           <NA> <NA> 3.928  1.982

## End(Not run)
```

holtsmark.timothy.uniformity
Uniformity trial of timothy

Description

Uniformity trial of timothy hay circa 1905

Usage

```
data("holtsmark.timothy.uniformity")
```

Format

A data frame with 240 observations on the following 3 variables.

row row

col column

yield yield per plot, kg

Details

Each plot was 5m x 5m = 1/4 are.

Field length: 6 plots * 5 m = 30 m

Field width: 40 plots * 5 m = 200 m.

Holtsmark & Larsen used this trial to compare standard deviations of different sized plots (combined from smaller plots).

Source

Holtsmark, G and Larsen, BR (1905). Om Muligheder for at indskraenke de Fejl, som ved Markforsog betinges af Jordens Uensartethed. *Tidsskrift for Landbrugets Planteavl*. 12, 330-351. (In Danish) <https://books.google.com/books?id=MdM0AQAAAMAJ&pg=PA330> <http://dca.au.dk/publikationer/historiske/planteavl>

Über die Fehler, welche bei Feldversuchen, durch die Ungleichartigkeit des Bodens bedingt werden. *Die Landwirtschaftlichen Versuchs-Stationen*, 65, 1–22. (In German) <https://books.google.com/books?id=eXA2AQAAAMAA>

References

Theodor Roemer (1920). *Der Feldversuch*.

Examples

```
data(holtsmark.timothy.uniformity)
dat <- holtsmark.timothy.uniformity

# Define diagonal 'check' plots like Holtsmark does
dat <- transform(dat,
                 check = ifelse(floor((row+col)/3)==(row+col)/3, "C", ""))

if(require(desplot)){
  desplot(yield ~ col*row, data=dat, flip=TRUE, text=check, show.key=FALSE,
         aspect=30/200, # true aspect
         main="holtsmark.timothy.uniformity")
}

sd(dat$yield) # 2.92 matches Holtsmark p. 348
```

huehn.wheat

Multi-environment trial of wheat to illustrate stability statistics

Description

Multi-environment trial to illustrate stability statistics

Usage

```
data("huehn.wheat")
```

Format

A data frame with 200 observations on the following 3 variables.

```
gen genotype
env environment
yield yield dt/ha
```

Details

Yields for a winter-wheat trial of 20 genotypes at 10 environments.

Note: Huehn 1979 does not use genotype-centered data when calculating stability statistics.

Source

Manfred Huehn (1979). Beitrage zur Erfassung der phanotypischen Stabilitat I. Vorschlag einiger auf Ranginformationen beruhenden Stabilitatsparameter. *EDV in Medizin und Biologie*, 10 (4), 112-117. Table 1. <http://nbn-resolving.de/urn:nbn:de:bsz:15-qucosa-145979>

References

Nassar, R and Huehn, M. (1987). Studies on estimation of phenotypic stability: Tests of significance for nonparametric measures of phenotypic stability. *Biometrics*, 43, 45-53.

Examples

```
data(huehn.wheat)
dat <- huehn.wheat

# Nassar & Huehn, p. 51 "there is no evidence for differences in stability
# among the 20 varieties".
if(require(gge)){
  m1 <- gge(yield ~ gen*env, dat)
  biplot(m1, main="huehn.wheat")
}

## Not run:

require(reshape2)
datm <- acast(dat, gen~env, value.var='yield')

apply(datm,1,mean) # Gen means match Huehn 1979 table 1
apply(datm,2,mean) # Env means
apply(datm, 2, rank) # Ranks match Huehn table 1

# Huehn 1979 did not use genotype-centered data, and his definition
# of S2 is different from later papers.
if(require(stable)){
  apply(huehn(datm, corrected=FALSE), 2, round,2) # S1 matches Huehn
##           MeanRank  S1
## Jubilar      6.70 3.62
## Diplomat     8.35 5.61
## Caribo      11.20 6.07
## Cbc710      13.65 6.70

# Very close match to Nassar & Huehn 1987 table 4.
apply(huehn(datm, corrected=TRUE), 2, round,2)
##           MeanRank  S1  Z1  S2  Z2
## Jubilar      10.2 4.00 5.51 11.29 4.29
```

```
## Diplomat      11.0 6.31 0.09 27.78 0.27
## Caribo        10.6 6.98 0.08 34.49 0.01
## Cbc710        10.9 8.16 1.78 47.21 1.73

}

## End(Not run)
```

hughes.grapes

Disease incidence on grape leaves in RCB experiment

Description

Disease incidence on grape leaves in RCB experiment with 6 different treatments.

Format

A data frame with 270 observations on the following 6 variables.

block Block factor, 1-3

trt Treatment factor, 1-6

vine Vine factor, 1-3

shoot Shoot factor, 1-5

di.seased Number of diseased leaves per shoot

total Number of total leaves per shoot

Details

These data come from a study of downy mildew on grapes. The experiment was conducted at Wooster, Ohio, on the experimental farm of the Ohio Agricultural Research and Development Center, Ohio State University.

There were 3 blocks with 6 treatments. Treatment 1 is the unsprayed control. On 30 Sep 1990, disease incidence was measured. For each plot, 5 randomly chosen shoots on each of the 3 vines were observed. The canopy was closed and shoots could be intertwined. On each shoot, the total number of leaves and the number of infected leaves were recorded.

Source

Hughes, G. and Madden, LV. 1995. Some methods allowing for aggregated patterns of disease incidence in the analysis of data from designed experiments. *Plant Pathology*, 44, 927–943. <http://doi.org/10.1111/j.1365-3059.1995.tb02651.x>

Used with permission of Larry Madden.

References

Hans-Pieter Piepho. 1999. Analysing disease incidence data from designed experiments by generalized linear mixed models. *Plant Pathology*, 48, 668–684. <http://doi.org/10.1046/j.1365-3059.1999.00383.x>

Examples

```

data(hughes.grapes)
dat <- hughes.grapes

dat <- transform(dat, rate = diseased/total, plot=trt:block)

# Trt 1 has higher rate, more variable, Trt 3 lower rate, less variable
require(lattice)
foo <- bwplot(rate ~ vine|block*trt, dat, main="hughes.grapes",
              xlab="vine")
if(require(latticeExtra)){
  useOuterStrips(foo)
}

# Table 1 of Piepho 1999
tapply(dat$rate, dat$trt, mean) # trt 1 does not match Piepho
tapply(dat$rate, dat$trt, max)

# -----
## Not run:
# Piepho model 3. Binomial data. May not be exactly the same model

# Use the binomial count data with lme4
require(lme4)
m1 <- glmer(cbind(diseased, total-diseased) ~ trt + block + (1|plot/vine),
            data=dat, family=binomial)
m1

# Switch from binomial counts to bernoulli data
require(aod)
bdat <- splitbin(cbind(diseased, total-diseased) ~ block+trt+plot+vine+shoot,
                 data=dat)$tab
names(bdat)[2] <- 'y'

# Using lme4
m2 <- glmer(y ~ trt + block + (1|plot/vine), data=bdat, family=binomial)
m2

# Now using MASS::glmPQL
require(MASS)
m3 <- glmPQL(y ~ trt + block, data=bdat,
             random=~1|plot/vine, family=binomial)
m3

## End(Not run)

```

hunter.corn

Multi-environment trial of corn with nitrogen fertilizer

Description

Corn yield response to nitrogen

Format

A data frame with 54 observations on the following 4 variables.

nitro nitrogen fertilizer, pound/acre

year year

loc location

yield yield, bu/ac

Details

Experiments were conducted in eastern Oregon during the years 1950-1952.

Planting rates varied from 15,000 to 21,000 planter per acre.

Source

Albert S. Hunter, John A. Yungen (1955). The Influence of Variations in Fertility Levels Upon the Yield and Protein Content of Field Corn in Eastern Oregon. *Soil Science Society of America Journal*, 19, 214-218. <http://doi.org/10.2136/sssaj1955.03615995001900020027x>

References

James Leo Paschal, Burton Leroy French (1956). A method of economic analysis applied to nitrogen fertilizer rate experiments on irrigated corn. Tech Bull 1141. United States Dept of Agriculture. books.google.com/books?id=gAdZtsEziCcC&pg=PP1

Examples

```
data(hunter.corn)
dat <- hunter.corn

dat <- transform(dat, env=factor(paste(loc,year)))

require(lattice)
xyplot(yield~nitro|env, dat, type='b',
main="hunter.corn - nitrogen response curves")
```

ilri.sheep

*Birth weight and weaning weight of Dorper x Red Maasi lambs***Description**

Birth weight and weaning weight of 882 lambs from a partial diallel cross of Dorper and Red Maasi breeds.

Format

A data frame with 882 observations on the following 12 variables.

year year of lamb birth, 1991-1996

lamb lamb id

sex sex of lamb, M=Male/F=Female

gen genotype, DD, DR, RD, RR

birthwt weight of lamb at birth, kg

weanwt weight of lamb at weaning, kg

weanage age of lamb at weaning, days

ewe ewe id

ewegen ewe genotype: D, R

damage ewe (dam) age in years

ram ram id

ramgen ram genotype: D, R

Details

Red Maasai sheep in East Africa are perceived to be resistant to certain parasites. ILRI decided in 1990 to investigate the degree of resistance exhibited by this Red Maasai breed and initiated a study in Kenya. A susceptible breed, the Dorper, was chosen to provide a direct comparison with the Red Maasai. The Dorper is well-adapted to this area and is also larger than the Red Maasai, and this makes these sheep attractive to farmers.

Throughout six years from 1991 to 1996 Dorper (D), Red Maasai (R) and Red Maasai x Dorper crossed ewes were mated to Red Maasai and Dorper rams to produce a number of different lamb genotypes. For the purposes of this example, only the following four offspring genotypes are considered (Sire x Dam): D x D, D x R, R x D and R x R.

Records are missing in 182 of the lambs, mostly because of earlier death.

Source

Case Study 4: Mixed model analysis for the estimation of components of genetic variation in lamb weaning weight. International Livestock Research Institute. <http://www.ilri.org/biometrics/CS/case%20study%204/case%20> Retrieved Dec 2011.

Licensed with Creative Commons BY-NC-SA 3.0 Unported license.

References

Baker, RL and Nagda, S. and Rodriguez-Zas, SL and Southey, BR and Audho, JO and Aduda, EO and Thorpe, W. 2003. Resistance and resilience to gastro-intestinal nematode parasites and relationships with productivity of Red Maasai, Dorper and Red Maasai x Dorper crossbred lambs in the sub-humid tropics. *Animal Science*, 76, 119-136.

Examples

```

data(ilri.sheep)
dat <- ilri.sheep
dat <- transform(dat, lamb=factor(lamb), ewe=factor(ewe), ram=factor(ram),
                 year=factor(year))
# dl is linear covariate, same as damage, but truncated to [2,8]
dat <- within(dat, {
  dl <- damage
  dl <- ifelse(dl < 3, 2, dl)
  dl <- ifelse(dl > 7, 8, dl)
  dq <- dl^2
})

dat <- subset(dat, !is.na(weanage))

# EDA
require(lattice)
## bwplot(weanwt ~ year, dat, main="ilri.sheep", xlab="year", ylab="Wean weight",
##       panel=panel.violin) # Year effect
bwplot(weanwt ~ factor(dl), dat,
       main="ilri.sheep", xlab="Dam age", ylab="Wean weight") # Dam age effect
# bwplot(weanwt ~ gen, dat,
#       main="ilri.sheep", xlab="Genotype", ylab="Wean weight") # Genotype differences
xyplot(weanwt ~ weanage, dat, type=c('p','smooth'),
       main="ilri.sheep", xlab="Wean age", ylab="Wean weight") # Age covariate

# case study page 4.18
lm1 <- lm(weanwt ~ year + sex + weanage + dl + dq + ewegen + ramgen, data=dat)
summary(lm1)
anova(lm1)

# -----

## Not run:
require(lme4)
lme1 <- lmer(weanwt ~ year + sex + weanage + dl + dq + ewegen + ramgen +
            (1|ewe) + (1|ram), data=dat)
print(lme1, corr=FALSE)
lme2 <- lmer(weanwt ~ year + sex + weanage + dl + dq + ewegen + ramgen +
            (1|ewe), data=dat)
lme3 <- lmer(weanwt ~ year + sex + weanage + dl + dq + ewegen + ramgen +
            (1|ram), data=dat)
anova(lme1, lme2, lme3)

```

```
## End(Not run)

# -----

## Not run:
# asreml3
require(asreml)
# case study page 4.20
m1 <- asreml(weanwt ~ year + sex + weanage + dl + dq + ramgen + ewegen,
             data=dat)
anova(m1)

# case study page 4.26
m2 <- asreml(weanwt ~ year + sex + weanage + dl + dq + ramgen + ewegen,
             random = ~ ram + ewe, data=dat)
anova(m2)

# case study page 4.37, year means
predict(m2, data=dat, classify="year")$predictions

## End(Not run)

# -----

## Not run:
## require(asreml4)
## # case study page 4.20
## m1 <- asreml(weanwt ~ year + sex + weanage + dl + dq + ramgen + ewegen,
##             data=dat)
## wald(m1)

## # case study page 4.26
## m2 <- asreml(weanwt ~ year + sex + weanage + dl + dq + ramgen + ewegen,
##             random = ~ ram + ewe, data=dat)
## wald(m2)

## # case study page 4.37, year means
## predict(m2, data=dat, classify="year")

## End(Not run)
```

immer.sugarbeet.uniformity

Uniformity trial of sugarbeets, measurements of yield, sugar, purity

Description

Uniformity trial of sugarbeets, at Minnesota, 1930, with measurements of yield, sugar, purity.

Format

A data frame with 600 observations on the following 5 variables.

```
row row
col column
yield yield, pounds
sugar sugar percentage
purity apparent purity
```

Details

Beets were planted in rows 22 inches apart, thinned to 1 plant per row. At harvest, the rows were marked into segments 33 feet long with 2 foot alleys between ends of plots. The harvested area was 60 rows 350 feet long.

Field width: 10 plots * 33 feet + 9 alleys * 2 feet = 348 feet

Field length: 60 plots/rows * 22 in/row / 12 in/feet = 110 feet

Planted in 1930. Field conditions were uniform. Beets were planted in rows 22 inches apart. After thinning, one beet was left in each 12-inch unit. At harvest, the field was marked out in plot 33 feet long, with a 2-foot alley between plots to minimize carover from the harvester. A sample of 10 beets was taken uniformly (approximately every third beet) and measured for sugar percentage and apparent purity. The beets were counted at weighing time and the yields were calculated on the basis of 33 beets per plot.

Immer found that aggregating the data from one row to two resulted in a dramatic reduction in the standard error (for yield).

Source

F. R. Immer. 1932. Size and shape of plot in relation to field experiments with sugar beets. *Jour. Agr. Research*, 44, 649–668. naldc.nal.usda.gov/download/IND43968078/PDF

Examples

```
data(immer.sugarbeet.uniformity)
dat <- immer.sugarbeet.uniformity

# Immer numbers rows from the top
if(require(desplot)){
  desplot(yield~col*row, data=dat,
          aspect=110/348, tick=TRUE, flip=TRUE, # true aspect
          main="immer.sugarbeet.uniformity - yield")
  desplot(sugar~col*row, data=dat,
          aspect=110/348, tick=TRUE, flip=TRUE,
          main="immer.sugarbeet.uniformity - sugar")
  desplot(purity~col*row, data=dat,
          aspect=110/348, tick=TRUE, flip=TRUE,
          main="immer.sugarbeet.uniformity - purity")
}
```

```
pairs(dat[,c('yield', 'sugar', 'purity')], main="immer.sugarbeet.uniformity")
```

 ivins.herbs

Percent ground cover of herbage species and nettles.

Description

Percent ground cover of herbage species and nettles.

Format

A data frame with 78 observations on the following 4 variables.

block block, 6 levels

gen genotype, 13 levels

nettle percent ground cover of nettles

herb percent ground cover of herbage species

Details

On the University of Nottingham farm, 13 different strains and species of herbage plants were sown on about 4 acres in an RCB design. Each grass species was sown together with white clover seed.

During establishment of the herbage plants, it became apparent that *Urtica dioica* (nettle) became established according to the particular herbage plant in each plot. In particular, nettle became established in plots sown with leguminous species and the two grass species. The graminaceous plots had less nettles.

The data here are the percentage ground cover of nettle and herbage plants in September 1951.

Note, some of the percent ground cover amounts were originally reported as 'trace'. These have been arbitrarily set to 0.1 in this data.

| gen | species | strain |
|-----|----------------------------|-----------------------------------|
| G01 | <i>Lolium perenne</i> | Irish perennial ryegrass |
| G02 | <i>Lolium perenne</i> | S. 23 perennial ryegrass |
| G03 | <i>Dactylis glomerata</i> | Danish cocksfoot |
| G04 | <i>Dactylis glomerata</i> | S. 143 cocksfoot |
| G05 | <i>Phleum pratense</i> | American timothy |
| G06 | <i>Phleum pratense</i> | S. 48 timothy |
| G07 | <i>Festuca pratensis</i> | S. 215 meadow fescue |
| G08 | <i>Poa trivialis</i> | Danish rough stalked meadow grass |
| G09 | <i>Cynosurus cristatus</i> | New Zealand crested dogstail |
| G10 | <i>Trifolium pratense</i> | Montgomery late red clover |
| G11 | <i>Medicago lupulina</i> | Commercial black medick |
| G12 | <i>Trifolium repens</i> | S. 100 white clover |
| G13 | <i>Plantago lanceolata</i> | Commercial ribwort plantain |

Source

Ivins, JD. (1952). Concerning the Ecology of *Urtica Dioica* L., *Journal of Ecology*, 40, 380-382. <http://doi.org/10.2307/2256806>

References

Ivins, JD (1950). Weeds in relation to the establishment of the Ley. *Grass and Forage Science*, 5, 237–242. <http://doi.org/10.1111/j.1365-2494.1950.tb01287.x>

O’Gorman, T.W. (2001). A comparison of the F-test, Friedman’s test, and several aligned rank tests for the analysis of randomized complete blocks. *Journal of agricultural, biological, and environmental statistics*, 6, 367–378. <http://doi.org/10.1198/108571101317096578>

Examples

```
data(ivins.herbs)
dat <- ivins.herbs

# Nettle is primarily established in legumes.
require(lattice)
xyplot(herb~nettle|gen, dat, main="ivins.herbs - herb yield vs weeds",
       xlab="Percent groundcover in nettles",
       ylab="Percent groundcover in herbs")

# O'Brien used first 7 species to test gen differences
dat7 <- droplevels(subset(dat, is.element(gen, c('G01','G02','G03','G04','G05','G06','G07'))))
m1 <- lm(herb ~ gen + block, data=dat7)
anova(m1) # gen p-value is .041
## Response: herb
##           Df  Sum Sq Mean Sq F value Pr(>F)
## gen         6 1083.24  180.540   2.5518 0.04072 *
## block       5  590.69  118.138   1.6698 0.17236
## Residuals 30 2122.48   70.749

friedman.test(herb ~ gen|block, dat7) # gen p-value .056
```

iyer.wheat.uniformity *Uniformity trials of wheat in India.*

Description

Uniformity trials of wheat in India.

Usage

```
data("iyer.wheat.uniformity")
```

Format

A data frame with 2000 observations on the following 3 variables.

```
row row
col column
yield yield, ounces per plot
```

Details

Data collected at the Agricultural Sub-station in Karnal, India, in April 1978. A net area of 400 ft x 125 ft was harvested by dividing it into 80x25 units 5 ft x 5 ft after eliminating a minimum border of 3.5 ft all around the net area.

Field width: 80 plots * 5 feet = 400 feet Field length: 25 rows * 5 feet = 125 feet

In a second paper, Iyer used this data to compare random vs. balanced arrangements of treatments to plots, with the conclusion that "it is very difficult to say which [method] is better. However, there is some tendency for the randomized arrangements to give more accurate results."

Source

P. V. Krishna Iyer (1942). Studies with wheat uniformity trial data. I. Size and shape of experimental plots and the relative efficiency of different layouts. *The Indian Journal of Agricultural Science*, 12, 240-262. Page 259-262. <https://archive.org/stream/in.ernet.dli.2015.7638/2015.7638.The-Indian-Journal-Of-Agricultural-Science-Vol-xii-1942#page/n267/mode/2up>

References

None.

Examples

```
data(iyer.wheat.uniformity)
dat <- iyer.wheat.uniformity

if(require(desplot)){
  desplot(yield ~ col*row, data=dat,
          main="iyer.wheat.uniformity", tick=TRUE,
          aspect=(25*5)/(80*5)) # true aspect
}

# not exactly the same as Iyer table 1, p. 241
var(subset(dat, col <= 20)$yield)
var(subset(dat, col > 20 & col <= 40)$yield)
var(subset(dat, col > 40 & col <= 60)$yield)
var(subset(dat, col > 60)$yield)

# cv for 1x1 whole-field
# sd(dat$yield)/mean(dat$yield)
# 18.3
```

jansen.apple

Infestation of apple shoots by apple canker.

Description

Infestation of apple shoots by apple canker.

Usage

```
data("jansen.apple")
```

Format

A data frame with 36 observations on the following 5 variables.

inoculum inoculum level

gen genotype/variety

block block

y number of inoculations developing canker

n number of inoculations

Details

Shoots of apple trees were infected with fungus *Nectria galligena*, which may cause apple canker.

The inoculum density treatment had 3 levels, measured in macroconidia per ml.

There were 4 blocks.

Source

J. Jansen & J.A. Hoekstra (1993). The analysis of proportions in agricultural experiments by a generalized linear mixed model. *Statistica Neerlandica*, 47(3), 161-174. <http://doi.org/10.1111/j.1467-9574.1993.tb01414.x>

Used with permission of J. Jansen. Electronic version supplied by Miroslav Zoric.

References

None.

Examples

```
data(jansen.apple)
dat <- jansen.apple
```

```
require(lattice)
dotplot(gen ~ y/n|block, data=dat, group=inoculum, auto.key=TRUE,
        main="jansen.apple",
```

```

      xlab="Proportion infected per block/inoculum", ylab="Genotype")

## Not run:
  require(lme4)
  # Tentative model. Needs improvement.
  m1 <- glmer(cbind(y,n-y) ~ gen + factor(inoculum) + (1|block),
             data=dat, family=binomial)
  summary(m1)

## End(Not run)

```

jansen.carrot

Infestation of carrots by fly larvae

Description

Infestation of 16 carrot genotypes by fly larvae, comparing 2 treatments in 16 blocks.

Usage

```
data("jansen.carrot")
```

Format

A data frame with 96 observations on the following 5 variables.

trt treatment

gen genotype

block block

n number of carrots sampled per plot

y number of carrots infested per plot

Details

This experiment was designed to compare different genotypes of carrots with respect to their resistance to infestation by larvae of the carrotfly.

There were 16 genotypes, 2 levels of pest-control treatments, conducted in 3 randomized complete blocks. About 50 carrots were sampled from each plot and evaluated. The data show the number of carrots and the number infested by fly larvae.

Source

J. Jansen & J.A. Hoekstra (1993). The analysis of proportions in agricultural experiments by a generalized linear mixed model. *Statistica Neerlandica*, 47(3), 161-174. <http://doi.org/10.1111/j.1467-9574.1993.tb01414.x>

Used with permission of J. Jansen. Electronic version supplied by Miroslav Zoric.

References

None.

Examples

```
data(jansen.carrot)
dat <- jansen.carrot

require(lattice)
dotplot(gen ~ y/n, data=dat, group=trt, auto.key=TRUE,
        main="jansen.carrot",
        xlab="Proportion of carrots infected per block", ylab="Genotype")

## Not run:
# Not run because CRAN wants < 5 seconds per example. This is close.
require(lme4)
# Tentative model. Needs improvement.
m1 <- glmer(cbind(y,n-y) ~ gen*trt + (1|block),
            data=dat, family=binomial)
summary(m1)
# Todo: Why are these results different from Jansen?
# Maybe he used ungrouped bernoulli data? Too slow with 4700 obs

## End(Not run)
```

jansen.strawberry *Ordered disease ratings on strawberry crosses.*

Description

Ordered disease ratings on strawberry crosses.

Usage

```
data("jansen.strawberry")
```

Format

A data frame with 144 observations on the following 5 variables.

male male parent

female female parent

block block

category disease damage, C1 < C2 < C3

count number of plants in each category

Details

In strawberries, red core disease is caused by a fungus, *Phytophthora fragariae*. This experiment evaluated different populations for damage caused by red core disease.

There were 3 male strawberry plants and 4 DIFFERENT female strawberry plants that were crossed to create 12 populations. Note: Jansen labeled the male parents 1,2,3 and the female parents 1,2,3,4. To reduce confusion, this data labels the female parents 5,6,7,8.

The experiment had four blocks with 12 plots each (one for each population). Plots usually had 10 plants, but some plots only had 9 plants. Each plant was assessed for damage from fungus and rated as belonging to category C1, C2, or C3 (increasing damage).

Source

J. Jansen, 1990. On the statistical analysis of ordinal data when extravariation is present. *Applied Statistics*, 39, 75-84, Table 1. <http://doi.org/10.2307/2347813>

Used with permission of Hans Jansen.

Examples

```
data(jansen.strawberry)
dat <- jansen.strawberry
dat <- transform(dat, category=ordered(category, levels=c('C1','C2','C3')))

dtab <- xtabs(count ~ male + female + category, data=dat)
ftable(dtab)

mosaicplot(dtab,
            color=c("lemonchiffon1","lightsalmon1","indianred"),
            main="jansen.strawberry disease ratings",
            xlab="Male parent", ylab="Female parent")

if(require(MASS) & require(vcd)){
  # Friendly suggests a minimal model is [MF][C]
  # m1 <- loglm( ~ 1*2 + 3, dtab) # Fails, only with devtools
  # mosaic(m1)
}
```

jenkyn.mildew

Yields from treatment for mildew control

Description

Yields from treatment for mildew control

Format

A data frame with 38 observations on the following 4 variables.

plot plot number
trt treatment factor, 4 levels
block block factor, 9 levels
yield grain yield, tons/ha

Details

There were four spray treatments: 0 (none), 1 (early), 2 (late), R (repeated).

Each treatment occurs once between each of the 9 ordered pairs of the other treatments.

The first and last plot are not assigned to a block.

Source

Norman Draper and Irwin Guttman (1980). Incorporating Overlap Effects from Neighboring Units into Response Surface Models. *Appl Statist*, 29, 128–134. <http://doi.org/10.2307/2986297>

References

Maria Durban, Christine Hackett, Iain Currie. Blocks, Trend and Interference in Field Trials.

Examples

```
data(jenkyn.mildew)
dat <- jenkyn.mildew

require(lattice)
bwplot(yield ~ trt, dat, main="jenkyn.mildew", xlab="Treatment")

# Residuals from treatment model show obvious spatial trends
m0 <- lm(yield ~ trt, dat)
xyplot(resid(m0)~plot, dat, ylab="Residual",
       main="jenkyn.mildew - treatment model")

# The blocks explain most of the variation
m1 <- lm(yield ~ trt + block, dat)
xyplot(resid(m1)~plot, dat, ylab="Residual",
       main="jenkyn.mildew - block model")
```

john.alpha

Alpha lattice design of spring oats

Description

Alpha lattice design of spring oats

Format

A data frame with 72 observations on the following 5 variables.

plot plot number

rep replicate

block incomplete block

gen genotype (variety)

yield dry matter yield (tonnes/ha)

Details

A spring oats trial grown in Craibstone, near Aberdeen. There were 24 varieties in 3 replicates, each consisting of 6 incomplete blocks of 4 plots. Planted in a resolvable alpha design. The plots were laid out in a single line.

Source

J. A. John & E. R. Williams (1995). *Cyclic and computer generated designs*, Chapman and Hall, London. Page 146.

References

Piepho, H.P. and Mohring, J. (2007), Computing heritability and selection response from unbalanced plant breeding trials, *Genetics*, 177, 1881-1888. <http://doi.org/10.1534/genetics.107.074229>

Examples

```
data(john.alpha)
dat <- john.alpha

# RCB (no incomplete block)
m0 <- lm(yield ~ 0 + gen + rep, data=dat)

# Block fixed (intra-block analysis) (bottom of table 7.4 in John)
m1 <- lm(yield ~ 0 + gen + rep + rep:block, dat)
anova(m1)

# Block random (combined inter-intra block analysis)
if(require(lme4) & require(lucid)){
```

```

m2 <- lmer(yield ~ 0 + gen + rep + (1|rep:block), dat)

anova(m2)
## Analysis of Variance Table
##      Df Sum Sq Mean Sq  F value
## gen  24  380.43  15.8513  185.9942
## rep   2    1.57   0.7851   9.2123
vc(m2)
##      grp      var1 var2   vcov  sdcor
## rep:block (Intercept) <NA> 0.06194 0.2489
## Residual              <NA> <NA> 0.08523 0.2919

# Variety means. John and Williams table 7.5. Slight, constant
# difference for each method as compared to John and Williams.
means <- data.frame(rcb=coef(m0)[1:24],
                   ib=coef(m1)[1:24],
                   intra=fixef(m2)[1:24])

head(means)
##      rcb      ib  intra
## genG01 5.201233 5.268742 5.146433
## genG02 4.552933 4.665389 4.517265
## genG03 3.381800 3.803790 3.537934
## genG04 4.439400 4.728175 4.528828
## genG05 5.103100 5.225708 5.075944
## genG06 4.749067 4.618234 4.575394

require(lattice)
splom(means, main="john.alpha - means for RCB, IB, Intra-block")

}

# -----

## Not run:
# asreml3
# Heritability calculation of Piepho & Mohring, Example 1
require(asreml)
m3 <- asreml(yield ~ 1 + rep, data=dat, random=~ gen + rep:block)
sg2 <- summary(m3)$varcomp['gen!gen.var', 'component'] # .142902

# Average variance of a difference of two adjusted means (BLUP)

p3 <- predict(m3, data=dat, classify="gen", sed=TRUE)
# Matrix of pair-wise SED values, squared
vdiff <- p3$predictions$sed^2
# Average variance of two DIFFERENT means (using lower triangular of vdiff)
vblup <- mean(vdiff[lower.tri(vdiff)]) # .05455038

# Note that without sed=TRUE, asreml reports square root of the average variance
# of a difference between the variety means, so the following gives the same value
# predict(m3, data=dat, classify="gen")$pred$avsed ^ 2 # .05455038

```

```

# Average variance of a difference of two adjusted means (BLUE)
m4 <- asreml(yield ~ 1 + gen + rep, data=dat, random = ~ rep:block)
p4 <- predict(m4, data=dat, classify="gen", sed=TRUE)
vdiff <- p4$predictions$sed^2
vblue <- mean(vdiff[lower.tri(vdiff)]) # .07010875
# Again, could use predict(m4, data=dat, classify="gen")$pred$avsed ^ 2

# H^2 Ad-hoc measure of heritability
sg2 / (sg2 + vblue/2) # .803

# H^2c Similar measure proposed by Cullis.
1-(vblup / 2 / sg2) # .809

## End(Not run)

# -----

## Not run:
## # Heritability calculation of Piepho & Mohring, Example 1
## require(asreml4)
## m3 <- asreml(yield ~ 1 + rep, data=dat, random=~ gen + rep:block)
## sg2 <- summary(m3)$varcomp['gen','component'] # .142902

## # Average variance of a difference of two adjusted means (BLUP)

## p3 <- predict(m3, data=dat, classify="gen", sed=TRUE)
## # Matrix of pair-wise SED values, squared
## vdiff <- p3$sed^2
## # Average variance of two DIFFERENT means (using lower triangular of vdiff)
## vblup <- mean(vdiff[lower.tri(vdiff)]) # .05455038

## # Note that without sed=TRUE, asreml reports square root of the average variance
## # of a difference between the variety means, so the following gives the same value
## # predict(m3, data=dat, classify="gen")$pred$avsed ^ 2 # .05455038

## # Average variance of a difference of two adjusted means (BLUE)
## m4 <- asreml(yield ~ 1 + gen + rep, data=dat, random = ~ rep:block)
## p4 <- predict(m4, data=dat, classify="gen", sed=TRUE)
## vdiff <- p4$sed^2
## vblue <- mean(vdiff[lower.tri(vdiff)]) # .07010875
## # Again, could use predict(m4, data=dat, classify="gen")$pred$avsed ^ 2

## # H^2 Ad-hoc measure of heritability
## sg2 / (sg2 + vblue/2) # .803

## # H^2c Similar measure proposed by Cullis.
## 1-(vblup / 2 / sg2) # .809

## End(Not run)

# -----

## Not run:

```

```

# Illustrate how to do the same calculations with lme4
# http://stackoverflow.com/questions/38697477

require(lme4)

cov2sed <- function(x){
  # Convert var-cov matrix to SED matrix
  # sed[i,j] = sqrt( x[i,i] + x[j,j]- 2*x[i,j] )
  n <- nrow(x)
  vars <- diag(x)
  sed <- sqrt( matrix(vars, n, n, byrow=TRUE) +
                  matrix(vars, n, n, byrow=FALSE) - 2*x )
  diag(sed) <- 0
  return(sed)
}

# Same as asreml model m4. Note 'gen' must be first term
m5blue <- lmer(yield ~ 0 + gen + rep + (1|rep:block), dat)

require(lsmmeans)
ls5blue <- lsmmeans(m5blue, "gen")
con <- ls5blue@linfoct[,1:24] # contrast matrix for genotypes
# The 'con' matrix is identity diagonal, so we don't need to multiply,
# but do so for a generic approach
# sed5blue <- cov2sed(con)
tmp <- tcrossprod( crossprod(t(con), vcov(m5blue)[1:24,1:24]), con)
sed5blue <- cov2sed(tmp)

# vblue Average variance of difference between genotypes
vblue <- mean(sed5blue[upper.tri(sed5blue)]^2) # .07010875 matches 'vblue' from asreml

# Now blups
m5blup <- lmer(yield ~ 0 + (1|gen) + rep + (1|rep:block), dat)
re5 <- ranef(m5blup,condVar=TRUE)
vv1 <- attr(re5$gen,"postVar")
vblup <- 2*mean(vv1) # .0577 not exactly same as 'vblup' above

# H^2 Ad-hoc measure of heritability
sg2 <- c(lme4::VarCorr(m5blup)[["gen"]]) # 0.142902
sg2 / (sg2 + vblue/2) # .803 matches asreml

# H^2c Similar measure proposed by Cullis.
1-(vblup / 2 / sg2) # .809 from asreml, .800 from lme4

## End(Not run)

# -----

```

`johnson.blight`*Potato blight due to weather in Prosser, Washington*

Description

Potato blight due to weather in Prosser, Washington

Format

A data frame with 25 observations on the following 6 variables.

`year` year

`area` area affected, hectares

`blight` blight detected, 0/1 numeric

`rain.am` number of rainy days in April and May

`rain.ja` number of rainy days in July and August

`precip.m` precipitation in May when temp > 5C, milimeters

Details

The variable 'blight detected' is 1 if 'area' > 0.

Source

Johnson, D.A. and Alldredge, J.R. and Vakoch, D.L. (1996). Potato late blight forecasting models for the semiarid environment of south-central Washington. *Phytopathology*, 86, 480–484. <http://doi.org/10.1094/Phyto-86-480>

References

Vinayanand Kandala, *Logistic Regression*

Examples

```
data(johnson.blight)
dat <- johnson.blight

# Define indicator for blight in previous year
dat$blight.prev[2:25] <- dat$blight[1:24]
dat$blight.prev[1] <- 0 # Need this to match the results of Johnson
dat$blight.prev <- factor(dat$blight.prev)
dat$blight <- factor(dat$blight)

# Johnson et al developed two logistic models to predict outbreak of blight

m1 <- glm(blight ~ blight.prev + rain.am + rain.ja, data=dat, family=binomial)
```

```

summary(m1)
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) -11.4699     5.5976  -2.049  0.0405 *
## blight.prev1  3.8796     1.8066   2.148  0.0318 *
## rain.am       0.7162     0.3665   1.954  0.0507 .
## rain.ja       0.2587     0.2468   1.048  0.2945
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## (Dispersion parameter for binomial family taken to be 1)

## Null deviance: 34.617 on 24 degrees of freedom
## Residual deviance: 13.703 on 21 degrees of freedom
## AIC: 21.703

m2 <- glm(blight ~ blight.prev + rain.am + precip.m, data=dat, family=binomial)
summary(m2)
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -7.5483     3.8070  -1.983  0.0474 *
## blight.prev1  3.5526     1.6061   2.212  0.0270 *
## rain.am       0.6290     0.2763   2.276  0.0228 *
## precip.m     -0.0904     0.1144  -0.790  0.4295
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## (Dispersion parameter for binomial family taken to be 1)

## Null deviance: 34.617 on 24 degrees of freedom
## Residual deviance: 14.078 on 21 degrees of freedom
## AIC: 22.078

require(lattice)
splom(dat[,c('blight','rain.am','rain.ja','precip.m')],
      main="johnson.blight - indicator of blight")

```

kadam.millet.uniformity

Uniformity trial of millet

Description

Uniformity trial of millet in India during 2 years

Usage

```
data("kadam.millet.uniformity")
```

Format

A data frame with 240 observations on the following 4 variables.

```
year year
row row
col column
yield yield, ounces
```

Details

Uniformity trials conducted during the kharip (monsoon) seasons of 1933 and 1934 at Kundewadi, Niphad, in the district of Nasik, India. Bajari (pearl millet) strain 54 was used.

In 1933:

Field length: 10 plots * 33 feet

Field width: 8 plots * 16.5 feet

In 1934:

Field length: 20 plots * 16.5 feet

Field width: 8 plots * 16.5 feet

Source

B. S. Kadam and S. M. Patel. (1937). Studies in Field-Plot Technique With *P. Typhoideum* Rich. The Empire Journal Of Experimental Agriculture, 5, 219-230. <https://archive.org/details/in.ernet.dli.2015.25282>

References

None.

Examples

```
data(kadam.millet.uniformity)

# similar to Kadam fig 1
dat1 <- subset(kadam.millet.uniformity, year==1933)
if(require(desplot)){
  desplot(yield ~ col*row, dat1, flip=TRUE,
          aspect=(10*33)/(8*16.5), # true aspect
          main="kadam.millet.uniformity 1933")
}

dat2 <- subset(kadam.millet.uniformity, year==1934)
if(require(desplot)){
  desplot(yield ~ col*row, dat2, flip=TRUE,
          aspect=(20*16.5)/(8*16.5), # true aspect
          main="kadam.millet.uniformity 1934")
}
```

kalamkar.potato.uniformity
Uniformity trial of potatoes

Description

Uniformity trial of potatoes at Saskatchewan, Canada, 1929.

Usage

```
data("kalamkar.potato.uniformity")
```

Format

A data frame with 576 observations on the following 3 variables.

```
row row  
col column  
yield yield of potatoes per plot
```

Details

The data is for potato yields in 96 rows, each 132 feet long, with 3 feet between rows.

Each row was harvested as six plots, each 22 feet long. Each hill had one seed piece. Hills were spaced 2 feet apart in each row.

Field width: 6 plots * 22 feet = 132 feet

Field length: 96 rows * 3 feet = 288 feet

Source

Kalamkar, R.J. (1932). Experimental Error and the Field-Plot Technique with Potatoes. *The Journal of Agricultural Science*, 22, 373-385. <http://doi.org/10.1017/S0021859600053697>

Examples

```
data(kalamkar.potato.uniformity)  
dat <- kalamkar.potato.uniformity  
  
# Similar to figure 1 of Kalamkar  
if(require(desplot)){  
  desplot(yield~col*row, data=dat,  
          flip=TRUE, tick=TRUE, aspect=288/132, # true aspect  
          main="kalamkar.potato.uniformity")  
}
```

kalamkar.wheat.uniformity

Uniformity trial of wheat

Description

Uniformity trial of wheat at Rothamsted, UK in 1931.

Usage

```
data("kalamkar.wheat.uniformity")
```

Format

A data frame with 1280 observations on the following 4 variables.

row row

col column

yield yield, grams/half-meter

ears ears per half-meter

Details

Kalamkar's paper published in 1932. Estimated crop year 1931.

Plot 18 of the Four Course Rotation Experiment, Great Hoos, at Rothamsted, UK was used. Sown with Yeoman II wheat.

Field length: 80 rows * 6 inches apart = 40 feet.

Field width = 16 segments * 0.5 meters = 8 meters.

The grain yield and number of ears for each half-meter length were recorded. This is quite a small field, only 1/40 acre in size.

Edge rows have higher yields. Row-end units have higher yields than interior units. These border effects are significant. Variation between rows is greater than variation within rows. Negative correlation between rows may indicate competition effects.

For ears, Kalamkar discarded 4 rows from each side and 3 half-meter lengths at each end.

Kalamkar suggested using four parallel half-meter rows as a sampling unit.

Source

Kalamkar, R. J (1932). A Study in Sampling Technique with Wheat. The Journal of Agricultural Science, Vol.22(4), pp.783-796. <https://doi.org/10.1017/S0021859600054599>

References

None.

Examples

```

data(kalamkar.wheat.uniformity)
dat <- kalamkar.wheat.uniformity
plot(yield ~ ears, dat, main="kalamkar.wheat.uniformity")

# totals match Kalamkar
sum(dat$yield)
sum(dat$ears)

if(require(desplot)){
  desplot(ears ~ col*row, dat,
          flip=TRUE, aspect=(80*0.5)/(16*1.64042), # true aspect
          main="kalamkar.wheat.uniformity - ears")
  desplot(yield ~ col*row, dat,
          flip=TRUE, aspect=(80*0.5)/(16*1.64042), # true aspect
          main="kalamkar.wheat.uniformity - yield")
}

# Show the negative correlation between rows
## require(asreml)
## require(lucid)
## dat <- dat %>% mutate(rowf=factor(row), colf=factor(col)) %>%
##   arrange(rowf,colf)
## m1=asreml(yield ~ 1, data=dat, rcov= ~ ar1(rowf):ar1(colf))
## vc(m1)
## ##      effect component std.error z.ratio constr
## ## R!variance  81.54      3.527      23      P
## ## R!rowf.cor  -0.09465  0.0277     -3.4      U
## ## R!colf.cor   0.2977   0.02628    11      U

```

kang.maize

*Multi-environment trial of maize in Louisiana at 4 locs in 3 years***Description**

Maize yields at 4 locs in 3 years in Louisiana.

Usage

```
data("kang.maize")
```

Format

gen genotype, 17 levels
 env environment, 12 levels
 yield yield, tonnes/ha
 environment, 13 levels

year year, 85-87
 loc location, 4 levels

Details

Yield trials were conducted at four locations (Alexandria, Baton Rouge, Bossier City, and St. Joseph) in Louisiana during 1985 to 1987. Each loc was planted as RCB design with 4 reps. Mean yields are given in this data.

Source

Kang, MS and Gorman, DP. (1989). Genotype x environment interaction in maize. *Agronomy Journal*, 81, 662-664. Table 2.

Used with permission of Dan Gorman.

Examples

```
data(kang.maize)
dat <- kang.maize

# Sweep out loc means, then show interaction plot.
if(require(reshape2)){
  mat <- acast(dat, gen~env, value.var='yield')
  mat <- sweep(mat, 2, colMeans(mat))
  dat2 <- melt(mat)
  names(dat2) <- c('gen','env','yield')
  require(lattice)
  xyplot(yield~env|gen, data=dat2, type='l', group=gen,
         panel=function(x,y,...){
           panel.abline(h=0,col="gray70")
           panel.xyplot(x,y,...)
         },
         ylab="Environment-centered yield",
         main="kang.maize - maize hybrid yields", scales=list(x=list(rot=90)))
}

# Weather covariates for each environment.
covs <- data.frame(env=c("AL85","AL86","AL87", "BR85","BR86","BR87",
                        "BC85","BC86","BC87", "SJ85","SJ86","SJ87"),
                  maxt=c(30.7,30.2,29.7,31.5,29.4,28.5, 31.9, 30.4,31.7, 32,29.6,28.9),
                  mint=c(18.7,19.3,18.5, 19.7,18,17.2, 19.1,20.4,20.3, 20.4,19.1,17.5),
                  rain=c(.2,.34,.22, .28,.36,.61, .2,.43,.2, .36,.41,.22),
                  humid=c(82.8,91.1,85.4, 88.1,90.9,88.6, 95.4,90.4,86.7, 95.6,89.5,85))
```

| | |
|-------------|---|
| kang.peanut | <i>Multi-environment trial of peanuts for 10 genotypes in 15 environments</i> |
|-------------|---|

Description

Peanut yields for 10 genotypes in 15 environments

Usage

```
data("kang.peanut")
```

Format

A data frame with 590 observations on the following 4 variables.

gen genotype factor, 10 levels

rep replicate factor, 4 levels

yield yield

env environment factor, 15 levels

Details

Florman, Tegua, mf484, mf485, mf487, mf489 have a long crop cycle. The others have a short crop cycle.

This data is also likely used in Casanoves et al 2005, "Evaluation of Multienvironment Trials of Peanut Cultivars", but this appears to be a slightly smaller subset (only 10 genotypes, and perhaps only the years 96,97,98,99). Based on the d.f. in their table 5, it appears that environment E13 was grown in 1998. ($5 \text{ loc} * (4-1) = 15$, but the table has 14, and 98-99 had only 3 reps instead of 4 reps.)

Source

Data from National Institute of Agricultural Technology, Argentina.

References

A. Saxton (2004). Genetic Analysis of Complex Traits Using SAS.

Examples

```
data(kang.peanut)
dat <- kang.peanut

# Table 5.1 of Kang et al. (Chapter 5 of Saxton)
if(require(reshape2)){
  round(acast(dat, env~gen, value.var='yield', fun=mean),2)
```

```
}  
  
# GGE biplot of Kang, p. 82.  
if(require(gge)){  
  m1 <- gge(yield~gen*env, data=dat, scale=FALSE)  
  biplot(m1, flip=c(1,1), main="kang.peanut - GGE biplot")  
}
```

karcher.turfgrass *Turfgrass ratings for different treatments*

Description

Turfgrass ratings for different treatments

Format

A data frame with 128 observations on the following 6 variables.

week week number
rep blocking factor
manage management factor, 4 levels
nitro nitrogen factor, 2 levels
rating turfgrass rating, 4 ordered levels
count number of samples for a given rating

Details

Turf color was assessed on a scale of Poor, Average, Good, Excellent.

The data are the number of times that a combination of management style and nitrogen level received a particular rating across four replicates and four sampling weeks. The eight treatments were in a completely randomized design.

Nitrogen level 1 is 2.5 g/m², level 2 is 5 g/m².

Management 1 = N applied with no supplemental water injection.

M2 = surface applied with supplemental water injection.

M3 = nitrogen injected 7.6 cm deep

M4 = nitrogen injected 12.7 cm deep.

Source

Schabenberger, Oliver and Francis J. Pierce. 2002. *Contemporary Statistical Models for the Plant and Soil Sciences*. CRC Press. Page 380.

Examples

```

data(karcher.turfgrass)
dat <- karcher.turfgrass

dat$rating <- ordered(dat$rating, levels=c('Poor', 'Average', 'Good', 'Excellent'))

ftable(xtabs(~manage+nitro+rating, dat)) # Table 6.19 of Schabenberger

# Probably would choose management M3, nitro N2
mosaicplot(xtabs(count ~ manage + rating + nitro, dat),
           shade=TRUE, dir=c('h', 'v', 'h'),
           main="karcher.turfgrass - turfgrass ratings")

# Multinomial logistic model. Probit Ordered Logistic Regression.
require(MASS)
m1 <- polr(rating ~ nitro*manage + week, dat, weights=count, Hess=TRUE, method='logistic')
summary(m1)

# Try to match the "predicted marginal probability distribution" of
# Schabenberger table 6.20. He doesn't define "marginal".
# Are the interaction terms included before aggregation?
# Are 'margins' calculated before/after back-transforming?
# At what level is the covariate 'week' included?

# Here is what Schabenberger presents:
##      M1  M2  M3  M4 |  N1  N2
## Poor  .668 .827 .001 .004 | .279 .020
## Avg   .330 .172 .297 .525 | .712 .826
## Good  .002 .001 .695 .008 | .008 .153
## Exc   .000 .000 .007 .003 | .001 .001

## We use week=3.5, include interactions, then average
newd <- expand.grid(manage=levels(dat$manage), nitro=levels(dat$nitro), week=3.5)
newd <- cbind(newd, predict(m1, newdata=newd, type='probs')) # probs
print(aggregate(. ~ manage, data=newd, mean), digits=2)
##  manage nitro week   Poor Average   Good Excellent
## 1     M1   1.5  3.5 0.67    0.33 0.0011 0.0000023
## 2     M2   1.5  3.5 0.76    0.24 0.00059 0.0000012
## 3     M3   1.5  3.5 0.0023  0.48 0.52   0.0042
## 4     M4   1.5  3.5 0.0086  0.57 0.42   0.0035

```

 keen.potatodamage

Damage to potato tubers from lifting rods.

Description

Damage to potato tubers from lifting rods.

Usage

```
data("keen.potatodamage")
```

Format

A data frame with 1152 observations on the following 6 variables.

energy energy factor

weight weight class

gen genotype/variety factor

rod rod factor

damage damage category

count count of tubers in each combination of categories

Details

Experiments performed at Wageningen, Netherlands.

Potatoes can be damaged by the lifter. In this experiment, eight types of lifting rod were compared. Two energy levels, six genotypes/varieties and three weight classes were used. Most combinations of treatments involved about 20 potato tubers. Tubers were rated as undamaged (D1) to severely damaged (D4).

The main interest is in differences between rods, and not in interactions. The other factors (besides rod) were introduced to create variety in experimental conditions and are not of interest.

Keen and Engle estimated the following rod effects.

```
# Rod: 1 2 3 4 5 6 7 8
```

```
# Effect: 0 -1.26 -0.42 0.55 -1.50 -1.85 -1.76 -2.09
```

Source

A. Keen and B. Engel. Analysis of a mixed model for ordinal data by iterative re-weighted REML. *Statistica Neerlandica*, 51, 129–144. Table 2. <http://doi.org/10.1111/1467-9574.00044>

Used with permission of Bas Engel.

Examples

```
data(keen.potatodamage)
dat <- keen.potatodamage
```

```
# Energy E1, Rod R4, Weight W1 have higher proportions of severe damage
# Rod 8 has the least damage
```

```
d2 <- xtabs(count~energy+rod+gen+weight+damage, data=dat)
mosaicplot(d2, color=c("lemonchiffon1", "moccasin", "lightsalmon1", "indianred"),
            xlab="Energy / Genotype", ylab="Rod / Weight", main="keen.potatodamage")
```

```
## Not run:
```

```
# Not run because it CRAN prefers examples faster than 5 seconds.
```



```

require(ordinal)
# Note, the clmm2 function can have only 1 random term. Results are
# similar to Keen & Engle, but necessarily different (they had multiple
# random terms).
m1 <- clmm2(damage ~ rod + energy + gen + weight, data=dat,
            weights=count, random=rod:energy, link='probit')
round(coef(m1)[4:10],2)
## rodR2 rodR3 rodR4 rodR5 rodR6 rodR7 rodR8
## -1.19 -0.41 0.50 -1.46 -1.73 -1.67 -1.99

# Alternative
# m2 <- clmm(damage ~ rod + energy + gen + weight +
#            (1|rod:energy), data=dat, weights=count, link='probit')

## End(Not run)

```

kempton.barley.uniformity

Uniformity trial of barley

Description

Uniformity trial of barley at Cambridge, England, 1978.

Format

A data frame with 196 observations on the following 3 variables.

row row

col column

yield grain yield, kg

Details

A uniformity trial of spring barley planted in 1978. Conducted by the Plant Breeding Institute in Cambridge, England.

Each plot is 5 feet wide, 14 feet long.

Field width: 7 plots * 14 feet = 98 feet

Field length: 28 plots * 5 feet = 140 feet

Source

R. A. Kempton and C. W. Howes (1981). The use of neighbouring plot values in the analysis of variety trials. *Applied Statistics*, 30, 59–70. <http://doi.org/10.2307/2346657>

References

McCullagh, P. and Clifford, D., (2006). Evidence for conformal invariance of crop yields, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science. 462, 2119–2143.

Examples

```

data(kempton.barley.uniformity)
dat <- kempton.barley.uniformity

if(require(desplot)){
  desplot(yield~col*row, dat,
          aspect=140/98, tick=TRUE, # true aspect
          main="kempton.barley.uniformity")
}

# -----

## Not run:

# Kempton estimated auto-regression coefficients b1=0.10, b2=0.91

dat <- transform(dat, xf = factor(col), yf=factor(row))
# asreml3
require(asreml)
m1 <- asreml(yield ~ 1, data=dat, rcov=~ar1(xf):ar1(yf))

require(lucid)
vc(m1)
##      effect component std.error z.ratio constr
## R!variance    0.1044   0.022     4.7   pos
## R!xf.cor      0.2458   0.07484    3.3  uncon
## R!yf.cor      0.8186   0.03822    21   uncon

# asreml estimates auto-regression correlations of 0.25, 0.82
# Kempton estimated auto-regression coefficients b1=0.10, b2=0.91

## End(Not run)

# -----

## Not run:

# Kempton estimated auto-regression coefficients b1=0.10, b2=0.91

## dat <- transform(dat, xf = factor(col), yf=factor(row))
## require(asreml4)
## m1 <- asreml(yield ~ 1, data=dat, residual = ~ar1(xf):ar1(yf))

## require(lucid)
## vc(m1)
## ##      effect component std.error z.ratio bound

```

```

## ##      xf:yf(R)      0.1044  0.022      4.7      P      0
## ##  xf:yf!xf!cor    0.2458  0.07484    3.3      U      0
## ##  xf:yf!yf!cor    0.8186  0.03822    21       U      0

## # asreml estimates auto-regression correlations of 0.25, 0.82
## # Kempton estimated auto-regression coefficients b1=0.10, b2=0.91

## End(Not run)

# -----

# Kempton defines 4 blocks, randomly assigns variety codes 1-49 in each block, fits
# RCB model, computes mean squares for variety and residual. Repeat 40 times.
# Kempton's estimate: variety = 1032, residual = 1013
# Our estimate: variety = 825, residual = 1080
fitfun <- function(dat){
  dat <- transform(dat, block=factor(ceiling(row/7)),
                    gen=factor(c(sample(1:49),sample(1:49),sample(1:49),sample(1:49))))
  m2 <- lm(yield*100 ~ block + gen, dat)
  anova(m2)[2:3, 'Mean Sq']
}
set.seed(251)
out <- replicate(50, fitfun(dat))
rowMeans(out) # 826 1079

# -----

```

kempton.competition *Sugar beet trial with competition effects*

Description

Yield of sugar beets for 36 varieties in a 3-rep RCB experiment. Competition effects are present.

Format

A data frame with 108 observations on the following 5 variables.

gen genotype, 36 levels

rep rep, 3 levels

row row

col column

yield yield, kg/plot

Details

Entries are grown in 12m rows, 0.5m apart. Guard rows were grown alongside replicate boundaries, but yields of these plots are not included.

Source

R Kempton, 1982. Adjustment for competition between varieties in plant breeding trials, *Journal of Agricultural Science*, 98, 599-611. <http://doi.org/10.1017/S0021859600054381>

Examples

```
data(kempton.competition)
dat <- kempton.competition

# Raw means in Kempton table 2
round(tapply(dat$yield, dat$gen, mean),2)

# Fixed genotype effects, random rep effects,
# Autocorrelation of neighboring plots within the same rep, phi = -0.22
require("nlme")
m1 <- lme(yield ~ -1+gen, random=~1|rep, data=dat,
          corr=corAR1(form=~col|rep))
# Lag 1 autocorrelation is negative--evidence of competition
plot(ACF(m1), alpha=.05, grid=TRUE, main="kempton.competition",
     ylab="Autocorrelation between neighboring plots")

# Genotype effects
round(fixef(m1),2)

# Variance of yield increases with yield
plot(m1, main="kempton.competition")
```

kempton.rowcol

Wheat row-column experiment

Description

Wheat yields of 35 genotypes, 2 reps, planted in a row-column design. Two plots are missing.

Format

A data frame with 68 observations on the following 5 variables.

rep replicate factor, 2 levels

row row

col column

gen genotype factor, 35 levels

yield yield

Details

Included to illustrate REML analysis of a row-column design.

Source

R A Kempton and P N Fox, *Statistical Methods for Plant Variety Evaluation*, Chapman and Hall, 1997.

Examples

```

data(kempton.rowcol)
dat <- kempton.rowcol
dat <- transform(dat, rowf=factor(row), colf=factor(col))

if(require(desplot)){
  desplot(yield~col*row|rep, dat,
          num=gen, out1=rep, # unknown aspect
          main="kempton.rowcol")
}

## Not run:
# Model with rep, row, col as random. Kempton, page 62.
# Use "-1" so that the vcov matrix doesn't include intercept
require(lme4)
m1 <- lmer(yield ~ -1 + gen + rep + (1|rep:rowf) + (1|rep:colf), data=dat)

# Variance components match Kempton.
print(m1, corr=FALSE)

# Standard error of difference for genotypes. Kempton page 62, bottom.
covs <- as.matrix(vcov(m1)[1:35, 1:35])
vars <- diag(covs)
vdiff <- outer(vars, vars, "+") - 2 * covs
sed <- sqrt(vdiff[upper.tri(vdiff)])
min(sed) # Minimum SED
mean(sed) # Average SED
max(sed) # Maximum SED

## End(Not run)

```

kempton.slatehall

Slate Hall Farm 1976 spring wheat

Description

Yields for a Slate Hall Farm 1976 spring wheat trial.

Format

A data frame with 150 observations on the following 5 variables.

rep rep, 6 levels

row row

col column

gen genotype, 25 levels

yield yield (grams/plot)

Details

The trial was a balanced lattice with 25 varieties in 6 replicates, 10 ranges of 15 columns. The plot size was 1.5 meters by 4 meters. Each row within a rep is an (incomplete) block.

Field length: 10 ranges * 4m = 40m

Field width: 15 columns * 1.5m = 22.5m

Source

R A Kempton and P N Fox. (1997). *Statistical Methods for Plant Variety Evaluation*, Chapman and Hall. Page 84.

Julian Besag and David Higdon. 1993. Bayesian Inference for Agricultural Field Experiments. Bull. Int. Statist. Table 4.1.

References

Gilmour, Arthur R and Robin Thompson and Brian R Cullis. (1994). Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models, *Biometrics*, 51, 1440-1450.

Examples

```
data(kempton.slatehall)
dat <- kempton.slatehall

# Besag 1993 figure 4.1 (left panel)
if(require(desplot)){
  grays <- colorRampPalette(c("#d9d9d9", "#252525"))
  desplot(yield ~ col * row, dat,
    aspect=40/22.5, # true aspect
    num=gen, out1=rep, col.regions=grays, # unknown aspect
    main="kempton.slatehall - spring wheat yields")
}

# -----

## Not run:
# Incomplete block model of Gilmour et al 1995
```

```

require(lme4)
require(lucid)
dat <- transform(dat, xf=factor(col), yf=factor(row))
m1 <- lmer(yield ~ gen + (1|rep) + (1|rep:yf) + (1|rep:xf), data=dat)
vc(m1)
##   groups      name variance stddev
## rep:xf (Intercept)  14810 121.7
## rep:yf (Intercept)  15600 124.9
## rep    (Intercept)   4262  65.29
## Residual                8062  89.79

## End(Not run)

# -----

## Not run:
# Incomplete block model of Gilmour et al 1995
# asreml3
require(asreml)
m2 <- asreml(yield ~ gen, random = ~ rep/(xf+yf), data=dat)

vc(m2)
##           effect component std.error z.ratio constr
## rep!rep.var      4262      6890   0.62   pos
## rep:xf!rep.var   14810      4865    3     pos
## rep:yf!rep.var   15600      5091   3.1    pos
## R!variance       8062      1340    6     pos

# Table 4
predict(m2, data=dat, classify="gen")$predictions$pvals

## End(Not run)

# -----

## Not run:
# Incomplete block model of Gilmour et al 1995
## require(asreml4)
## require(lucid)
## m2 <- asreml(yield ~ gen, random = ~ rep/(xf+yf), data=dat)

## vc(m2)
## ##   effect component std.error z.ratio bound
## ##   rep      4262      6890   0.62   P   0
## ##   rep:yf  15600      5091   3.1    P   0
## ##   rep:xf  14810      4865    3     P   0
## ##   units(R) 8062      1340    6     P   0

## # Table 4
## # predict(m2, data=dat, classify="gen")$pvals
## ##   gen predicted.value std.error  status
## ## 1 G01          1280      60.2 Estimable
## ## 2 G02          1550      60.2 Estimable

```

```

## ## 3 G03          1420      60.2 Estimable
## ## 4 G04          1450      60.2 Estimable
## ## 5 G05          1530      60.2 Estimable

## End(Not run)

# -----

```

kenward.cattle *Repeated measurement of weights of calves with two treatments.*

Description

Repeated measurements of the weights of calves from a trial on the control of intestinal parasites.

Usage

```
data("kenward.cattle")
```

Format

A data frame with 660 observations on the following 4 variables.

```

animal animal factor
trt treatment factor, A or B
day day, numeric, 0-133
weight bodyweight, kg

```

Details

Grazing cattle can ingest larvae, which deprives the host animal of nutrients and weakens the immune system, affecting the growth of the animal.

Two treatments A and B were applied randomly to 60 animals (30 each in two groups) to control the disease.

Each animal was weighed 11 times at two-week intervals (one week between the final two measurements).

Is there a difference in treatments, and when does that difference first become manifest?

Source

Kenward, Michael G. (1987). A Method for Comparing Profiles of Repeated Measurements. *Applied Statistics*, 36, 296-308. Table 1. <http://doi.org/10.2307/2347788>

References

W. Zhang, C. Leng and C. Y. Tang (2015). A joint modelling approach for longitudinal studies *J. R. Statist. Soc. B*, 77 (2015), 219–238. <http://doi.org/10.1111/rssb.12065>

Examples

```

data(kenward.cattle)
dat <- kenward.cattle

# Profile plots
require(lattice)
foo1 <- xyplot(weight~day|trt, data=dat, type='l', group=animal,
               xlab="Day", ylab="Animal weight", main="kenward.cattle")
print(foo1)

# -----

## Not run:

# lme4. Fixed treatment intercepts, treatment polynomial trend.
# Random deviation for each animal
require(lme4)
m1a <-lmer(weight ~ trt*poly(day, 4) + (1|animal), data=dat,
           REML = FALSE)
# Change separate polynomials into common polynomial
m1b <-lmer(weight ~ trt + poly(day, 4) + (1|animal), data=dat,
           REML = FALSE)
# Drop treatment differences
m1c <-lmer(weight ~ poly(day, 4) + (1|animal), data=dat,
           REML = FALSE)
anova(m1a, m1b, m1c) # Significant differences between trt polynomials

# Overlay polynomial predictions on plot
require(latticeExtra)
dat$pred <- predict(m1a, re.form=NA)
foo1 + xyplot(pred ~ day|trt, data=dat,
              lwd=2, col="black", type='l')

# A Kenward-Roger Approximation and Parametric Bootstrap
library(pbkrtest)
KRmodcomp(m1b, m1c) # Non-signif
# Model comparison of nested models using parametric bootstrap methods
# PBmodcomp(m1b, m1c, nsim=500)
## Parametric bootstrap test; time: 13.20 sec; samples: 500 extremes: 326;
## large : weight ~ trt + poly(day, 4) + (1 | animal)
## small : weight ~ poly(day, 4) + (1 | animal)
##          stat df p.value
## LRT     0.2047  1  0.6509
## PBtest  0.2047   0.6527

## End(Not run)

# -----

## Not run:

```

```

# ASREML approach to model. Not final by any means.
# Maybe a spline curve for each treatment, plus random deviations for each time
# asreml3
require(asreml)
m1 <- asreml(weight ~ 1 + lin(day) +      # overall line
             trt + trt:lin(day),         # different line for each treatment
             data=dat,
             random = ~ spl(day) +      # overall spline
                 trt:spl(day) +         # different spline for each treatment
                 dev(day) + trt:dev(day) ) # non-spline deviation at each time*trt

p1 <- predict(m1, data=dat, classify="trt:day")
p1 <- p1$predictions$pvals
foo2 <- xyplot(predicted.value ~ day|trt, p1, type='l', lwd=2, lty=1, col="black")

require(latticeExtra)
print(foo1 + foo2)

# Not much evidence for treatment differences

anova(m1)
##              Df Sum of Sq Wald statistic Pr(Chisq)
## (Intercept)  1 37128459      139060 <2e-16 ***
## trt          1    455          2 0.1917
## lin(day)     1  570798        2138 <2e-16 ***
## trt:lin(day) 1    283          1 0.3031
## residual (MS)      267

```

| effect component | std.error | z.ratio | constr |
|----------------------|-----------|---------|----------|
| spl(day) | 25.29 | 24.09 | 1 pos |
| dev(day) | 1.902 | 4.923 | 0.39 pos |
| trt:spl(day)!trt.var | 0.00003 | 0.00002 | 18 bnd |
| trt:dev(day)!trt.var | 0.00003 | 0.00002 | 18 bnd |
| R!variance | 267 | 14.84 | 18 pos |

```

require(lucid)
vc(m1)
##              effect component std.error z.ratio constr
##              spl(day) 25.29 24.09 1 pos
##              dev(day) 1.902 4.923 0.39 pos
## trt:spl(day)!trt.var 0.00003 0.00002 18 bnd
## trt:dev(day)!trt.var 0.00003 0.00002 18 bnd
##              R!variance 267 14.84 18 pos

## End(Not run)

# -----

## Not run:
# ASREML approach to model. Not final by any means.
# Maybe a spline curve for each treatment, plus random deviations for each time
## require(asreml4)
## m1 <- asreml(weight ~ 1 + lin(day) +      # overall line
##          trt + trt:lin(day),         # different line for each treatment
##          data=dat,
##          random = ~ spl(day) +      # overall spline
##              trt:spl(day) +         # different spline for each treatment
##              dev(day) + trt:dev(day) ) # non-spline deviation at each time*trt

## p1 <- predict(m1, data=dat, classify="trt:day",

```

```

##           design.points=list(day=0:133))
## p1 <- p1$pvals
## foo2 <- xyplot(predicted.value ~ day|trt, p1, type='l', lwd=2, lty=1, col="black")

## require(latticeExtra)
## print(foo1 + foo2)

## # Not much evidence for treatment differences

## wald(m1)
## ##           Df Sum of Sq Wald statistic Pr(Chisq)
## ## (Intercept)  1 37128459          139060 <2e-16 ***
## ## trt          1     455             2    0.1917
## ## lin(day)     1  570798           2138    <2e-16 ***
## ## trt:lin(day) 1     283             1    0.3031
## ## residual (MS)          267

## require(lucid)
## vc(m1)
## ##           effect component std.error z.ratio constr
## ##           spl(day) 25.29  24.09      1      pos
## ##           dev(day)  1.902  4.923    0.39    pos
## ## trt:spl(day)!trt.var 0.00003 0.000002 18     bnd
## ## trt:dev(day)!trt.var 0.00003 0.000002 18     bnd
## ##           R!variance 267      14.84    18     pos

## End(Not run)

# -----

```

khin.rice.uniformity *Uniformity trial of rice*

Description

Uniformity trial of rice in Burma, 1948.

Usage

```
data("khin.rice.uniformity")
```

Format

A data frame with 1080 observations on the following 3 variables.

row row

col column

yield yield, oz/plot

Details

A uniformity trial of rice. Conducted at the Mudon Agricultural Station, Burma, in 1947-48. Basic plots were 3 feet square.

Field width: 30 plots * 3 feet.

Field length: 36 plots * 3 feet.

Source

Khin, San. 1950. Investigation into the relative costs of rice experiments based on the efficiency of designs. Dissertation: Imperial College of Tropical Agriculture (ICTA). Appendix XV. <http://hdl.handle.net/2139/42422>

References

None.

Examples

```
data(khin.rice.uniformity)
dat <- khin.rice.uniformity

if(require(desplot)){
  desplot(yield ~ col*row, data=dat, flip=TRUE,
          main="khin.rice.uniformity",
          aspect=(36*3)/(30*3)) # true aspect
}
```

kiesselbach.oats.uniformity
Uniformity trial of oats

Description

Uniformity trial of oats at Nebraska in 1916.

Usage

```
data("kiesselbach.oats.uniformity")
```

Format

A data frame with 207 observations on the following 3 variables.

row row

col column

yield yield bu/ac

Details

Experiment conducted in 1916. Crop was Kerson oats. Each plot covered 1/30th acre. Oats were drilled in plats 66 inches wide by 16 rods long. The drill was 66 inches wide. Plats were separated by a space of 16 inches between outside drill rows.

The source document includes three photographs of the field.

1 acre = 43560 sq feet

1/30 acre = 1452 sq feet = 16 rods * 16.5 ft/rod * 5.5 ft

Field length: 69 plats * 5.5 ft + 68 gaps * 1.33 feet = 469 feet

Field width: 3 plats * 16 rods/plat * 16.5 ft/rod = 792 feet

Source

Kiesselbach, Theodore A. (1917). Studies Concerning the Elimination of Experimental Error in Comparative Crop Tests. University of Nebraska Agricultural Experiment Station Research Bulletin No. 13. Pages 51-72.

<https://archive.org/details/StudiesConcerningTheEliminationOfExperimentalErrorInComparativeCrop>
<http://digitalcommons.unl.edu/extensionhist/430/>

References

None.

Examples

```
data(kiesselbach.oats.uniformity)
dat <- kiesselbach.oats.uniformity
range(dat$yield) # 56.7 92.8 match Kiesselbach p 64.

if(require(desplot)){
  desplot(yield ~ col*row, dat,
          tick=TRUE, flip=TRUE, aspect=792/469, # true aspect
          main="kiesselbach.oats.uniformity")
}
```

kreusler.maize

Growth of maize plants in Germany during 1875-1878

Description

Growth of maize plants in Germany during 1875-1878.

Usage

```
data("kreusler.maize")
```

Format

A data frame with 165 observations on the following 17 variables.

gen genotype

year year

date calendar date

raindays number of days of rain per week (zahl der regenstage)

rain rain amount (mm)

temp temperature mean (deg C) (temperatur mittel)

parentseed weight of parent seed (g) (alte korner)

roots weight of roots (g) (wurzel)

leaves weight of leaves (g) (blatter)

stem weight of stem (g) (stengel)

tassel weight of tassel (g) (blutenstande)

grain weight of grain (korner)

plantweight weight of entire plant (ganze pflanze)

plantheight plant height (cm) (mittlere hohe der pflanzen)

leafcount number of leaves (anzahl der blatter)

leafarea leaf area (cm²) (flachenmaass der blatter)

Details

Experiments were performed at Poppelsdorf, Germany (near Bonn) during the years 1875 to 1878. Observations were collected weekly throughout the growing season.

Five varieties were grown in 1875. Two in 1876, and one in 1877 and 1878.

The plants were selected by eye as representative, with the number of plants chosen decreasing during the growing season. For example, the dry-weight data was based on the following number of plants:

In 1875 the number sampled began at 20 and dropped to 10.

In 1876 the number sampled began at 45 and dropped to 24.

In 1877 the number sampled began at 90 and dropped to 36.

In 1878 the number sampled began at 120 and dropped to 40.

Most of the observations included fresh weight and dry weight of entire plants, along with leaf area, date of inflorescence, fertilization, and kernel development.

The data of Hornberger 71 are the same as Kreuzler/Hornberger, but more complete.

The temperature data was originally given in degrees Reaumur in 1875 and 1876, and degrees Celsius in 1877 and 1878. All temperatures in this data are degrees Celsius. Note: deg C = 1.25 deg R. Briggs, Kidd & West (1920) give all temperature in Celsius.

Source

The 1875-1876 data are from:

A. Prehn & G. Becker. (1878) *Jahresbericht fur Agrikultur-chemie*, Vol 20, p. 216-220. <https://books.google.com/books?id=Z>

The 1877 data are from:

A. Kreusler, A. Prehn, Hornberger. (1880) *Jahresbericht fur Agrikultur-Chemie*, Vol 21, p. 248. <https://books.google.com/books?id=U3IYAQAIAAJ&pg=248>

The 1878 data are from:

U. Kreusler, A. Prehn, R. Hornberger. (1880). *Jahresbericht fur Agrikultur-Chemie*, Vol 22, p. 211. <https://books.google.com/books?id=9HIYAQAIAAJ&pg=211>

Dry plant weight and leaf area for all genotypes and years are repeated by:

G. E. Briggs, Franklin Kidd, Cyril West. (1920). A Quantitative Analysis of Plant Growth. Part I. *Annals of Applied Biology*, 7, 103-123.

G. E. Briggs, Franklin Kidd, Cyril West. (1920). A Quantitative Analysis of Plant Growth. Part II. *Annals of Applied Biology*, 7, 202-223.

References

Roderick Hunt, G. Clifford Evans. 1980. Classical Data on the Growth of Maize: Curve Fitting With Statistical Analysis. *New Phytol*, 86, 155-180.

Examples

```
data(kreusler.maize)
dat <- kreusler.maize
dat$date2 <- as.Date(dat$date,"%d %b %Y")
dat$doy <- as.numeric(strftime(dat$date2, format="%j"))

# Hunt & Evans Fig 2a
library(lattice)
xyplot(log10(plantweight)~doy|factor(year), data=dat, group=gen,
        type=c('p','smooth'), span=.4, as.table=TRUE,
        xlab="Day of year", main="kreusler.maize - growth of maize",
        auto.key=list(columns=5))

## Not run:

# Hunt & Evans Fig 2b
xyplot(log10(plantweight)~doy|gen, data=dat, group=factor(year),
        type=c('p','smooth'), span=.5, as.table=TRUE,
        xlab="Day of year",
        auto.key=list(columns=4))

# Hunt & Evans Fig 3a
xyplot(log10(leafarea)~doy|factor(year), data=dat, group=gen,
        type=c('p','smooth'), span=.5, as.table=TRUE,
        xlab="Day of year",
        auto.key=list(columns=5))

# Hunt & Evans Fig 3a
```

```

xyplot(log10(leafarea)~doy|gen, data=dat, group=factor(year),
       type=c('p','smooth'), span=.5, as.table=TRUE,
       xlab="Day of year",
       auto.key=list(columns=4))

# All traits

xyplot(raindays~doy|factor(year), data=dat, group=gen,
       type='l', auto.key=list(columns=5), as.table=TRUE, layout=c(1,4))
xyplot(rain~doy|factor(year), data=dat, group=gen,
       type='l', auto.key=list(columns=5), as.table=TRUE, layout=c(1,4))
xyplot(temp~doy|factor(year), data=dat, group=gen,
       type='l', auto.key=list(columns=5), as.table=TRUE, layout=c(1,4))
xyplot(parentseed~doy|factor(year), data=dat, group=gen,
       type='l', auto.key=list(columns=5), as.table=TRUE, layout=c(1,4))
xyplot(roots~doy|factor(year), data=dat, group=gen,
       type='l', auto.key=list(columns=5), as.table=TRUE, layout=c(1,4))
xyplot(leaves~doy|factor(year), data=dat, group=gen,
       type='l', auto.key=list(columns=5), as.table=TRUE, layout=c(1,4))
xyplot(stem~doy|factor(year), data=dat, group=gen,
       type='l', auto.key=list(columns=5), as.table=TRUE, layout=c(1,4))
xyplot(grain~doy|factor(year), data=dat, group=gen,
       type='l', auto.key=list(columns=5), as.table=TRUE, layout=c(1,4))
xyplot(plantweight~doy|factor(year), data=dat, group=gen,
       type='l', auto.key=list(columns=5), as.table=TRUE, layout=c(1,4))
xyplot(plantheight~doy|factor(year), data=dat, group=gen,
       type='l', auto.key=list(columns=5), as.table=TRUE, layout=c(1,4))
xyplot(leafcount~doy|factor(year), data=dat, group=gen,
       type='l', auto.key=list(columns=5), as.table=TRUE, layout=c(1,4))
xyplot(leafarea~doy|factor(year), data=dat, group=gen,
       type='l', auto.key=list(columns=5), as.table=TRUE, layout=c(1,4))
xyplot(tassel~doy|factor(year), data=dat, group=gen,
       type='l', auto.key=list(columns=5), as.table=TRUE, layout=c(1,4))

## End(Not run)

```

kristensen.barley.uniformity

Uniformity trial of barley

Description

Uniformity trial of barley conducted in Denmark, 1905.

Usage

```
data("kristensen.barley.uniformity")
```


Format

A data frame with 718 observations on the following 3 variables.

row row

col column

yield yield, hectograms/plot

Details

Experiment conducted in 1905 at Askov, Denmark. Harvested plot size was 10 x 14 'alen', 6.24 x 8.79 meters. The soil was uniform, but an attack of mildew spread from an adjacent field. Yield is measured in hectograms/plot for straw and grain together. (Page 468).

Orientation of the plots dimensions is not clear from the text, but the aspect used in the example below aligns well with Kristensen figure 1.

Notes from Kristensen: Fig 5 is a 3x3 moving average, Fig 6 is deviation from the trend, Fig 7 is the field average added to the deviation. Fig 13 is another uniformity trial of barley in 1924, Fig 14 is a uniformity trial of oats in 1924.

Source

R. K. Kristensen (1925). Anlaeg og Opgørelse af Markforsøeg. *Tidsskrift for landbrugets planteavl*, Vol 31, 464. Fig 1, pg. 467. <http://dca.au.dk/publikationer/historiske/planteavl/>

References

J. Neyman, K. Iwazskiewicz, St. Kolodziejczyk. (1935). Statistical Problems in Agricultural Experimentation. *Supplement to the Journal of the Royal Statistical Society*, Vol. 2, No. 2 (1935), pp. 107-180. <http://doi.org/10.2307/2983637>

Examples

```
data(kristensen.barley.uniformity)
dat <- kristensen.barley.uniformity

if(require(desplot)){
  desplot(yield ~ col*row, dat,
          flip=TRUE, aspect=(11*6.24)/(22*8.79),
          main="kristensen.barley.uniformity")
}
```

lambert.soiltemp *Average monthly soil temperature near Zurich*

Description

Average monthly soil temperature near Zurich, at seven depths, averaged over four years.

Format

A data frame with 84 observations on the following 3 variables.

month month

depth depth in soil (feet)

temp temperature (the units are "du Crest")

Details

This is one of the earliest time series in scientific literature.

These data show the monthly soil temperature near Zurich, averaged over four years (beginning in 1762), at 7 different depths.

The temperature measurements are related to the 'du Crest' scale. (The measurements do not seem to be exactly according to the du Crest scale. If you can read German, use the Google books link to see if you can figure out why.) Even the scale on Lambert's own graph doesn't match the data.

Greater depths show less variation and a greater lag in temperature responsiveness to the air temperature.

This data also appears in *Pedometrics*, issue 23, December 2007. But, the formula for converting the temperature does not make sense and the data in Table 1 do not directly match the corresponding figure.

Source

Johann Heinrich Lambert (1779), *Pyrometrie*. Page 358. http://books.google.com/books?id=G5I_AAAAcAAJ&pg=PA358

Graph: http://www.fisme.science.uu.nl/wiskrant/artikelen/hist_grafieken/begin/images/pyrometrie.gif

Examples

```
# Reproduce Lambert figure 39.

data(lambert.soiltemp)
dat <- lambert.soiltemp

# Make 3 cycles of the data so that the loess line bends back up at
# month 1 and month 12
dat <- rbind(dat, transform(dat, month=month-12),
             transform(dat, month=month+12))
require(lattice)
```

```
xyplot(temp ~ month, dat, group=depth, type=c('p','smooth'),
        main="lambert.soiltemp",
        xlim=c(-3,15), ylab="Soil temperature (du Crest) at depth (feet)",
        span=.2, auto.key=list(columns=4))

# To do: Find a good model for this data
```

 lasrosas.com

Yield monitor data for a corn field in Argentina with variable nitrogen.

Description

Yield monitor data for a corn field in Argentina with variable nitrogen.

Usage

```
data("lasrosas.com")
```

Format

A data frame with 3443 observations on the following 8 variables.

year year, 1999 or 2001

lat latitude

long longitude

yield yield, quintals/ha

nitro nitrogen fertilizer, kg/ha

topo topographic factor

bv brightness value (proxy for low organic matter content)

rep rep factor

nf nitrogen as a factor, N0-N4

Details

Corn yield and nitrogen fertilizer treatment with field characteristics for the Las Rosas farm, Rio Cuarto, Cordoba, Argentina.

Data has 6 nitro treatments, 3 reps, in strips.

Data collected using yield monitor, for harvests in 1999 and 2001.

The points within each long strip have been averaged so that the distance between points `_within_` a strip is the same as the distance `_between_` strips (9.8 meters).

The topographic factor a factor with levels W = West slope, HT = Hilltop, E = East slope, LO = Low East.

The 'rep' factor in this data was added by hand and did not appear in the original data.

Slightly different levels of nitrogen were used in the two years, so the nitrogen factor 'nf' was created to have common levels across years.

Published descriptions of the data describe the experiment design as having randomized nitrogen treatments. The nitrogen treatments were randomized within one rep, but the same randomization was used in the other two reps.

Anselin et al. used corn grain price of \$6.85/quintal and nitrogen cost of \$0.4348/kg.

The corners of the field in 1999 are: <https://www.google.com/maps/place/-33.0501258,-63.8488636>
<https://www.google.com/maps/place/-33.05229635,-63.84181819>

Anselin et al. found a significant response to nitrogen for slope. However, Bongiovanni and Lowenberg-DeBoer (2002) found that slope position was NOT significant in 2001.

Source

The Las Rosas data files were obtained from <https://geodacenter.asu.edu/sdata> and converted from ESRI shape files to a flat data.frame.

Used with permission of the ASU GeoDa Center.

References

Bongiovanni and Lowenberg-DeBoer (2000). Nitrogen management in corn with a spatial regression model. *Proceedings of the Fifth International Conference on Precision Agriculture*

Anselin, L., R. Bongiovanni, J. Lowenberg-DeBoer (2004). A spatial econometric approach to the economics of site-specific nitrogen management in corn production. *American Journal of Agricultural Economics*, 86, 675–687.

Lambert, Lowenberg-Deboer, Bongiovanni (2004). A Comparison of Four Spatial Regression Models for Yield Monitor Data: A Case Study from Argentina. *Precision Agriculture*, 5, 579-600.

Examples

```
data(lasrosas.corn)
dat <- lasrosas.corn

# yield map
require(lattice)
if(require(latticeExtra)){ # for panel.levelplot.points
  redblue <- colorRampPalette(c("firebrick", "lightgray", "#375997"))
  levelplot(yield ~ long*lat|factor(year), data=dat,
            main="lasrosas.corn grain yield", xlab="Longitude", ylab="Latitude",
            scales=list(alternating=FALSE),
            prepanel = prepanel.default.xyplot,
            panel = panel.levelplot.points,
            type = c("p", "g"), aspect = "iso", col.regions=redblue)
}

d1 <- subset(dat, year==1999)

# Experiment design
xyplot(lat~long, data=d1, col=as.numeric(as.factor(d1$nitro)), pch=d1$topo,
```

```

main="lasrosas.corn experiment layout 1999")

# A quadratic response to nitrogen is suggested
xyplot(yield~nitro|topo, data=d1, type=c('p','smooth'), layout=c(4,1),
       main="lasrosas.corn yield by topographic zone 1999")

# Full-field quadratic response to nitrogen. Similar to Bongiovanni 2000,
# table 1.
m1 <- lm(yield ~ 1 + nitro + I(nitro^2), data=d1, subset=year==1999)
coef(m1)

```

lavoranti.eucalyptus *Height of Eucalyptus trees in southern Brazil*

Description

Height of Eucalyptus trees in southern Brazil

Format

A data frame with 490 observations on the following 4 variables.

gen genotype (progeny) factor

origin origin of progeny

loc location

height height, meters

Details

The genotypes originated from three different locations in Queensland, Australia, and were tested in southern Brazil. The experiment was conducted as a randomized complete block design with 6 plants per plot and 10 blocks. Mean tree height is reported.

The testing locations are described in the following table:

| Loc | City | Lat (S) | Long (W) | Altitude | Avg min temp | Avg max temp | Avg temp (C) | Preci |
|-----|--------------------------|---------|----------|----------|--------------|--------------|--------------|-------|
| L1 | Barra Ribeiro, RS | 30.33 | 51.23 | 30 | 9 | 25 | 19 | 1400 |
| L2 | Telemaco Borba, PR | 24.25 | 20.48 | 850 | 11 | 26 | 19 | 1480 |
| L3 | Boa Esperanca de Sul, SP | 21.95 | 48.53 | 540 | 15 | 23 | 21 | 1300 |
| L4 | Guanhaes, MG | 18.66 | 43 | 900 | 14 | 24 | 19 | 1600 |
| L5 | Ipatinga, MG | 19.25 | 42.33 | 250 | 15 | 24 | 22 | 1250 |
| L6 | Aracruz, ES | 19.8 | 40.28 | 50 | 15 | 26 | 24 | 1360 |
| L7 | Cacapva, SP | 23.05 | 45.76 | 650 | 14 | 24 | 20 | 1260 |

Arciniegas-Alarcon (2010) used the 'Ravenshoe' subset of the data to illustrate imputation of missing values.

Source

O J Lavoranti (2003). Estabilidade e adaptabilidade fenotípica através da reamostragem bootstrap no modelo AMMI, PhD thesis, University of Sao Paulo, Brazil.

References

Arciniegas-Alarcon, S. and Garcia-Pena, M. and dos Santos Dias, C.T. and Krzanowski, W.J. (2010). An alternative methodology for imputing missing data in trials with genotype-by-environment interaction, *Biometrical Letters*, 47, 1-14. <http://doi.org/10.2478/bile-2014-0006>

Examples

```
# Arciniegas-Alarcon et al use SVD and regression to estimate missing values.
# Partition the matrix X as a missing value xm, row vector xr1, column
# vector xc1, and submatrix X11
# X = [ xm  xr1 ]
#      [ xc1 X11 ] and let X11 = UDV'.
# Estimate the missing value xm = xr1 V D^{-1} U' xc1

data(lavoranti.eucalyptus)
dat <- lavoranti.eucalyptus

require(lattice)
grays <- colorRampPalette(c("#d9d9d9", "#252525"))
levelplot(height~loc*gen, dat, main="lavoranti.eucalyptus - GxE heatmap",
          col.regions=grays)

dat <- droplevels(subset(dat, origin=="Ravenshoe"))
if(require(reshape2)){
  dat <- acast(dat, gen~loc, value.var='height')

  dat[1,1] <- NA
  x11 <- dat[-1,][,-1]
  X11.svd <- svd(x11)
  xc1 <- dat[-1,][,1]
  xr1 <- dat[, -1][1,]
  xm <- xr1
  xm # = 18.29, Original value was 17.4
}
```

 lee.potatobligh

Repeated measurements of resistance to potato blight

Description

Repeated measurements of resistance to potato blight.

Usage

```
data("lee.potatobligh")
```

Format

A data frame with 14570 observations on the following 7 variables.

```
year  planting year
gen   genotype / cultivar factor
col   column
row   row
rep   replicate block (numeric)
date  date for data collection
y     score 1-9 for blight resistance
```

Details

These data were collected from biennial screening trials conducted by the New Zealand Institute of Crop and Food Research at the Pukekohe Field Station. The trials evaluate the resistance of potato cultivars to late blight caused by the fungus *Phytophthora infestans*. In each trial, the damage to necrotic tissue was rated on a 1-9 scale at multiple time points during the growing season.

Lee (2009) used a Bayesian model that extends the ordinal regression of McCullagh to include spatial variation and sigmoid logistic curves to model the time dependence of repeated measurements on the same plot.

Data from 1989 were not included due to a different trial setup being used. All the trials here were laid out as latinized row-column designs with 4 or 5 reps. Each plot consisted of four seed tubers planted with two Ilam Hardy spread plants in a single row 2 meters long with 76 centimeter spacing between rows.

In 1997, 18 plots were lost due to flooding. In 2001, by the end of the season most plants were nearly dead.

Note, in plant-breeding, it is common to use a "breeder code" for each genotype, which after several years of testing is changed to a registered commercial variety name. For this R package, the Potato Pedigree Database, <http://www.plantbreeding.wur.nl/potatopedigree/reverselookup.php>, was used to change breeder codes (in early testing) to the variety names used in later testing. For example, among the changes made were the following:

| | |
|------------|---------------------|
| Driver | 287.12 |
| Kiwitea | 064/56 |
| Gladiator | 1308.66 |
| Karaka | 221.17 |
| Kiwitea | 064.56 maybe 064.54 |
| Moonlight | 511.1 |
| Pacific | 177.3 |
| Red Rascal | 1830.11 |
| Rua | 155.05 |
| Summit | 517.12 |

White Delight 1949.64

Source

Lee, Arier Chi-Lun (2009). Random effects models for ordinal data. Ph.D. thesis, The University of Auckland. <https://researchspace.auckland.ac.nz/handle/2292/4544>.

Used with permission of Arier Chi-Lun Lee and John Anderson.

Data retrieved from <https://researchspace.auckland.ac.nz/handle/2292/5240>.

Licensed via Open Database License 1.0. (allows sub-licensing). See: <http://opendatacommons.org/licenses/dbcl/1.0/>

Examples

```
data(lee.potatobligh)
dat <- lee.potatobligh

# Common cultivars across years.
# Based on code from here: http://stackoverflow.com/questions/20709808
gg <- tapply(dat$gen, dat$year, function(x) as.character(unique(x)))
tab <- outer(1:11, 1:11,
             Vectorize(function(a, b) length(Reduce(intersect, gg[c(a, b)]))))
head(tab) # Matches Lee page 27.
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11]
## [1,]  20  10   7   5   3   2   3   2   3   3   2
## [2,]  10  30  17   5   4   3   4   4   5   4   2
## [3,]   7  17  35   9   6   3   4   5   6   4   3
## [4,]   5   5   9  35  16   8   9  14  15  13  11
## [5,]   3   4   6  16  40  12  11  18  18  16  14

# Note the progression to lower scores as time passes in each year
skp <- c(rep(0,10),
         rep(0,7),1,1,1,
         rep(0,8),1,1,
         rep(0,6),1,1,1,1,
         rep(0,5),1,1,1,1,1,
         rep(0,5),1,1,1,1,1,
         rep(0,6),1,1,1,1,
         rep(0,5),1,1,1,1,1,
         rep(0,5),1,1,1,1,1,
         rep(0,5),1,1,1,1,1)

if(require(desplot)){
  desplot(y ~ col*row|date, dat,
          ylab="Year of testing", # unknown aspect
          layout=c(10,11),skip=as.logical(skp),
          main="lee.potatobligh - maps of blight resistance over time")
}
```



```
# 1983 only. I.Hardy succumbs to blight quickly
require("lattice")
xyplot(y ~ date|gen, dat, subset=year==1983, group=rep,
       xlab="Date", ylab="Blight resistance score",
       main="lee.potatobligh 1983", as.table=TRUE,
       auto.key=list(columns=5),
       scales=list(alternating=FALSE, x=list(rot=90, cex=.7)))
```

lehner.soybeanmold *Yield, white mold, and sclerotia for soybeans in Brazil*

Description

Yield, white mold, and sclerotia for soybeans in Brazil

Usage

```
data("lehner.soybeanmold")
```

Format

A data frame with 382 observations on the following 9 variables.

study study number
year year of harvest
loc location name
elev elevation
region region
trt treatment number
yield crop yield, kg/ha
mold white mold incidence, percent
sclerotia weight of sclerotia g/ha

Details

Data are the mean of 4 reps.

Original source (Portuguese) <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/101371/1/Ensaios-cooperativos-de-controle-quimico-de-mofo-branco-na-cultura-da-soja-safras-2009-a-2012.pdf>

Data included here via GPL3 license.

Source

Lehner, M. S., Pethybridge, S. J., Meyer, M. C., & Del Ponte, E. M. (2016). Meta-analytic modelling of the incidence-yield and incidence-sclerotial production relationships in soybean white mould epidemics. *Plant Pathology*. doi:10.1111/ppa.12590

References

Full commented code and analysis <https://emdelponete.github.io/paper-white-mold-meta-analysis/>

Examples

```

data(lehner.soybeanmold)
dat <- lehner.soybeanmold

## Not run:
op <- par(mfrow=c(2,2))
hist(dat$mold, main="White mold incidence")
hist(dat$yield, main="Yield")
hist(dat$sclerotia, main="Sclerotia weight")
par(op)

## End(Not run)

require(lattice)
xyplot(yield ~ mold|study, dat, type=c('p','r'),
       main="lehner.soybeanmold")
# xyplot(sclerotia ~ mold|study, dat, type=c('p','r'))

# Example meta-analysis. Could use metafor package to construct the forest
# plot, but it is easier with lattice and ggplot is slow/clumsy
if(require(latticeExtra) & require(metafor)){
  # calculate correlation & confidence for each loc
  cors <- split(dat, dat$study)
  cors <- sapply(cors,
                 FUN=function(X){
                   res <- cor.test(X$yield, X$mold)
                   c(res$estimate, res$parameter[1],
                     conf.low=res$conf.int[1], conf.high=res$conf.int[2])
                 })
  cors <- as.data.frame(t(as.matrix(cors)))
  cors$study <- rownames(cors)
  # Fisher Z transform
  cors <- transform(cors, ri = cor)
  cors <- transform(cors, ni = df + 2)
  cors <- transform(cors,
                    yi = 1/2 * log((1 + ri)/(1 - ri)),
                    vi = 1/(ni - 3))
  # Overall correlation across studies
  overall <- rma.uni(yi, vi, method="ML", data=cors) # metafor package
  # back transform
  overall <- predict(overall, transf=transf.ztor)

  # weight and size for forest plot
  wi <- 1/sqrt(cors$vi)
  size <- 0.5 + 3.0 * (wi - min(wi))/(max(wi) - min(wi))

  # now the forest plot
  segplot(factor(study) ~ conf.low+conf.high, data=cors,

```

```

        draw.bands=FALSE, level=size, centers=ri, cex=size,
        col.regions=colorRampPalette(c("gray85", "dodgerblue4")),
        main="White mold vs. soybean yield",
        xlab=paste("Study correlation, confidence, and study weight (blues)\n",
                  "Overall (black)"),
        ylab="Study ID") +
    layer(panel.abline(v=overall$pred, lwd=2)) +
    layer(panel.abline(v=c(overall$cr.lb, overall$cr.ub), lty=2, col="gray"))
}

# Meta-analyses are typically used when the original data is not available.
# Since the original data is available, a mixed model is probably better.
## Not run:
library(lme4)
m1 <- lmer(yield ~ mold # overall slope
           + (1+mold |study), # random intercept & slope per study
           data=dat)
summary(m1)

## End(Not run)

```

```

lessman.sorghum.uniformity
           Uniformity trial of sorghum

```

Description

Uniformity trial of sorghum at Ames, Iowa, 1959.

Usage

```
data("lessman.sorghum.uniformity")
```

Format

A data frame with 2640 observations on the following 3 variables.

row row

col column

yield yield, ounces

Details

The uniformity trial was conducted at the Agronomy Farm at Ames, Iowa, in 1959. The field was planted to grain sorghum in rows spaces 40 inches apart, thinned to a stand of three inches between plants. The entire field was 48 rows (40 inches apart), each 300 feet long and harvested in 5-foot lengths. Threshed grain was dried to 8-10 percent moisture before weighing. Weights are ounces. Average yield for the field was 95.3 bu/ac.

Field width: 48 rows * 40 inches / 12in/ft = 160 feet

Field length: 300 feet

Plot yields from the two outer rows on each side of the field were omitted from the analysis.

CV values from this data do not quite match Lessman's value. The first page of Table 17 was manually checked for correctness and there were no problems with the optical character recognition (other than obvious errors like 0/o).

Source

Lessman, Koert James, (1962) *Comparisons of methods for testing grain yield of sorghum* Retrospective Theses and Dissertations. Paper 2063. Appendix Table 17.

References

None.

Examples

```
data(lessman.sorghum.uniformity)
dat <- lessman.sorghum.uniformity

if(require(desplot)){
  desplot(yield ~ col*row, dat,
          aspect=300/160, tick=TRUE, flip=TRUE, # true aspect
          main="lessman.sorghum.uniformity")
}

# Omit outer two columns (called 'rows' by Lessman)
dat <- subset(dat, col > 2 & col < 47)
nrow(dat)
var(dat$yield) # 9.09
sd(dat$yield)/mean(dat$yield) # CV 9.2%

if(require(reshape2) & require(agricolae)){
  dmat <- acast(dat, row~col, value.var='yield')
  index.smith(dmat,
              main="lessman.sorghum.uniformity",
              col="red") # Similar to Lessman Table 1
  # Lessman said that varying the width of plots did not have an appreciable
  # effect on CV, and optimal row length was 3.2 basic plots, about 15-20
}
```

li.millet.uniformity *Uniformity trial of millet*

Description

Uniformity trial of millet at China in 1934.

Format

A data frame with 600 observations on the following 3 variables.

row row

col column

yield yield (grams)

Details

Crop date estimated to be 1934.

Field was 100 ft x 100 ft. Plots were 15 feet long by 1 foot wide.

Field width: 100 feet

Field length: 100 feet

Li found the most efficient use of land was obtained with plats 15 feet long and two rowss wide. Also satisfactory would be one row 30 feet long.

Source

Li, HW and Meng, CJ and Liu, TN. 1936. Field Results in a Millet Breeding Experiment. *Agronomy Journal*, 28, 1-15. Table 1. <http://doi.org/10.2134/agronj1936.00021962002800010001x>

Examples

```
data(li.millet.uniformity)
dat <- li.millet.uniformity

mean(dat$yield) # matches Li et al.
if(require(desplot)){
  desplot(yield~col*row, dat,
          aspect=100/100, # true aspect
          main="li.millet.uniformity")
}
```

 lillemo.wheat

Multi-environment trial of wheat susceptible to powdery mildew

Description

Resistance of wheat to powdery mildew

Usage

```
data("lillemo.wheat")
```

Format

A data frame with 408 observations on the following 4 variables.

gen genotype, 24 levels

env environment, 13 levels

score score

scale scale used for score

Details

The data are means across reps of the original scores. Lower scores indicate better resistance to mildew.

Each location used one of four different measurement scales for scoring resistance to powdery mildew: 0-5 scale, 1-9 scale, 0-9 scale, percent.

Environment codes consist of two letters for the location name and two digits for the year of testing. Location names: CA=Cruz Alta, Brazil. Ba= Bawburgh, UK. Aa=As, Norway. Ha=Hamar, Norway. Ch=Choryn, Poland. Ce=Cerekwica, Poland. Ma=Martonvasar, Hungary. Kh=Kharkiv, Ukraine. BT=Bila Tserkva, Ukraine. Gl=Glevakha, Ukraine. Bj=Beijing, China.

Note, Lillemo et al. did not remove genotype effects as is customary when calculating Huehn's non-parametric stability statistics.

In the examples below, the results do not quite match the results of Lillemo. This could easily be the result of the original data table being rounded to 1 decimal place. For example, environment 'Aa03' had 3 reps and so the mean for genotype 1 was probably 16.333, not 16.3.

Source

Morten Lillemo, Ravi Sing, Maarten van Ginkel. (2011). Identification of Stable Resistance to Powdery Mildew in Wheat Based on Parametric and Nonparametric Methods *Crop Sci.* 50:478-485. <http://doi.org/10.2135/cropsci2009.03.0116>

Used with permission of Morten Lillemo.

Electronic data supplied by Miroslav Zoric.

References

None.

Examples

```

data(lillemo.wheat)
dat <- lillemo.wheat

# Change factor levels to match Lillemo
dat$env <- as.character(dat$env)
dat$env <- factor(dat$env,
                 levels=c("Bj03", "Bj05", "CA03", "Ba04", "Ma04",
                          "Kh06", "G105", "BT06", "Ch04", "Ce04",
                          "Ha03", "Ha04", "Ha05", "Ha07", "Aa03", "Aa04", "Aa05"))

# Interesting look at different measurement scales by environment
require(lattice)
qqmath(~score|env, dat, group=scale,
       as.table=TRUE, scales=list(y=list(relation="free")),
       auto.key=list(columns=4),
       main="lillemo.wheat - QQ plots by environment")

# -----

## Not run:
# Change data to matrix format
require(reshape2)
datm <- acast(dat, gen~env, value.var='score')

# Environment means. Matches Lillemo Table 3
apply(datm, 2, mean)

# Two different transforms within envts to approximate 0-9 scale
datt <- datm
datt[, "CA03"] <- 1.8 * datt[, "CA03"]
ix <- c("Ba04", "Kh06", "G105", "BT06", "Ha03", "Ha04", "Ha05", "Ha07", "Aa03", "Aa04", "Aa05")
datt[, ix] <- apply(datt[, ix], 2, sqrt)

# Genotype means of transformed data. Matches Lillemo table 3.
round(rowMeans(datt), 2)

# Biplot of transformed data like Lillemo Fig 2
require(gge)
biplot(gge(datt, scale=FALSE), main="lillemo.wheat")

# Median polish of transformed table
m1 <- medpolish(datt)
# Half-normal prob plot like Fig 1
# require(faraway)
# halfnorm(abs(as.vector(m1$resid)))

# Nonparametric stability statistics. Lillemo Table 4.

```

```

huehn <- function(mat){
  # Gen in rows, Env in cols
  nenv <- ncol(mat)
  # Corrected yield. Remove genotype effects
  # Remove the following line to match Table 4 of Lillemo
  mat <- sweep(mat, 1, rowMeans(mat)) + mean(mat)
  # Ranks in each environment
  rmat <- apply(mat, 2, rank)

  # Mean genotype rank across envts
  MeanRank <- apply(rmat, 1, mean)

  # Huehn S1
  gfun <- function(x){
    oo <- outer(x,x,"-")
    sum(abs(oo)) # sum of all absolute pairwise differences
  }
  S1 <- apply(rmat, 1, gfun)/(nenv*(nenv-1))

  # Huehn S2
  S2 <- apply((rmat-MeanRank)^2,1,sum)/(nenv-1)

  out <- data.frame(MeanRank,S1,S2)
  rownames(out) <- rownames(mat)
  return(out)
}
round(huehn(datm),2) # Matches table 4

# I do not think phenability package gives correct values for S1
# require(phenability)
# nahu(datm)

## End(Not run) # dontrun

```

lin.superiority

Multi-environment trial of 33 barley genotypes in 12 locations

Description

Multi-environment trial of 33 barley genotypes in 12 locations

Usage

```
data("lin.superiority")
```


Format

A data frame with 396 observations on the following 4 variables.

```
gen genotype/cultivar
region region
loc location
yield yield (kg/ha)
```

Details

Yield of six-row barley from the 1983 annual report of Eastern Cooperative Test in Canada.

The named cultivars Bruce, Conquest, Laurier, Leger are checks, while the other cultivars were tests.

Source

C. S. Lin, M. R. Binns (1985). Procedural approach for assessing cultivar-location data: Pairwise genotype-environment interactions of test cultivars with checks *Canadian Journal of Plant Science*, 1985, 65(4): 1065-1071. Table 1. <http://doi.org/10.4141/cjps85-136>

References

C. S. Lin, M. R. Binns (1988). A Superiority Measure Of Cultivar Performance For Cultivar x Location Data. *Canadian Journal of Plant Science*, 68, 193-198. <http://doi.org/10.4141/cjps88-018>

Mohammed Ali Hussein, Asmund Bjornstad, and A. H. Aastveit (2000). SASG x ESTAB: A SAS Program for Computing Genotype x Environment Stability Statistics. *Agronomy Journal*, 92; 454-459. <http://doi.org/10.2134/agronj2000.923454x>

Examples

```
data(lin.superiority)
dat <- lin.superiority

if(require(latticeExtra) & require(reshape2)){
  # calculate the superiority measure of Lin & Binns 1988

  dat2 <- acast(dat, gen ~ loc, value.var="yield")
  locmean <- apply(dat2, 2, mean)
  locmax <- apply(dat2, 2, max)
  P <- apply(dat2, 1, function(x) {
    sum((x-locmax)^2)/(2*length(x))
  })/1000
  P <- sort(P)
  round(P) # match Lin & Binns 1988 table 2, column Pi

  # atlantic & quebec regions overlap
  # require(gge)
  # m1 <- gge(yield ~ gen*loc, env.group=region, data=dat,
```

```

#           main="lin.superiority")
# biplot(m1)

# create a figure similar to Lin & Binns 1988

# add P, locmean, locmax back into the data
dat$locmean <- locmean[match(dat$loc, names(locmean))]
dat$locmax <- locmax[match(dat$loc, names(locmax))]
dat$P <- P[match(dat$gen, names(P))]
dat$gen <- reorder(dat$gen, dat$P)
xyplot(locmax ~ locmean|gen, data=dat,
       type=c('p','r'), as.table=TRUE, col="gray",
       main="lin.superiority - Superiority index",
       xlab="Location Mean",
       ylab="Yield of single cultivars (blue) & Maximum (gray)") +
xyplot(yield ~ locmean|gen, data=dat,
       type=c('p','r'), as.table=TRUE, pch=19)
}

```

lin.unbalanced

Multi-environment trial of 33 barley genotypes in 18 locations

Description

Multi-environment trial of 33 barley genotypes in 18 locations

Usage

```
data("lin.unbalanced")
```

Format

A data frame with 405 observations on the following 4 variables.

gen genotype/cultivar

loc location

yield yield (kg/ha)

region region

Details

Yield of six-row barley from the 1986 Eastern Cooperative trial

The named cultivars Bruce, Conquest, Laurier, Leger are checks, while the other cultivars were tests.

Source

C. S. Lin, M. R. Binns (1988). A Method for Assessing Regional Trial Data When The Test Cultivars Are Unbalanced With Respect to Locations. *Canadian Journal of Plant Science*, 68(4): 1103-1110. <http://doi.org/10.4141/cjps88-130>

References

None

Examples

```
data(lin.unbalanced)
dat <- lin.unbalanced

# location maximum, lin & binns table 1
aggregate(yield ~ loc, data=dat, FUN=max)

# location 'index', lin & binns, table 1
dat2 <- subset(dat, is.element(dat$gen, c('Bruce', 'Laurier', 'Leger', 'S1', 'S2',
                                         'S3', 'S4', 'S5', 'S6', 'S7', 'T1', 'T2')))
aggregate(yield ~ loc, data=dat2, FUN=mean)

if(require(reshape2)){
  # calculate the superiority measure of Lin & Binns 1988

  dat3 <- acast(dat, gen ~ loc, value.var="yield")
  #locmean <- apply(dat3, 2, mean)
  locmax <- apply(dat3, 2, max, na.rm=TRUE)
  P <- apply(dat3, 1, function(x) {
    sum((x-locmax)^2, na.rm=TRUE)/(2*length(na.omit(x)))
  })/1000
  P <- sort(P)
  round(P) # match Lin & Binns 1988 table 2, column Pi
}
```

little.splitblock *Split-block experiment of sugar beets*

Description

Split-block experiment of sugar beets.

Usage

```
data("little.splitblock")
```

Format

A data frame with 80 observations on the following 6 variables.

```
row row
col column
yield sugar beet yield, tons/acre
harvest harvest date, weeks after planting
nitro nitrogen, pounds/acre
block block
```

Details

Four rates of nitrogen, laid out as a 4x4 Latin-square experiment.

Within each column block, the sub-plots are strips (across 4 rows) of 5 different harvest dates.

The use of sub-plots as strips necessitates care when determining the error terms in the ANOVA table.

Note, Little has yield value of 22.3 for row 3, column I-H3. This data uses 23.3 in order to match the marginal totals given by Little.

Source

Thomas M. Little, F. Jackson Hills. (1978) *Agricultural Experimentation*

References

None.

Examples

```
data(little.splitblock)
dat <- little.splitblock

# Match marginal totals given by Little.
## sum(dat$yield)
## with(dat, tapply(yield,col,sum))
## with(dat, tapply(yield,row,sum))

# Layout shown by Little figure 10.2
if(require(desplot)){
  desplot(yield ~ col*row, data=dat,
          out1=block, out2=col, col=nitro, cex=1, num=harvest,
          main="little.splitblock")
}

# Convert continuous traits to factors
dat <- transform(dat, R=factor(row), C=factor(block),
                 H=factor(harvest), N=factor(nitro))
```

```
if(FALSE){  
  library(lattice)  
  xyplot(yield ~ nitro|H,dat)  
  xyplot(yield ~ harvest|N,dat)  
}  
  
# Anova table matches Little, table 10.3  
m1 <- aov(yield ~ R + C + N + H + N:H +  
          Error(R:C:N + C:H + C:N:H), data=dat)  
summary(m1)
```

lonnquist.maize

Multi-environment trial of maize, half diallel

Description

Half diallel of maize

Usage

```
data("lonnquist.maize")
```

Format

A data frame with 78 observations on the following 3 variables.

p1 parent 1 factor

p2 parent 2 factor

yield yield

Details

Twelve hybrids were selfed/crossed in a half-diallel design planted in 3 reps at 2 locations in 2 years. The data here are means adjusted for block effects.

Source

J. H. Lonquist, C. O. Gardner. (1961) Heterosis in Intervarietal Crosses in Maize and Its Implication in Breeding Procedures. *Crop Science*, 1, 179-183. Table 1.

References

Mohring, Melchinger, Piepho. (2011). REML-Based Diallel Analysis. *Crop Science*, 51, 470-478. <http://doi.org/10.2135/cropsci2010.05.0272>

C. O. Gardner and S. A. Eberhart. 1966. Analysis and Interpretation of the Variety Cross Diallel and Related Populations. *Biometrics*, 22, 439-452. <http://doi.org/10.2307/2528181>

Examples

```

data(lonnquist.maize)
dat <- lonnquist.maize
dat <- transform(dat,
  p1=factor(p1,
    levels=c("C","L","M","H","G","P","B","RM","N","K","R2","K2")),
  p2=factor(p2,
    levels=c("C","L","M","H","G","P","B","RM","N","K","R2","K2")))
require(lattice)
redblue <- colorRampPalette(c("firebrick", "lightgray", "#375997"))
levelplot(yield ~ p1*p2, dat, col.regions=redblue,
  main="lonnquist.maize - yield of diallel cross")

# Calculate the F1 means in Lonquist, table 1
if(require(reshape2)){
  mat <- acast(dat, p1~p2)
  mat[upper.tri(mat)] <- t(mat)[upper.tri(mat)] # make symmetric
  diag(mat) <- NA
  round(rowMeans(mat, na.rm=TRUE),1)
  ##   C    L    M    H    G    P    B    RM    N    K    R2    K2
  ## 94.8 89.2 95.0 96.4 95.3 95.2 97.3 93.7 95.0 94.0 98.9 102.4
}

# -----

## Not run:
# Mohring 2011 used 6 varieties to calculate GCA & SCA
# Matches Table 3, column 2
d2 <- subset(dat, is.element(p1, c("M","H","G","B","K","K2")) &
  is.element(p2, c("M","H","G","B","K","K2")))
d2 <- droplevels(d2)
# asreml4
require(asreml)
m2 <- asreml(yield~ 1, data=d2, random = ~ p1 + and(p2))
require(lucid)
vc(m2)
##   effect component std.error z.ratio    con
## p1!p1.var      3.865   3.774     1 Positive
## R!variance    15.93    5.817    2.7 Positive

# Calculate GCA effects
m3 <- asreml(yield~ p1 + and(p2), data=d2)
coef(m3)$fixed-1.462
# Matches Gardner 1966, Table 5, Griffing method

## End(Not run)

# -----

## Not run:

```

```

# Mohring 2011 used 6 varieties to calculate GCA & SCA
# Matches Table 3, column 2
## d2 <- subset(dat, p1
##           p2)
## d2 <- droplevels(d2)
## require(asreml4)
## m2 <- asreml(yield~ 1, data=d2, random = ~ p1 + and(p2))
## require(lucid)
## vc(m2)
## ##      effect component std.error z.ratio      con
## ## p1!p1.var      3.865      3.774      1 Positive
## ## R!variance      15.93      5.817      2.7 Positive

## # Calculate GCA effects
## m3 <- asreml(yield~ p1 + and(p2), data=d2)
## coef(m3)$fixed-1.462
## # Matches Gardner 1966, Table 5, Griffing method

## End(Not run)

# -----

```

love.cotton.uniformity

Uniformity trial of cotton

Description

Uniformity trial of cotton

Usage

```
data("love.cotton.uniformity")
```

Format

A data frame with 170 observations on the following 3 variables.

row row

col column

yield yield

Details

Within each 100-foot row, the first 20 feet were harvested as a single plot, and then the rest of the row was harvested in 5-foot lengths.

Crop location is unknown, but likely Southeast Asia given Love's work teaching in China. Date circa 1930.

Possibly more information would be in the collected papers of Harry Love at Cornell: <http://rmc.library.cornell.edu/EAD/html/Cotton - Plot Technic Study 1930-1932. Box 3, Folder 34>

Source

Harry Love (1936). *Application of Statistical Methods to Agricultural Research*. The Commercial Press, Shanghai.

References

None.

Examples

```
data(love.cotton.uniformity)

# omit first column which has 20-foot plots
dat <- subset(love.cotton.uniformity, col > 1)

if(require(desplot)){
  desplot(yield ~ col*row, dat, flip=TRUE,
          aspect=20/80, # just a guess
          main="love.cotton.uniformity")
}
```

lu.stability

Multi-environment trial of maize, to illustrate stability statistics

Description

Multi-environment trial to illustrate stability statistics

Usage

```
data("lu.stability")
```


Format

A data frame with 120 observations on the following 4 variables.

yield yield
 gen genotype factor, 5 levels
 env environment factor, 6 levels
 block block factor, 4 levels

Details

Data for 5 maize genotypes in 2 years x 3 sites = 6 environments.

Source

H.Y. Lu and C. T. Tien. (1993) Studies on nonparametric method of phenotypic stability: II. Selection for stability of agro-economic concept. *J. Agric. Assoc. China* 164:1-17.

References

Hsiu Ying Lu. 1995. PC-SAS Program for Estimating Huehn's Nonparametric Stability Statistics. *Agron J.* 87:888-891.

Kae-Kang Hwu and Li-yu D Liu. (2013) Stability Analysis Using Multiple Environment Trials Data by Linear Regression. (In Chinese) *Crop, Environment & Bioinformatics* 10:131-142.

Examples

```
data(lu.stability)
dat <- lu.stability

## Not run:

# GxE means. Match Lu 1995 table 1
require(reshape2)
datm <- acast(dat, gen~env, fun=mean, value.var='yield')
round(datm, 2)
# Gen/Env means. Match Lu 1995 table 3
apply(datm, 1, mean)
apply(datm, 2, mean)

# Traditional ANOVA. Match Hwu table 2
# F value for gen,env
m1 = aov(yield~env+gen+Error(block:env+env:gen), data=dat)
summary(m1)
# F value for gen:env, block:env
m2 <- aov(yield ~ gen + env + gen:env + block:env, data=dat)
summary(m2)

# Finlay Wilkinson regression coefficients
```

```

# First, calculate env mean, merge in
require(dplyr)
dat2 <- group_by(dat, env)
dat2 <- mutate(dat2, loc.mean=mean(yield))
m4 <- lm(yield ~ gen -1 + gen:loc.mean, data=dat2)
coef(m4) # Match Hwu table 4

# Table 6: Shukla's heterogeneity test
dat2$ge = gl(5,6) # Create a separate ge interaction term
m6 <- lm(yield ~ gen + env + ge + ge:loc.mean, data=dat2)
m6b <- lm( yield ~ gen + env + ge + loc.mean, data=dat2)
anova(m6, m6b) # Non-significant difference

# Table 7 - Shukla stability
# First, environment means
emn <- group_by(dat2, env)
emn <- summarize(emn, ymn=mean(yield))
# Regress GxE terms on envt means
getab = (model.tables(m2,"effects")$tables)$'gen:env'
getab
for (ll in 1:nrow(getab)){
  m7l <- lm(getab[ll, ] ~ emn$ymn)
  cat("\n\n***** Gen ",ll," *****\n")
  cat("Regression coefficient: ",round(coefficients(m7l)[2],5),"\n")
  print(anova(m7l))
} # Match Hwu table 7.

## End(Not run) # dontrun

```

lucas.switchback

Switchback trial in dairy with three treatments

Description

Switchback trial in dairy with three treatments

Usage

```
data("lucas.switchback")
```

Format

A data frame with 36 observations on the following 5 variables.

cow cow factor, 12 levels

trt treatment factor, 3 levels

period period factor, 3 levels

yield yield (FCM = fat corrected milk), pounds/day

block block factor

Details

Lucas says "because no data from feeding trials employing the present designs are yet available, uniformity data will be used".

Six cows were started together in block 1, then three cows in block 2 and three cows in block 3.

Source

Lucas, HL. 1956. Switchback trials for more than two treatments. *Journal of Dairy Science*, 39, 146-154. [http://doi.org/10.3168/jds.S0022-0302\(56\)94721-X](http://doi.org/10.3168/jds.S0022-0302(56)94721-X)

References

Sanders, WL and Gaynor, PJ. 1987. Analysis of Switchback Data Using Statistical Analysis System. *Journal of Dairy Science*, 70, 2186-2191. [http://doi.org/10.3168/jds.S0022-0302\(87\)80273-4](http://doi.org/10.3168/jds.S0022-0302(87)80273-4)

Examples

```
data(lucas.switchback)
dat <- lucas.switchback

# Create a numeric period variable
dat$per <- as.numeric(substring(dat$period,2))

require(lattice)
xyplot(yield ~ period|block, data=dat, group=cow, type=c('l','r'),
       auto.key=list(columns=6),
       main="lucas.switchback - (actually uniformity data)")

# Need to use 'terms' to preserve the order of the model terms
# Really, cow(block), per:cow(block), period(block)
m1 <- aov(terms(yield ~ block + cow:block + per:cow:block +
               period:block + trt, keep.order=TRUE), data=dat)
anova(m1) # Match Sanders & Gaynor table 3
## Analysis of Variance Table
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## block      2   30.93  15.464  55.345 5.132e-05 ***
## block:cow   9 1700.97  188.997  676.426 1.907e-09 ***
## block:cow:per 12  120.47  10.040  35.932 4.137e-05 ***
## block:period  3   14.85   4.950  17.717 0.001194 **
## trt         2    1.58   0.789   2.825 0.126048
## Residuals   7    1.96   0.279

coef(m1) # trtT2 and trtT3 match Sanders table 3 trt diffs
```

 lyon.potato.uniformity

Uniformity trial of potatoes

Description

Uniformity trial of potatoes at Nebraska Experiment Station, 1909.

Format

A data frame with 204 observations on the following 3 variables.

row row

col column, section

yield yield, pounds

Details

In 1909, potatoes were harvested from uniform land at Nebraska Experiment Station.

There were 34 rows, 34 inches apart. Lyon, page 97 says "He harvested each row in six sections, each of which was seventy-two feet and seven inches long." It is not clear if each section is 72 feet long, or if the entire row is 72 feet long. Yield of potato is roughly 0.5 to 0.8 pounds per square foot, so it seems more plausible the entire row is 72 feet long (see calculations below).

Field width: 6 plots = 72 feet

Field length: 34 rows * 34 in / 12in/ft = 96 ft

Source

Lyon, T.L. (1911). Some experiments to estimate errors in field plat tests. *Proc. Amer. Soc. Agron*, 3, 89-114. Table III. <http://doi.org/10.2134/agronj1911.00021962000300010016x>

References

None.

Examples

```
data(lyon.potato.uniformity)
dat <- lyon.potato.uniformity

# Yield per square foot, assuming 72 foot rows
sum(dat$yield)/(72*96) # 0.67 # seems about right
# Yield per square foot, assuming 72 foot plots
sum(dat$yield)/(6*72*96) # 0.11

if(require(desplot)){
```

```
desplot(yield ~ col*row, dat,  
        tick=TRUE, flip=TRUE, aspect=96/72, # true aspect  
        main="lyon.potato.uniformity")  
}
```

lyons.wheat

Multi-environment trial of winter wheat at 12 sites in 4 years.

Description

Yield of winter wheat at 12 sites in 4 years.

Format

A data frame with 48 observations on the following 3 variables.

loc location, 12 levels

year year, numeric

yield yield (kg)

Details

Krzanowski uses this briefly for multi-dimensional scaling.

Source

R. Lyons (1980). A review of multidimensional scaling. Unpublished M.Sc. dissertation, University of Reading.

References

Krzanowski, W.J. (1988) *Principles of multivariate analysis*. Oxford University Press.

Examples

```
data(lyons.wheat)  
dat <- lyons.wheat  
  
require(lattice)  
xyplot(yield~factor(year), dat, group=loc,  
        main="lyons.wheat",  
        auto.key=list(columns=4), type=c('p','l'))
```

masood.rice.uniformity

Uniformity trial of rice

Description

Uniformity trial of rice at Lahore, Punjab, circa 2011.

Usage

```
data("masood.rice.uniformity")
```

Format

A data frame with 288 observations on the following 3 variables.

row row

col column

yield yield, kg/m²

Details

Data by collected from the Rice Research Institute on a paddy yield trial. A single variety of rice was harvested in an area 12m x 24 m. Yield in kilograms was measured for each square meter. Masood et al report a low degree of similarity for neighboring plots.

Note, the Smith index calculations below match the results in the Pakistan Journal of Agricultural Research, but do not match the results in the American-Eurasian Journal, which seems to be the same paper and seems to refer to the same data. The results may simply differ by a scaling factor.

Field length: 24 plots x 1m = 24m.

Field width: 12 plots x 1m = 12m.

Source

Masood, M Asif and Raza, Irum. 2012. Estimation of optimum field plot size and shape in paddy yield trial. *Pakistan J. Agric. Res.*, Vol. 25 No. 4, 2012

Used with permission of Asif Masood.

References

Masood, M Asif and Raza, Irum. 2012. Estimation of optimum field plot size and shape in paddy yield trial. *American-Eurasian Journal of Scientific Research*, 7, 264-269. Table 1. <http://doi.org/10.5829/idosi.aejrsr.2012.7.6>

Examples

```
data(masood.rice.uniformity)
dat <- masood.rice.uniformity

if(require(desplot)) {
  desplot(yield ~ col*row, data=dat,
          flip=TRUE, tick=TRUE, aspect=24/12, # true aspect
          main="masood.rice.uniformity - yield heatmap")
}

## Not run:
require(agricolae) & require(reshape2)
dmat <- acast(dat, row~col, value.var='yield')
index.smith(dmat,
            main="masood.rice.uniformity",
            col="red") # CVs match Table 3

## End(Not run)
```

mcclelland.corn.uniformity

Uniformity trial of corn

Description

Uniformity trial of corn at Arkansas Experiment Station, 1925.

Usage

```
data("mcclelland.corn.uniformity")
```

Format

A data frame with 438 observations on the following 3 variables.

```
row row
col column
yield yield
```

Details

A uniformity trial of Paymaster corn in 1925 at the Arkansas Experimental Station.

Field length = 219 rows * 44 inches / 12 inches/ft = 803 ft.

Field width = 66ft * 2 = 132 feet.

Note: In the source document, table 2, first 'west' column and second-to-last row (page 822), the value 1.40 is assumed to be a typographical error and was changed to 14.0 for this data.

Source

McClelland, Chalmer Kirk (1926). Some determinations of plat variability. *Agronomy Journal*, 18, 819-823. <http://doi.org/10.2134/agronj1926.00021962001800090009x>

Converted from paper to electronic format by Kevin Wright

References

None.

Examples

```
data(mcclelland.corn.uniformity)
dat <- mcclelland.corn.uniformity

# McClelland table 3, first row, gives 11.2% error for 432 vales.
# Probable error = 0.67449 * sd(). Relative to mean.
0.67449 * sd(dat$yield)/mean(dat$yield) # 11.2%

if(require(desplot)){
  desplot(yield ~ col*row, dat,
          flip=TRUE,
          aspect=(219*44/12)/132, # true aspect, 219 rows * 44 inches x 132 feet
          main="mcclelland.corn.uniformity")
}
```

mcclelland.turnip

Yields of turnips grown for winter fodder

Description

Yields of turnips grown for winter fodder with two treatments for planting date and density, planted as RCB.

Format

A data frame with 64 observations on the following 6 variables.

```
gen genotype
date planting date, levels 21Aug1990 28Aug1990
density planting density, 1, 2, 4, 8 kg/ha
block block, 4 levels
yield yield
```


Details

This is a randomized block experiment with 16 treatments allocated at random to each of four blocks. The 16 treatments were combinations of two varieties, two planting dates, and four densities.

Lee et al (2008) proposed an analysis using mixed models with changing treatment variances.

Piepho (2009) proposed an ordinary ANOVA using transformed data.

Source

Statistical Modelling Using Genstat, K. J. McConway, M. C. Jones, P. C. Taylor.

Used with permission of Kevin McConway.

References

Michael Berthold, D. J. Hand. *Intelligent data analysis: an introduction*, 1998. Pages 75–82.

Lee, C.J. and O'Donnell, M. and O'Neill, M. (2008). Statistical analysis of field trials with changing treatment variance. *Agronomy Journal*, 100, 484–489.

Piepho, H.P. (2009), Data transformation in statistical analysis of field trials with changing treatment variance. *Agronomy Journal*, 101, 865–869.

Examples

```
data(mconway.turnip)
dat <- mconway.turnip
dat$densf <- factor(dat$density)

# Table 2 of Lee et al.
m0 <- aov( yield ~ gen * densf * date + block, dat )
summary(m0)
##              Df Sum Sq Mean Sq F value    Pr(>F)
## gen              1   84.0    83.95    8.753 0.00491 **
## densf            3  470.4   156.79   16.347 2.51e-07 ***
## date             1  233.7   233.71   24.367 1.14e-05 ***
## block            3  163.7    54.58    5.690 0.00216 **
## gen:densf        3    8.6     2.88    0.301 0.82485
## gen:date         1   36.5    36.45    3.800 0.05749 .
## densf:date       3  154.8    51.60    5.380 0.00299 **
## gen:densf:date   3   18.0     6.00    0.626 0.60224
## Residuals       45  431.6     9.59
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Boxplots suggest heteroskedasticity for date, density
if(require("HH")){
  interaction2wt(yield ~ gen + date + densf +block, dat,
    x.between=0, y.between=0,
    main="mconway.turnip - yield")
}
```

```

require(nlme)
# Random block model
m1 <- lme(yield ~ gen * date * densf, random= ~1|block, data=dat)
summary(m1)
anova(m1)

# Multiplicative variance model over densities and dates
m2 <- update(m1,
             weights=varComb(varIdent(form=~1|densf),
                             varIdent(form=~1|date)))
summary(m2)
anova(m2)

# Unstructured variance model over densities and dates
m3 <- update(m1, weights=varIdent(form=~1|densf*date))
summary(m3)
anova(m3)

# Table 3 of Piepho, using transformation
m4 <- aov( yield^.235 ~ gen * date * densf + block, dat )
summary(m4)

```

mcleod.barley

Multi-environment trial of barley in South Canterbury with yield and yield components

Description

Yield and yield components for barley with different seeding rates.

Format

A data frame with 40 observations on the following 10 variables.

year year, numeric
site site factor
rate rate, numeric
plants plants per sq meter
tillers tillers per plant
heads heads per plant
surviving percent surviving tillers
grains grains per head
weight weight of 1000 grains
yield yield tons/hectare

Details

Trials were conducted at 5 sites, 3 years in South Canterbury. (not all sites in every year). Values are the average of 6 blocks. In 1974 there was a severe drought. The other years had favorable growing conditions.

Source

C. C. McLeod (1982). Effects of rates of seeding on barley sown for grain. *New Zealand Journal of Experimental Agriculture*, 10, 133-136. <http://doi.org/10.1080/03015521.1982.10427857>.

References

Maindonald (1992).

Examples

```
data(mcleod.barley)
dat <- mcleod.barley

# Table 3 of McLeod. Across-environment means by planting rate
d1 <- aggregate(cbind(plants, tillers, heads, surviving, grains,
                      weight, yield) ~ rate, dat, FUN=mean)
# Calculate income based on seed cost of $280/ton, grain $140/ton.
d1 <- transform(d1, income=140*yield-280*rate/1000)
signif(d1,3)
##  rate plants tillers heads surviving grains weight yield
##   50 112.12   5.22 4.36   83.95 21.25 46.11 3.97
##   75 162.75   4.04 3.26   80.89 19.95 45.10 4.26
##  100 202.62   3.69 2.73   74.29 19.16 44.66 4.38
##  125 239.00   3.28 2.33   71.86 18.45 43.45 4.41
##  150 293.62   2.90 2.00   69.54 17.94 42.77 4.47

# Even though tillers/plant, heads/plant, surviving tillers,
# grains/head, weight/1000 grains are all decreasing as planting
# rate increases, the total yield is still increasing.
# But, income peaks around seed rate of 100.

require(lattice)
xyplot(yield +income +surviving +grains +weight +plants +tillers +heads ~ rate,
       data=d1, outer=TRUE, type=c('p','l'),
       scales=list(y=list(relation="free")),
       xlab="Nitrogen rate", ylab="Trait value",
       main="mcleod.barley - nitrogen response curves" )
```

mead.cauliflower *Leaves for cauliflower plants at different times*

Description

Leaves for cauliflower plants at different times in two years.

Format

A data frame with 14 observations on the following 4 variables.

year year factor
degdays degree days above 32F
leaves number of leaves

Details

Numbers of leaves for 10 cauliflower plants in each of two years, and temperature degree-days above 32F, divided by 100.)

The year is 1956-57 or 1957-58.

Over the data range shown, the number of leaves is increasing linearly. Extrapolating backwards shows that a linear model is inappropriate, and so a glm is used.

Source

Roger Mead, Robert N Curnow, Anne M Hasted. 2002. *Statistical Methods in Agriculture and Experimental Biology*, 3rd ed. Chapman and Hall. Page 251.

References

Mick O'Neill. *Regression & Generalized Linear (Mixed) Models*. Statistical Advisory & Training Service Pty Ltd.

Examples

```
data(mead.cauliflower)
dat <- mead.cauliflower

dat <- transform(dat, year=factor(year))

m1 <- glm(leaves ~ degdays + year, data=dat, family=poisson)
coef(m1)
## (Intercept)    degdays    year1957
## 3.49492453    0.08512651    0.21688760

dat$pred <- predict(m1, type="response")
require(lattice)
```

```
if(require(latticeExtra)){
xyplot(leaves~degdays, data=dat, groups=year, type=c('p'),
      auto.key=list(columns=2), main="mead.cauliflower - observed (symbol) & fitted (line)",
      xlab="degree days", ylab="Number of leaves", ) +
xyplot(pred~degdays, data=dat, groups=year, type=c('l'), col="black")
}
```

mead.cowpeamaize *Maize/Cowpea intercropping experiment*

Description

Maize/Cowpea Intercropping experiment with multiple nitrogen treatments.

Format

A data frame with 72 observations on the following 6 variables.

block block, 3 levels

nitro nitrogen, 4 levels

cowpea cowpea variety, 2 levels

maize maize variety, 3 levels

cyield cowpea yield, kg/ha

myield maize yield, kg/ha

Details

An intercropping experiment conducted in Nigeria. The four nitrogen treatments were 0, 40, 80, 120 kg/ha.

Source

Roger Mead. 1990. A Review of Methodology For The Analysis of Intercropping Experiments. Training Working Document No. 6. CIMMYT. <http://repository.cimmyt.org/xmlui/handle/10883/868>

References

Roger Mead, Robert N Curnow, Anne M Hasted. 2002. Statistical Methods in Agriculture and Experimental Biology, 3rd ed. Chapman and Hall. Page 390.

Examples

```

data(mead.cowpeamaize)
dat <- mead.cowpeamaize

# Cowpea and maize yields are clearly in competition
if(require("latticeExtra")){
  useOuterStrips(xyplot(myield ~ cyield|maize*cowpea, dat, group=nitro,
                        main="mead.cowpeamaize - intercropping",
                        xlab="cowpea yield",
                        ylab="maize yield", auto.key=list(columns=4)))
}

# Mead Table 2 Cowpea yield anova...strongly affected by maize variety.
anova(aov(cyield ~ block + maize + cowpea + nitro +
          maize:cowpea + maize:nitro + cowpea:nitro +
          maize:cowpea:nitro, dat))
# Cowpea mean yields for nitro*cowpea
aggregate(cyield ~ nitro+cowpea, dat, FUN=mean)
# Cowpea mean yields for each maize variety
aggregate(cyield ~ maize, dat, FUN=mean)

# Bivariate analysis
aov.c <- anova(aov(cyield/1000 ~ block + maize + cowpea + nitro +
                  maize:cowpea + maize:nitro + cowpea:nitro +
                  maize:cowpea:nitro, dat))

aov.m <- anova(aov(myield/1000 ~ block + maize + cowpea + nitro +
                  maize:cowpea + maize:nitro + cowpea:nitro +
                  maize:cowpea:nitro, dat))

aov.cm <- anova(aov(cyield/1000 + myield/1000 ~ block + maize + cowpea + nitro +
                   maize:cowpea + maize:nitro + cowpea:nitro +
                   maize:cowpea:nitro, dat))

biv <- cbind(aov.m[,1:2], aov.c[,2], aov.cm[,2])
names(biv) <- c('df', 'maize ss', 'cowpea ss', 'ss for sum')
biv$'sum of prod' <- (biv[,4] - biv[,2] - biv[,3] ) /2
biv$cor <- biv[,5]/(sqrt(biv[,2] * biv[,3]))
signif(biv,2)
##
## df maize ss cowpea ss ss for sum sum of prod cor
## block 2 0.290 0.0730 0.250 -0.058 -0.400
## maize 2 18.000 0.4100 13.000 -2.600 -0.980
## cowpea 1 0.027 0.0060 0.058 0.013 1.000
## nitro 3 29.000 0.1100 25.000 -1.800 -0.980
## maize:cowpea 2 1.100 0.0099 0.920 -0.099 -0.950
## maize:nitro 6 1.300 0.0680 0.920 -0.200 -0.680
## cowpea:nitro 3 0.240 0.1700 0.150 -0.130 -0.640
## maize:cowpea:nitro 6 1.300 0.1400 1.300 -0.033 -0.079
## Residuals 46 16.000 0.6000 14.000 -1.400 -0.460

```

| | |
|------------------|--|
| mead.germination | <i>Seed germination with different temperatures/concentrations</i> |
|------------------|--|

Description

Seed germination with different temperatures/concentrations

Format

A data frame with 64 observations on the following 5 variables.

temp temperature regimen
 rep replication factor (not blocking)
 conc chemical concentration
 germ number of seeds germinating
 seeds number of seeds tested = 50

Details

The rep factor is NOT a blocking factor.

Source

Roger Mead, Robert N Curnow, Anne M Hasted. 2002. *Statistical Methods in Agriculture and Experimental Biology*, 3rd ed. Chapman and Hall. Page 350-351.

Used with permission of Roger Mead, Robert Curnow, and Anne Hasted.

References

Schabenberger, O. and Pierce, F.J., 2002. *Contemporary statistical models for the plant and soil sciences*, CRC.

Examples

```
data(mead.germination)
dat <- mead.germination
dat <- transform(dat, concf=factor(conc))
require(lattice)
xyplot(germ~log(conc+.01)|temp, dat, layout=c(4,1),
       main="mead.germination", ylab="number of seeds germinating")

m1 <- glm(cbind(germ, seeds-germ) ~ 1, dat, family=binomial)
m2 <- glm(cbind(germ, seeds-germ) ~ temp, dat, family=binomial)
m3 <- glm(cbind(germ, seeds-germ) ~ concf, dat, family=binomial)
m4 <- glm(cbind(germ, seeds-germ) ~ temp + concf, dat, family=binomial)
m5 <- glm(cbind(germ, seeds-germ) ~ temp * concf, dat, family=binomial)
```

```

anova(m1,m2,m3,m4,m5)
##   Resid. Df Resid. Dev Df Deviance
## 1      63   1193.80
## 2      60    430.11  3   763.69
## 3      60    980.10  0  -549.98
## 4      57    148.11  3   831.99
## 5      48     55.64  9    92.46

# Show logit and fitted values. T2 has highest germination
subset(cbind(dat, predict(m5), fitted(m5)), rep=="R1")

```

mead.lamb

Number of lambs born to 3 breeds on 3 farms

Description

Number of lambs born to 3 breeds on 3 farms

Usage

```
data("mead.lamb")
```

Format

A data frame with 36 observations on the following 4 variables.

farm farm: F1, F2, F3

breed breed: B1, B2, B3

lambclass lambing class: L0, L1, L2, L3

y count of ewes in class

Details

The data 'y' are counts of ewes in different lambing classes. The classes are number of live lambs per birth for 0, 1, 2, 3+ lambs.

Source

Roger Mead, Robert N Curnow, Anne M Hasted. 2002. Statistical Methods in Agriculture and Experimental Biology, 3rd ed. Chapman and Hall. Page 359.

References

None

Examples

```

data(mead.lamb)
dat <- mead.lamb

# farm 1 has more ewes in lambclass 3
d2 <- xtabs(y ~ farm+breed+lambclass, data=dat)
mosaicplot(d2, color=c("lemonchiffon1","moccasin","lightsalmon1","indianred"),
           xlab="farm/lambclass", ylab="breed", main="mead.lamb")

names(dat) <- c('F','B','L','y') # for compactness
# Match totals in Mead example 14.6
if(require(dplyr)){
  dat <- group_by(dat, F,B)
  summarize(dat, y=sum(y))
  ##           F           B           y
  ## <fctr> <fctr> <int>
  ## 1     F1         A     150
  ## 2     F1         B      46
  ## 3     F1         C      78
  ## 4     F2         A      72
  ## 5     F2         B      79
  ## 6     F2         C      28
  ## 7     F3         A     224
  ## 8     F3         B     129
  ## 9     F3         C      34
}

# Models
m1 <- glm(y ~ F + B + F:B,
          data=dat,
          family=poisson(link=log))
m2 <- update(m1, y ~ F + B + F:B + L)
m3 <- update(m1, y ~ F + B + F:B + L + B:L)
m4 <- update(m1, y ~ F + B + F:B + L + F:L)
m5 <- update(m1, y ~ F + B + F:B + L + B:L + F:L)

AIC(m1, m2, m3, m4, m5) # Model 4 has best AIC
##      df      AIC
## m1  9 852.9800
## m2 12 306.5457
## m3 18 303.5781
## m4 18 206.1520
## m5 24 213.8873

# Change contrasts for Miroslav
m4 <- update(m4,
             contrasts=list(F=contr.sum,B=contr.sum,L=contr.sum))
summary(m4)

# Match deviance table from Mead
if(require(broom)){

```

```

all <- do.call(rbind, lapply(list(m1, m2, m3, m4, m5), broom::glance))
rownames(all) <- unlist(lapply(list(m1, m2, m3, m4, m5),
                              function(x) as.character(formula(x)[3])))
all[,c('deviance', 'df.residual')]
##           deviance df.residual
## F + B + F:B      683.67257      27
## F + B + L + F:B   131.23828      24
## F + B + L + F:B + B:L 116.27069      18
## F + B + L + F:B + F:L  18.84460      18
## F + B + L + F:B + B:L + F:L 14.57987      12
}

# Using MASS::loglm
if(require(MASS)){
  # Note: without 'fitted=TRUE', devtools::run_examples has an error
  m4b <- MASS::loglm(y ~ F + B + F:B + L + F:L, data = dat, fitted=TRUE)
  # Table of farm * class interactions. Match Mead p. 360
  round(coef(m4b)$F.L,2)
  fitted(m4b)
  resid(m4b)
  # require(vcd)
  # mosaic(m4b, shade=TRUE,
  # formula = ~ F + B + F:B + L + F:L,
  # residual_type="rstandard", keep_aspect=FALSE)
}

```

mead.strawberry

Strawberry yields in RCB experiment

Description

Strawberry yields in RCB experiment

Format

A data frame with 32 observations on the following 5 variables.

row row

col column

block block, 4 levels

gen genotype, 8 levels

yield yield, pounds

Details

A hedge along the right side (column 8) caused shading and lower yields.

R. Mead said (in a discussion of the Besag & Higdon paper), "the blocks defined (as given to me by the experimenter) are the entire horizontal rows...the design of the trial is actually (and unrecognized by me also) a checker-board of eight half-blocks with two groups of split-plot varieties".

The two sub-groups of genotypes are G, V, R1, F and Re, M, E, P.

Source

Unknown, but prior to 1968 according to Besag. Probably via R. Mead.

References

R. Mead, 1990, *The Design of Experiments*.

Julian Besag and D Higdon, 1999. Bayesian Analysis of Agricultural Field Experiments, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61, 691–746. Table 4.

Examples

```
data(mead.strawberry)
dat <- mead.strawberry
dat$sub <- ifelse(is.element(dat$gen, c('G', 'V', 'R1', 'F')),
                "S1", "S2")

if(require(desplot)){
  desplot(yield~col*row, dat,
          text=gen, cex=1, out1=block, out2=sub, # unknown aspect
          main="mead.strawberry")
}
```

mercer.mangold.uniformity

Uniformity trial of mangolds

Description

Uniformity trial of mangolds at Rothamsted Experiment Station, England, 1910.

Usage

```
data("mercer.mangold.uniformity")
```

Format

A data frame with 200 observations on the following 4 variables.

```
row row
col column
roots root yields, pounds
leaves leaf yields, pounds
```

Details

Grown in 1910. Each plot was 3 drills, each drill being 2.4 feet wide. Plots were 1/200 acres, 7.2 feet by 30.25 feet long The "length of the plots runs with the horizontal lines of figures [in Table I], this being also the direction of the drills across the field."

Field width: 10 plots * 30.25ft = 302.5 feet

Field length: 20 plots * 7.25 ft = 145 feet

Source

Mercer, WB and Hall, AD, 1911. The experimental error of field trials *The Journal of Agricultural Science*, 4, 107-132. Table 1. <http://doi.org/10.1017/S002185960000160X>

References

McCullagh, P. and Clifford, D., (2006). Evidence for conformal invariance of crop yields, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 462, 2119–2143. <http://doi.org/10.1098/rspa.2006.1667>

Examples

```
data(mercerc.mangold.uniformity)
dat <- mercerc.mangold.uniformity

if(require(desplot)){
  desplot(leaves~col*row, data=dat,
          aspect=145/302, # true aspect
          main="mercerc.mangold.uniformity - leaves")
}
if(require(desplot)){
  desplot(roots~col*row, data=dat,
          aspect=145/302, # true aspect
          main="mercerc.mangold.uniformity - roots")
}

## Not run:
require(lattice)
xyplot(roots~leaves, data=dat)

## End(Not run)
```

mercer.wheat.uniformity

Uniformity trial of wheat

Description

Uniformity trial of wheat at Rothamsted Experiment Station, England, 1910.

Format

A data frame with 500 observations on the following 4 variables.

row row

col column

grain grain yield, pounds

straw straw yield, pounds

Details

The wheat crop was grown in the summer of 1910 at Rothamsted Experiment Station (Harpenden, Hertfordshire, England). In the Great Knott, a seemingly uniform area of 1 acre was harvested in separate plots, each 1/500th acre in size. The grain and straw from each plot was weighed separately.

McCullagh gives more information about the plot size.

Field width: 25 plots * 8 ft = 200 ft

Field length: 20 plots * 10.82 ft = 216 ft

D. G. Rossiter (2014) uses this data for an extensive data analysis tutorial.

Source

Mercer, WB and Hall, AD, (1911). The experimental error of field trials *The Journal of Agricultural Science*, 4, 107-132. Table 5. <http://doi.org/10.1017/S002185960000160X>

References

McCullagh, P. and Clifford, D., (2006). Evidence for conformal invariance of crop yields, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 462, 2119–2143. <http://doi.org/10.1098/rspa.2006.1667>

D. G. Rossiter (2014). Tutorial: Using the R Environment for Statistical Computing An example with the Mercer & Hall wheat yield dataset.

G. A. Baker (1941). Fundamental Distribution of Errors for Agricultural Field Trials. *National Mathematics Magazine*, 16, 7-19. <http://doi.org/10.2307/3028105>

The 'spdep' package includes the grain yields (only) and spatial positions of plot centres in its example dataset 'wheat'.

Note, checked that all '4.03' values in this data match the original document.

Examples

```

data(mercer.wheat.uniformity)
dat <- mercer.wheat.uniformity

if(require(desplot)){
  desplot(grain ~ col*row, dat,
          aspect=216/200, # true aspect
          main="mercer.wheat.uniformity - grain yield")
}

require("lattice")
xyplot(straw ~ grain, data=dat, type=c('p','r'),
       main="mercer.wheat.uniformity - regression")

## Not run:
require(hexbin)
hexbinplot(straw ~ grain, data=dat)

## End(Not run)

if(require(sp) & require(gstat)) {
  plot.wid <- 2.5
  plot.len <- 3.2
  nr <- length(unique(dat$row))
  nc <- length(unique(dat$col))

  xy <- expand.grid(x = seq(plot.wid/2, by=plot.wid, length=nc),
                  y = seq(plot.len/2, by=plot.len, length=nr))
  dat.sp <- dat
  coordinates(dat.sp) <- xy

  # heatmap
  splot(dat.sp, zcol = "grain", cuts=8,
        cex = 1.6,
        col.regions = bpy.colors(8),
        main = "Grain yield", key.space = "right")

  # variogram
  vg <- variogram(grain ~ 1, dat.sp, cutoff = plot.wid * 10, width = plot.wid)
  plot(vg, plot.numbers = TRUE,
       main="mercer.wheat.uniformity - variogram")
}

```

Description

This is monthly weather summaries for the 6 sites where barley yield trials were conducted.

Format

A data frame with 719 observations on the following 8 variables.

site site, 6 levels
 year year, 1927-1936
 mo month, 1-12, numeric
 cdd monthly cooling degree days, Fahrenheit
 hdd monthly heating degree days, Fahrenheit
 precip monthly precipitation, inches
 min monthly average daily minimum temp, Fahrenheit
 max monthly average daily maximum temp, Fahrenheit

Details

When the weather data was extracted from the National Climate Data Center, the following weather stations were chosen, based on availability of weather data in the given time frame (1927-1936) and the proximity to the town (site) for the barley data.

| site | station name | station |
|-------------|-------------------------------------|-------------|
| Morris | MORRIS WC EXPERIMENTAL STATION | USC00215638 |
| StPaul | MINNEAPOLIS WEATHER BUREAU DOWNTOWN | USC00215433 |
| Crookston | CROOKSTON NW EXPERIMENTAL STATION | USC00211891 |
| GrandRapids | GRAND RAPIDS FRS LAB | USC00213303 |
| Waseca | WASECA EXPERIMENTAL STATION | USC00218692 |
| Duluth | SUPERIOR | USC00478349 |

'cdd' are cooling degree days, which is the number of degree days with a temperature above 65 Fahrenheit.

'hdd' are heating degree days, below 65 Fahrenheit.

No data is available for Duluth in Dec, 1931.

Source

National Climate Data Center, <http://www.ncdc.noaa.gov/>.

References

Kevin Wright. 2013. Revisiting Immer's Barley Data. *The American Statistician*, 67, 129-133. <http://doi.org/10.1080/00031305.2013.801783>

Examples

```

data(minnesota.barley.yield)
dat <- minnesota.barley.yield
data( minnesota.barley.weather)
datw <- minnesota.barley.weather

# Weather trends over time
if(require(latticeExtra)){
  useOuterStrips(xyplot(cdd~mo|year*site, datw, groups=year,
                        main="minnesota.barley",
                        xlab="month", ylab="Cooling degree days",
                        subset=(mo > 3 & mo < 10),
                        scales=list(alternating=FALSE),
                        type='l', auto.key=list(columns=5)))
}

# Total cooling/heating/precip in Apr-Aug for each site/yr
ww <- subset(datw, mo>=4 & mo<=8)
ww <- aggregate(cbind(cdd,hdd,precip)~site+year, data=ww, sum)

# Average yield per each site/env
yy <- aggregate(yield~site+year, dat, mean)

minn <- merge(ww, yy)

# Higher yields generally associated with cooler temps, more precip
if(require(reshape2)){
  me <- melt(minn, id.var=c('site','year'))
  mey <- subset(me, variable=="yield")
  mey <- mey[,c('site','year','value')]
  names(mey) <- c('site','year','y')
  mec <- subset(me, variable!="yield")
  names(mec) <- c('site','year','covar','x')
  mecy <- merge(mec, mey)
  mecy$yr <- factor(mecy$year)
  oldpar <- tpg <- trellis.par.get()
  tpg$superpose.symbol$pch <- substring(levels(mecy$yr),4) # Last digit of year
  trellis.par.set(tpg)
  foo <- xyplot(y~x|covar*site, data=mecy, groups=yr, cex=1, ylim=c(5,65),
               xlab="", ylab="yield", main="minnesota.barley",
               panel=function(x,y,...) {
                 panel.lmline(x,y,..., col="gray")
                 panel.superpose(x,y,...)
               },
               scales=list(x=list(relation="free")))
  if(require(latticeExtra)){
    foo <- useOuterStrips(foo, strip.left = strip.custom(par.strip.text=list(cex=.7)))
    combineLimits(foo, margin.x=2L) # Use a common x axis for all rows
  }
} # reshape2

```

 minnesota.barley.yield

Multi-environment trial of barley in Minnesoty at 6 sites in 1927-1936.

Description

This is an expanded version of the barley data that is often used to illustrate dot plots.

Format

A data frame with 647 observations on the following 4 variables.

yield yield in bu/ac

gen genotype (variety) factor

year year

site site factor, 6 levels

Details

The `lattice` package contains a smaller version of this data for the years 1931 and 1932.

The following comments are in reference to the mentioned source documents.

Notes about Immer (1934).

The University Farm location is at the city of Saint Paul.

This source provides the yield data for each of the three blocks at each location in 1931 and 1932.

The following registration numbers and names are given:

| C.I. number | Variety name |
|-------------|------------------------|
| Minn 184 | Manchuria |
| Minn 445 | Glabron |
| Minn 440 | Svansota |
| Minn 447 | Velvet |
| Minn 448 | Trebi |
| Minn 457 | Manchuria x Smooth Awn |
| Minn 462 | Smooth Awn x Manchuria |
| Minn 452 | Peatland |
| Minn 475 | Svanhals x Lion |
| Minn 529 | Wisconsin No 38 |

Notes about Harlan (1935):

The 1931 yields match the average values of Immer (1934).

The Minnesota 474 and 475 cultivars are both 'Svanhals x Lion' crosses.

No yields are reported at Crookston in 1928 because of a crop failure. (Page 20)

Also, in the report for North Dakota it says "the zero yields at Williston, ND in 1931 were caused by drought". (Page 31)

Notes about Wiebe (1940):

The 1932 data generally match the average values from Immer (1934) with the following notes.

The data for Glabron at St Paul in 1932 are missing, but given as 36.8 in Immer (1934). This value is treated as missing in this R dataset.

The data for Svansota at Morris in 1932 are missing, but given as 35.0 in Immer (1934). This value is treated as missing in this R dataset.

The yield for 'Wisconsin 38' at St Paul in 1932 is shown as 3.80, but 38 in Immer (1934). The latter value is used in this R dataset.

The yields for No475 in 1932 are not reported in Wiebe (1940), but are reported in Immer (1934).

No yields are reported at Morris in 1933 and 1934, because of a crop failure owing to drought.

Notes about Hayes (1942).

This sources gives the block-level yield data for 5 cultivars at 4 sites in 1932 and 1935. Cultivar 'Barbless' is the same as 'Wisconsin No38'.

Source

H V Harlan and P R Cowan and Lucille Reinbach. (1935). *Yields of barley varieties in the United States and Canada, 1927-1931*. United States Dept of Agriculture. <http://naldc.nal.usda.gov/download/CAT86200440/PDF>

Gustav A. Wiebe, Philip Russell Cowan, Lucille Reinbach-Welch. (1940). *Yields of barley varieties in the United States and Canada, 1932-36*. United States Dept of Agriculture. <http://books.google.com/books?id=OUfxLocnp>

References

Immer, R. F. and H. K. Hayes and LeRoy Powers. (1934). Statistical Determination of Barley Varietal Adaptation. *Journal of the American Society of Agronomy*, 26, 403-419.

Hayes, H.K. and Immer, F.R. (1942). *Methods of plant breeding*. McGraw Hill.

Kevin Wright. (2013). Revisiting Immer's Barley Data. *The American Statistician*, 67, 129-133. <http://doi.org/10.1080/00031305.2013.801783>

Examples

```
data(minnesota.barley.yield)
dat <- minnesota.barley.yield
dat$yr <- factor(dat$year)

# Drop Dryland, Jeans, CompCross, MechMixture because they have less than 5
# year-loc values
dat <- droplevels(subset(dat, !is.element(gen, c("CompCross", "Dryland", "Jeans", "MechMixture"))))

# 1934 has huge swings from one loc to the next
require(lattice)
dotplot(gen~yield|site, dat, groups=yr,
```

```
main="minnesota.barley.yield",  
auto.key=list(columns=5), scales=list(y=list(cex=.5)))
```

montgomery.wheat.uniformity

Uniformity trial of wheat, 2 years on the same land

Description

Uniformity trial of wheat at Nebraska Experiment Station, 1909 & 1911.

Usage

```
data("montgomery.wheat.uniformity")
```

Format

A data frame with 448 observations on the following 3 variables.

year year

col column

row row

yield yield, grams

Details

Experiments were conducted by the Nebraska Experiment Station.

A field was sown to Turkey winter wheat in the fall of 1908 and harvested in 1909. The drill, 5.5 feet wide, was driven across the first series of 14 blocks, the boundaries of the blocks being later established. Each series was sown the same way, no space was allowed between the blocks. Each block was 5.5 ft square.

The experiment was done 3 times with harvests in 1909, 1910, 1911. A simple heatmap of the 3 years' yields are shown in Montgomery (1912), figure 3, p. 178.

The 1909 data are given by Montgomery (1913), figure 10, page 37. NOTE: North is at the right side of this diagram.

The 1910 data are not available.

The 1911 data are given by Montgomery (1912), figure 1, page 165. NOTE: North is at the top of this diagram.

Surface & Pearl (1916) give a simple method for adjusting yield due to fertility effects using the 1909 data.

Field length: 16 blocks

Field width: 14 plots

Source

E. G. Montgomery (1912). Variation in Yield and Methods of Arranging Plats To Secure Comparative Results. *Twenty-Fifth Annual Report of the Agricultural Experiment Station of Nebraska*, 164-180. <https://books.google.com/books?id=M-5BAQAAMAAJ&pg=RA4-PA164>

E. G. Montgomery (1913). Experiments in Wheat Breeding: Experimental Error In The Nursery and Variation in Nitrogen and Yield. U.S. Dept of Agriculture, Bureau of Plant Industry, Bulletin 269. Figure 10, page 37. <http://doi.org/10.5962/bhl.title.43602>

References

Surface & Pearl, (1916). A method of correcting for soil heterogeneity in variety tests. *Journal of Agricultural Research*, 5, 22, 1039-1050. Figure 2. <https://books.google.com/books?id=BVNyoZXFVSkC&pg=PA1039>

Examples

```
data(montgomery.wheat.uniformity)
dat <- montgomery.wheat.uniformity
dat09 <- subset(dat, year==1909)
dat11 <- subset(dat, year==1911)

# Match the figures of Montgomery Fig 3, p. 178
if(require(desplot)){
  desplot(yield ~ col*row, dat09,
          aspect=1, # true aspect
          main="montgomery.wheat.uniformity - 1909 yield")
  desplot(yield ~ col*row, subset(dat, year==1911),
          aspect=1, # true aspect
          main="montgomery.wheat.uniformity - 1911 yield")
}

## Not run:
# Surface & Pearl adjust 1909 yield for fertility effects.
# They calculate smoothed yield as (row sum)*(column sum)/(total)
# and subtract this from the overall mean to get 'deviation'.
# We can do something similar with a linear model with rows and columns
# as factors, then predict yield to get the smooth trend.
# Corrected yield = observed - deviation = observed - (smooth-mean)

m1 <- lm(yield ~ factor(col) + factor(row), data=dat09)
dev1 <- predict(m1) - mean(dat09$yield)
# Corrected. Similar (but not exact) to Surface, fig 2.
dat09$correct <- round(dat09$yield - dev1,0)

require(desplot)
desplot(yield ~ row*col, dat09, main="observed", shorten="none", text=yield, flip=TRUE)
desplot(correct ~ row*col, dat09, text=correct,
        cex=0.8, shorten="none", flip=TRUE,
        main="corrected")
# Corrected yields are slightly shrunk toward overall mean
```

```

plot(correct~yield,dat09, xlim=c(350,1000), ylim=c(350,1000))
abline(0,1)

## End(Not run)

```

| | |
|------------------|--|
| moore.uniformity | <i>Uniformity trials of pole beans, bush beans, sweet corn, carrots, spring and fall cauliflower</i> |
|------------------|--|

Description

Uniformity trials of pole beans, bush beans, sweet corn, carrots, spring and fall cauliflower at Washington, 1952-1955.

Format

Each data frame has the following columns at a minimum. Some datasets have an additional trait column.

```

row row
col column
yield yield

```

Details

All trials were grown on sandy loam soil in the Puyallup valley of Washington. In most experiments a gradient in soil fertility was evident. Moore & Darroch appear to have assigned 4 treatments to the plots and used the residual variation to calculate a CV. In the examples below a 'raw' CV is calculated and is always higher than the CV given by Moore & Darroch.

Blue Lake Pole Beans.

Conducted 1952. Seven pickings were made at about 5-day intervals. Table 26.

Field width: 12 rows x 5 feet = 60 feet.

Field length: 12 ranges x 10 feet = 120 feet.

Bush Beans.

Conducted in 1955. Two harvests. Table 27.

Field width: 24 rows x 3 feet = 72 feet.

Field length: 24 ranges x 5 feet = 120 feet.

Sweet Corn.

Conducted 1952. Table 28-29.

Field width: 24 rows x 3 feet = 72 feet.

Field length: 12 ranges x 10 feet = 120 feet.

Carrots.

Conducted 1952. Table 30.

Field width: 24 rows * 1.5 feet = 36 feet.

Field length: 12 ranges * 5 feet = 60 feet.

Spring Cauliflower.

Conducted spring 1951. Five harvests. Table 31-32.

Field width: 12 rows x 3 feet = 36 feet.

Field length: 10 plants * 1.5 feet * 20 ranges = 300 feet.

Fall Cauliflower.

Conducted fall 1951. Five harvests. Table 33-34.

Field width: 12 rows x 3 feet = 36 feet.

Field length: 10 plants * 1.5 feet * 20 ranges = 300 feet.

Source

Moore, John F and Darroch, JG. (1956). Field plot technique with Blue Lake pole beans, bush beans, carrots, sweet corn, spring and fall cauliflower. Washington Agricultural Experiment Stations, Institute of Agricultural Sciences, State College of Washington.

References

None.

Examples

```
cv <- function(x) sd(x)/mean(x)

if(require(desplot)){
  # Pole Bean
  data(moore.polebean.uniformity)
  cv(moore.polebean.uniformity$yield) # 8.00. Moore says 6.73.
  desplot(yield~col*row, moore.polebean.uniformity,
          flip=TRUE, tick=TRUE, aspect=120/60, # true aspect
          main="moore.polebean.uniformity - yield")
}

if(require(desplot)){
  # Bush bean
  data(moore.bushbean.uniformity)
  cv(moore.bushbean.uniformity$yield) # 12.1. Moore says 10.8
  desplot(yield~col*row, moore.bushbean.uniformity,
          flip=TRUE, tick=TRUE, aspect=120/72, # true aspect
          main="moore.bushbean.uniformity - yield")
}

if(require(desplot)){
  # Sweet corn
  data(moore.sweetcorn.uniformity)
  cv(moore.sweetcorn.uniformity$yield) # 17.5. Moore says 13.6
```

```

desplot(yield~col*row, moore.sweetcorn.uniformity,
        flip=TRUE, tick=TRUE, aspect=120/72, # true aspect
        main="moore.sweetcorn.uniformity - yield")
## desplot(ears~col*row, moore.sweetcorn.uniformity,
##        flip=TRUE, tick=TRUE, aspect=120/72, # true aspect
##        main="moore.sweetcorn.uniformity - ears")
## library(lattice)
## xyplot(yield ~ ears, moore.sweetcorn.uniformity)
}

if(require(desplot)){
# Carrots
data(moore.carrots.uniformity)
cv(moore.carrots.uniformity$yield) # 33.4. Moore says 27.6
desplot(yield~col*row, moore.carrots.uniformity,
        flip=TRUE, tick=TRUE, aspect=60/36, # true aspect
        main="moore.carrots.uniformity - yield")
}

if(require(desplot)){
# Spring cauliflower
data(moore.springcauliflower.uniformity)
cv(moore.springcauliflower.uniformity$yield) # 21. Moore says 19.5
desplot(yield~col*row, moore.springcauliflower.uniformity,
        flip=TRUE, tick=TRUE, aspect=300/36, # true aspect
        main="moore.springcauliflower.uniformity - yield")
## desplot(heads~col*row, moore.springcauliflower.uniformity,
##        flip=TRUE, tick=TRUE, aspect=300/36, # true aspect
##        main="moore.springcauliflower.uniformity - heads")
## library(lattice)
## xyplot(yield ~ heads, moore.springcauliflower.uniformity)
}

if(require(desplot)){
# Fall cauliflower
data(moore.fallcauliflower.uniformity)
cv(moore.fallcauliflower.uniformity$yield) # 17.7. Moore says 17.0
desplot(yield~col*row, moore.fallcauliflower.uniformity,
        flip=TRUE, tick=TRUE, aspect=300/36, # true aspect
        main="moore.fallcauliflower.uniformity - yield")
## desplot(heads~col*row, moore.fallcauliflower.uniformity,
##        flip=TRUE, tick=TRUE, aspect=300/36, # true aspect
##        main="moore.fallcauliflower.uniformity - heads")
## library(lattice)
## xyplot(yield ~ heads, moore.fallcauliflower.uniformity)
}

```

Description

Yields and acres harvested in each state for the major agricultural crops in the United States, from approximately 1900 to 2011. Crops include: barley, corn, cotton, hay, rice, sorghum, soybeans, wheat.

Usage

```
nass.barley
nass.corn
nass.cotton
nass.hay
nass.sorghum
nass.wheat
nass.rice
nass.soybean
```

Format

```
year year
state state factor
acres acres harvested
yield average yield
```

Details

Be cautious with yield values for states with small acres harvested.

Yields are in bushels/acre, except: cotton pounds/acre, hay tons/acre, rice pounds/acre.

Each crop is in a separate dataset: nass.barley, nass.corn, nass.cotton, nass.hay, nass.sorghum, nass.wheat, nass.rice, nass.soybean.

Source

United States Department of Agriculture, National Agricultural Statistics Service. <http://quickstats.nass.usda.gov/>

Examples

```
data(nass.corn)
dat <- nass.corn

# Use only states that grew at least 100K acres of corn in 2011
keep <- droplevels(subset(dat, year == 2011 & acres > 100000))$state
dat <- droplevels(subset(dat, is.element(state, keep)))

# Acres of corn grown each year
require(lattice)
xyplot(acres ~ year|state, dat, type='l', as.table=TRUE,
       main="nass.corn: state trends in corn acreage")
```



```
## Plain levelplot, using only states
## require(reshape2)
## datm <- acast(dat, year~state, value.var='yield')
## redblue <- colorRampPalette(c("firebrick", "lightgray", "#375997"))
## levelplot(datm, aspect=.7, col.regions=redblue,
##           main="nass.corn",
##           scales=list(x=list(rot=90, cex=.7)))

# Model the rate of genetic gain in Illinois as a piecewise regression
# Breakpoints define periods of open-pollinated varieties, double-cross,
# single-cross, and transgenic hybrids.

dil <- subset(nass.corn, state=="Illinois" & year >= 1900)
m1 <- lm(yield ~ pmin(year,1932) + pmax(1932, pmin(year, 1959)) +
         pmax(1959, pmin(year, 1995)) + pmax(1995, year), dil)
signif(coef(m1)[-1],3) # Rate of gain for each segment
plot(yield ~ year, dil, main="nass.corn: piecewise linear model of Illinois corn yields")
lines(dil$year, fitted(m1))
abline(v=c(1932,1959,1995), col="wheat")
```

nebraska.farmincome *Nebraska farm income in 2007 by county*

Description

Nebraska farm income in 2007 by county

Format

A data frame with 93 observations on the following 4 variables.

county county

crop crop income, thousand dollars

animal livestock and poultry income, thousand dollars

area area of each county, square miles

Details

The variables for each county are:

Value of farm products sold - crops (NAICS) 2007 (adjusted)

Value of farm products sold - livestock, 2007 (adjusted).

Area in square miles.

Note: Cuming county is a very important beef-producing county. Some counties are not reported to protect privacy. Western Nebraska is dryer and has lower income. South-central Nebraska is irrigated and has higher crop income per square mile.

Source

U.S. Department of Agriculture-National Agriculture Statistics Service. <http://censtats.census.gov/usa/usa.shtml>

Examples

```

data(nebraska.farmincome)
dat <- nebraska.farmincome

if(require("maps") & require("mapproj") & require("latticeExtra")) {
  # latticeExtra for mapplot

  dat$stco <- paste0('nebraska,', dat$county)
  # Scale to million dollars per county
  dat <- transform(dat, crop=crop/1000, animal=animal/1000)

  # Raw, county-wide incomes. Note the outlier Cuming county
  redblue <- colorRampPalette(c("firebrick", "lightgray", "#375997"))
  mapplot(stco ~ crop + animal, data = dat, colramp=redblue,
    main="nebraska.farmincome",
    xlab="Farm income from animals and crops (million $ per county)",
    scales = list(draw = FALSE),
    map = map('county', 'nebraska', plot = FALSE, fill = TRUE,
      projection = "mercator") )

  # Now scale to income/mile^2
  dat <- within(dat, {
    crop.rate <- crop/area
    animal.rate <- animal/area
  })
  # And use manual breakpoints.
  mapplot(stco ~ crop.rate + animal.rate, data = dat, colramp=redblue,
    main="nebraska.farmincome: income per square mile (percentile breaks)",
    xlab="Farm income (million $ / mi^2) from animals and crops",
    scales = list(draw = FALSE),
    map = map('county', 'nebraska', plot = FALSE, fill = TRUE,
      projection = "mercator"),
    # Percentile break points
    # breaks=quantile(c(dat$crop.rate, dat$animal.rate),
    #   c(0,.1,.2,.4,.6,.8,.9,1), na.rm=TRUE)
    # Fisher-Jenks breakpoints via classInt package
    # breaks=classIntervals(na.omit(c(dat$crop.rate, dat$animal.rate)),
    #   n=7, style='fisher')$brks
    breaks=c(0,.049, .108, .178, .230, .519, .958, 1.31))
}

```

Description

Uniformity trials of sweet corn in Alberta, 1956.

Usage

```
data("nonnecke.corn.uniformity")
```

Format

A data frame:

loc location

row row

col column

yield yield of marketable ears, pounds

Details

Experiments were conducted at three locations in Southern Alberta at Lethbridge, Vauxhall, and Cranford in 1956. Plot layout was 32 rows, each 179 feet long, allowing 18 ten-foot plots per row. Rows were 3 feet apart, thinned to one foot between plants. A double guard row surrounded the entire plot. The same two persons were assigned to harvest the corn from all locations. All 576 plots were harvested in one day. Optimal plot sizes were found to be 10ft x 6ft or 20ft by 3ft. The R data uses row/column for plot/row.

Field width: 18 plots * 10 ft = 180 feet

Field length: 32 rows * 3 ft = 96 feet

Source

Ib Libner Nonnecke. 1958. Yield variability of sweet corn and canning peas as affected by plot size and shape. Thesis at Oregon State College. <http://hdl.handle.net/1957/23367>

References

I. L. Nonnecke, 1959. The precision of field experiments with vegetable crops as influenced by plot and block size and shape: I. Sweet corn. Canadian Journal of Plant Science, 39(4): 443-457. Tables 1-7. <http://doi.org/10.4141/cjps59-061>

Examples

```
# Corn 1
data(nonnecke.corn.uniformity)
dat <- nonnecke.corn.uniformity

if(require(desplot)){
  desplot(yield~col*row|loc, dat,
          flip=TRUE, tick=TRUE, aspect=96/180, # true aspect
          main="nonnecke.corn.uniformity")
}
```

}

`nonnecke.peas.uniformity`*Uniformity trial of canning peas*

Description

Uniformity trial of canning peas in southern Alberta, 1957.

Usage

```
data("nonnecke.peas.uniformity")
```

Format

A data frame with 540 observations on the following 5 variables.

block block factor

row row

col column

vines vines weight, pounds

peas shelled peas weight, pounds

Details

Width of basic plot was 10 feet, length was 5 feet, as limited by the viner. At each of two blocks/locations, planting consisted of 18 rows (only 15 rows were harvested) that were 10 feet wide and 90 feet long. Rows were separated by 7 foot bare ground to facilitate harvesting. Nonnecke 1960 shows a map of one block.

Plots were harvested with a five foot mower. Vines from each plot were weighed, then shelled. The two blocks/locations were side by side and combined by Nonnecke. The optimum plot size was found to be 5 feet long and 10 feet wide.

Field width: 15 rows * 10 ft/row + 14 gaps * 7 ft/gap = 248 feet

Field length: 18 plots * 5 ft/plot = 90 feet

Source

Ib Libner Nonnecke. 1958. Yield variability of sweet corn and canning peas as affected by plot size and shape. Thesis at Oregon State College. <http://hdl.handle.net/1957/23367>

References

I. L. Nonnecke, 1960. The precision of field experiments with vegetable crops as influenced by plot and block size and shape: II. Canning peas. *Canadian Journal of Plant Science*, 40(2): 396-404. <http://doi.org/10.4141/cjps60-053>

Examples

```

data(nonnecke.peas.uniformity)
dat <- nonnecke.peas.uniformity

if(require(desplot)){
  desplot(vines~col*row|block, data=dat,
          tick=TRUE, flip=TRUE, aspect=248/90, # true aspect
          main="nonnecke.peas.uniformity - vines")
  desplot(peas~col*row|block, data=dat,
          tick=TRUE, flip=TRUE, aspect=248/90, # true aspect
          main="nonnecke.peas.uniformity - peas")
}

## Not run:
require(lattice)
xyplot(peas~vines|block,dat,
       xlab="vine weight", ylab="shelled pea weight",
       main="nonnecke.peas.uniformity")

## End(Not run)

```

odland.soybean.uniformity

Uniformity trials of soy hay and soybeans

Description

Uniformity trials of soy hay and soybeans at Virginia Experiment Station, 1925-1926.

Format

Data frames with 3 variables.

row row

col column

yield yield: hay in tons, beans in bushels

Details

Grown at West Virginia Experiment Station in 1925 & 1926. Soybeans were planted in rows 30 inches apart.

Soy hay

In 1925 the crop was harvested for forage, 42 rows, each 200 feet long. Yields of 8-foot plats recorded.

Field width: 42 plots * 30 in / 12in/ft = 105 ft

Field length: 24 plots * 8 feet = 192 feet + border = total 200 feet.

Soy beans

In 1926 the crop was harvested for seed, 55 rows, each 232 feet long. Yields of 8-foot plats were recorded. In 1926, data for the last row on page 96 seems to be missing.

Field width: 55 plots * 30 in / 12in/ft = 137.5 feet

Field length: 28 plots * 8 feet = 224 feet + border = total 232 feet.

Odland and Garber provide no agronomic context for the yield variation.

Source

Odland, TE and Garber, RJ. 1928. Size of Plat and Number of Replications in Field Experiments with Soybeans. *Agronomy Journal*, 20, 93–108. <http://doi.org/10.2134/agronj1928.00021962002000020002x>

Examples

```
data(odland.soyhay.uniformity)
dat1 <- odland.soyhay.uniformity

if(require(desplot)){
  desplot(yield ~ col*row, dat1,
          flip=TRUE, aspect=200/105, # true aspect
          main="odland.soyhay.uniformity")
}

data(odland.soybean.uniformity)
dat2 <- odland.soybean.uniformity
if(require(desplot)){
  desplot(yield ~ col*row, dat2, flip=TRUE,
          aspect = 232/137,
          main="odland.soybean.uniformity")
}
```

 omer.sorghum

Yield of sorghum in 3 years, 2 locations

Description

Yield of sorghum in 3 years, 2 locations

Usage

```
data("omer.sorghum")
```

Format

A data frame with 432 observations on the following 4 variables.

env environment
 rep replication
 gen genotype factor
 yield yield, kg/ha

Details

Trials were conducted in Sudan, 3 years at 2 locations, 4 reps in RCBD at each location.

Source

Siraj Osman Omer, Abdel Wahab Hassan Abdalla, Mohammed Hamza Mohammed, Murari Singh (2015). Bayesian estimation of genotype-by-environment interaction in sorghum variety trials *Communications in Biometry and Crop Science*, 10 (2), 82-95.

Electronic data provided by Siraj Osman Omer.

References

None.

Examples

```
data(omer.sorghum)
dat <- omer.sorghum

# gge biplot
if(require(gge)){
  m0 <- gge(yield ~ gen* env, data=dat)
  biplot(m0, main="omer.sorghum")
}

## Not run:
# REML approach
require(lme4)
require(lucid)

# 1 loc, 2 years. Match Omer table 1.
m1 <- lmer(yield ~ 1 + env + (1|env:rep) + (1|gen) + (1|gen:env),
  data=subset(dat, is.element(env, c('E2','E4'))))
vc(m1)
##      grp      var1 var2  vcov  sdcor
## gen:env (Intercept) <NA> 17050  130.6
##      gen (Intercept) <NA>  2760   52.54
## env:rep (Intercept) <NA>   959.1  30.97
## Residual          <NA> <NA> 43090  207.6
```

```

# 1 loc, 3 years. Match Omer table 1.
m2 <- lmer(yield ~ 1 + env + (1|env:rep) + (1|gen) + (1|gen:env),
          data=subset(dat, is.element(env, c('E2','E4','E6'))))
vc(m2)
##      grp      var1 var2  vcov  sdcov
## gen:env (Intercept) <NA> 22210 149
##      gen (Intercept) <NA>  9288 96.37
## env:rep (Intercept) <NA>  1332 36.5
## Residual      <NA> <NA> 40270 200.7

# all 6 locs. Match Omer table 3, frequentist approach
m3 <- lmer(yield ~ 1 + env + (1|env:rep) + (1|gen) + (1|gen:env),
          data=dat)
vc(m3)
##      grp      var1 var2  vcov  sdcov
## gen:env (Intercept) <NA> 21340 146.1
## env:rep (Intercept) <NA>  1152 33.95
##      gen (Intercept) <NA>  1169 34.2
## Residual      <NA> <NA> 24660 157

## End(Not run)

```

onofri.winterwheat *Yield of winter wheat varieties across 7 years.*

Description

Yield of 8 durum winter wheat varieties across 7 years with 3 reps.

Usage

```
data("onofri.winterwheat")
```

Format

A data frame with 168 observations on the following 5 variables.

```

year  year, numeric
block block, 3 levels
plot  plot, numeric
gen   genotype, 7 levels
yield yield for each plot

```

Details

Yield of 8 durum winter wheat varieties across 7 years with 3 reps.

Source

Andrea Onofri, Egidio Ciricifolo (2007). Using R to Perform the AMMI Analysis on Agriculture Variety Trials. R News, Vol. 7, No. 1, pp. 14-19.

Downloaded electronic version from here Nov 2015: <http://www.casaonofri.it/Biometry/index.html>

Used with permission of Andrea Onofri.

References

F. Mendiburu. AMMI. <http://tarwi.lamolina.edu.pe/~fmendiburu/AMMI.htm>

A. Onofri. <http://accounts.unipg.it/~onofri/RTutorial/CaseStudies/WinterWheat.htm>

Examples

```
data(onofri.winterwheat)
dat <- onofri.winterwheat
dat <- transform(dat, year=factor(dat$year))

m1 <- aov(yield ~ year + block:year + gen + gen:year, dat)
anova(m1) # Matches Onofri figure 1

if(require(agricolae)){
  m2 <- AMMI(dat$year, dat$gen, dat$block, dat$yield)
  plot(m2)
  title("onofri.winterwheat - AMMI biplot")
}
```

ortiz.tomato

Multi-environment trial of tomato in Latin America, weight/yield and environmental covariates

Description

Multi-environment trial of tomato in Latin America, weight/yield and environmental covariates

Usage

```
data("ortiz.tomato.covs")
data("ortiz.tomato.yield")
```

Format

The ortiz.tomato.covs data frame has 18 observations on the following 18 variables.

env environment

Day degree days (base 10)

Dha days to harvest

Driv drivings (0/1)
 ExK extra potassium (kg / ha)
 ExN extra nitrogen (kg / ha)
 ExP extra phosphorous (kg / ha)
 Irr irrigation (0/1)
 K potassium (me/100 g)
 Lat latitude
 Long longitude
 MeT mean temperature (C)
 MnT min temperature (C)
 MxT max temperature (C)
 OM organic matter (percent)
 P phosphorous (ppm)
 pH soil pH
 Prec precipitation (mm)
 Tri trimming (0/1)

The `ortiz.tomato.yield` data frame has 270 observations on the following 4 variables.

env environment
 gen genotype
 yield marketable fruit yield t/ha
 weight fruit weight, g

Details

The environment locations are:

| | |
|-----|-----------------------------------|
| E04 | Estanzuela, Guatemala |
| E05 | Baja Verapaz, Guatemala |
| E06 | Cogutepeque, El Salvador |
| E07 | San Andres, El Salvador |
| E11 | Comayagua, Honduras |
| E14 | Valle de Sabaco, Nicaragua |
| E15 | San Antonio de Belen, Costa Rica |
| E20 | San Cristobal, Dominican Republic |
| E21 | Constanza, Dominican Republic |
| E27 | Palmira, Colombia |
| E40 | La Molina, Peru |
| E41 | Santiago, Chile |
| E42 | Chillan, Chile |
| E43 | Curacavi, Chile |
| E44 | Colina, Chile |
| E50 | Belem, Brazil |
| E51 | Caacupe, Paraguay |
| E53 | Centeno, Trinidad Tobago |

Source

Rodomiro Ortiz and Jose Crossa and Mateo Vargas and Juan Izquierdo, 2007. Studying the Effect of Environmental Variables On the Genotype x Environment Interaction of Tomato. *Euphytica*, 153, 119–134. <http://doi.org/10.1007/s10681-006-9248-7>

Used with permission of Rodomiro Ortiz.

Examples

```
data(ortiz.tomato.covs)
data(ortiz.tomato.yield)

if(require(pls) & require(reshape2))
# Double-centered yield matrix
Y <- acast(ortiz.tomato.yield, env ~ gen, value.var='yield')
Y <- sweep(Y, 1, rowMeans(Y, na.rm=TRUE))
Y <- sweep(Y, 2, colMeans(Y, na.rm=TRUE))

# Standardized covariates
X <- ortiz.tomato.covs
rownames(X) <- X$env
X <- X[,c("MxT", "MnT", "MeT", "Prec", "Day", "pH", "OM", "P", "K",
         "ExN", "ExP", "ExK", "Trim", "Driv", "Irr", "Dha")]
X <- scale(X)

# Now, PLS relating the two matrices.
# Note: pls deletes observations with missing values

m1 <- pls(Y~X)
# Inner-product relationships similar to Ortiz figure 1.
biplot(m1, which="x", var.axes=TRUE, main="ortiz.tomato - env*cov biplot")
#biplot(m1, which="y", var.axes=TRUE)
```

pacheco.soybean

Multi-environment trial of soybean in Brazil.

Description

Yields of 18 soybean genotypes at 11 environments in Brazil.

Format

gen genotype, 18 levels
 env environment, 11 levels
 yield yield, kg/ha

Details

In each environment was used an RCB design with 3 reps. The means of the reps are shown here.

Source

R M Pacheco, J B Duarte, R Vencovsky, J B Pinheiro, A B Oliveira, (2005). Use of supplementary genotypes in AMMI analysis. *Theor Appl Genet*, 110, 812-818. <http://doi.org/10.1007/s00122-004-1822-6>

Used with permission of Robert Pacheco.

Examples

```
data(pacheco.soybean)
dat <- pacheco.soybean

# AMMI biplot similar to Fig 2 of Pacheco et al.
if(require(agricolae)){
  m1 <- with(dat, AMMI(env, gen, REP=1, yield))
  bip <- m1$biplot[,1:3]
  # Fig 1 of Pacheco et al.
  with(bip, plot(yield, PC1, cex=0.0,
                text(yield,PC1,labels=row.names(bip), col="blue"),
                xlim=c(1000,3000),main="pacheco.soybean - AMMI biplot",frame=TRUE))
  with(bip[19:29,], points(yield, PC1, cex=0.0,
                          text(yield,PC1,labels=row.names(bip[19:29,]),
                              col="darkgreen"))))
}
```

parker.orange.uniformity

Uniformity trial of oranges

Description

Uniformity trial of oranges at Riverside, CA, 1921-1927.

Usage

```
data("parker.orange.uniformity")
```

Format

A data frame with 1364 observations on the following 4 variables.

year year

row row

col column

yield yield, pounds/tree for plot

Details

An orchard of naval oranges was planted in 1917 at the University of California Citrus Experiment Station at Riverside. The orchard was maintained under uniform conditions for 10 years.

Eight Washington Navel orange trees in a single row constituted a plot. The planting distance is 20 feet between trees within the row and 24 feet between rows. Every other row was a guard row, so row 2 and row 4 were observational units, while row 3 was a guard row. For example, from row 2 to row 4 is $2 \times 24 = 48$ feet. Another way to think of this is that each plot was 48 feet wide, but only the middle 24 feet was harvested. At each end of the plot was one guard tree. Including guard trees at the row ends, each row plot was 10 trees \times 20 feet = 200 feet long.

Field length (north-south) 27 plots \times 48 feet = 1296 feet.

Field width (west-east) 10 plots \times 200 feet = 2000 feet.

An investigation into the variability between plots included systematic soil surveys, soil moisture, soil nitrates, and inspection for differences in infestation of the citrus nematode. None of these factors was considered to be the primary cause of the variations in yield.

After the 7 years of uniformity trials, different treatments were applied to the plots.

Parker et al. state that soil heterogeneity is considerable and first-year yields are not predictive of future yields.

Table 25 has mean top volume per tree for each plot in 1926. Table 26 has mean area of trunk cross section.

Data were extracted from the source document by OCR and all values were checked by hand.

Source

E. R. Parker & L. D. Batchelor. (1932). Variation in the Yields of Fruit Trees in Relation to the Planning of Future Experiments. *Hilgardia*, 7(2), 81-161. Tables 3-9. <http://doi.org/10.3733/hilg.v07n02p081>

References

Batchelor, L. D. (Leon Dexter), b. 1884; Parker, E. R. (Edwin Robert), 1896-1952; McBride, Robert, d. 1927. (1928) Studies preliminary to the establishment of a series of fertilizer trials in a bearing citrus grove. Vol B451. Berkeley, Cal. : Agricultural Experiment Station <https://archive.org/details/studiesprelimina4>

Examples

```
data(parker.orange.uniformity)
dat <- parker.orange.uniformity

# Parker fig 2, field plan
if(require(desplot)){
  dat$year <- factor(dat$year)
  # 27 rows * 48 ft x 10 cols * 200 feet
  desplot(yield ~ col*row|year, data = dat,
          flip = TRUE, aspect = 27*48/(10*200), # true aspect
          main = "parker.orange.uniformity")
}
```

```

# CV across plots in each year. Similar to Parker table 11
cv <- function(x) {
  x <- na.omit(x)
  sd(x)/mean(x)
}
round(100*tapply(dat$yield, dat$year, cv),2)

# Correlation of plot yields across years. Similar to Parker table 15.
# Paker et al may have calculated correlation differently.
if(require(reshape2) & require(corrgram)){
  dat2 <- acast(dat, row+col ~ year, value.var = 'yield')
  round(cor(dat2, use = "pair"),3)
  corrgram(dat2, lower = panel.pts, upper = panel.conf,
           main="parker.orange.uniformity")
}

# Fertility index. Mean across years (ignoring 1921). Parker table 16
dat3 <- aggregate(yield ~ row+col, data = subset(dat, year !=1921 ),
                 FUN = mean, na.rm = TRUE)
round(acastr(dat3, row ~ col, value.var = 'yield'),0)

if(require(desplot)){
  desplot(yield ~ col*row, data = dat3,
         flip = TRUE, aspect = 27*48/(10*200), # true aspect
         main = "parker.orange.uniformity - mean across years")
}

```

patterson.switchback *A switchback trial in dairy cattle*

Description

A switchback trial in dairy cattle

Usage

```
data("patterson.switchback")
```

Format

A data frame with 36 observations on the following 4 variables.

y response, milk FCM

trt treatment factor, 4 levels

period period factor, 3 levels

cow cow factor, 12 levels

Details

There are three periods. Each cow is assigned to one treatment cycle like T1-T2-T1, where T1 is the treatment in period P1 and P3, and T2 is the treatment in period P2.

There are four treatments.

All $4 \times 3 = 12$ treatment cycles are represented.

Data were extracted from Lowry, page 70.

Source

Patterson, HD and Lucas, HL. 1962. Change-over designs. *Technical Bulletin 147*, North Carolina Agricultural Experimental Station.

References

Lowry, S.R. 1989. Statistical design and analysis of dairy nutrition experiments to improve detection of milk response differences. *Proceedings of the Conference on Applied Statistics in Agriculture*, 1989. <http://newprairiepress.org/agstatconference/1989/proceedings/7/>

Examples

```
data(patterson.switchback)
dat <- patterson.switchback

# Create groupings for first treatment, second treatment
datp1 <- subset(dat, period=="P1")
datp2 <- subset(dat, period=="P2")
dat$p1trt <- datp1$trt[match(dat$cow, datp1$cow)]
dat$p2trt <- datp2$trt[match(dat$cow, datp2$cow)]

if(require(latticeExtra)){
  useOuterStrips(xyplot(y ~ period|p1trt*p2trt, data=dat,
    group=cow, type=c('l','r'),
    auto.key=list(columns=5),
    main="patterson.switchback",
    xlab="First/Third period treatment",
    ylab="Second period treatment"))
}

# Create a numeric period variable
dat$per <- as.numeric(substring(dat$period,2))

# Need to use 'terms' to preserve the order of the model terms
m1 <- aov(terms(y ~ cow + per:cow + period + trt, keep.order=TRUE), data=dat)
anova(m1) # Match table 2 of Lowry
## Analysis of Variance Table
##           Df Sum Sq Mean Sq F value    Pr(>F)
## cow        11 3466.0  315.091  57.1773 2.258e-06 ***
## cow:per    12   953.5   79.455  14.4182 0.0004017 ***
## period     1    19.7   19.740   3.5821 0.0950382 .
```

```
## trt      3  58.3  19.418  3.5237  0.0685092 .
## Residuals 8  44.1   5.511
```

```
pearce.apple      Apple tree yields for 6 treatments with covariate
```

Description

Apple tree yields for 6 treatments with covariate of previous yield.

Format

A data frame with 24 observations on the following 4 variables.

```
block block factor, 4 levels
trt treatment factor, 6 levels
prev previous yield in boxes
yield yield per plot
```

Details

Treatment 'S' is the standard practice in English apple orchards of keeping the land clean in the summer.

The previous yield is the number of boxes of fruit, for the four seasons previous to the application of the treatments.

Source

S. C. Pearce (1953). *Field Experiments With Fruit Trees and Other Perennial Plants*. Commonwealth Bureau of Horticulture and Plantation Crops, Farnham Royal, Slough, England, App. IV.

References

James G. Booth, Walter T. Federer, Martin T. Wells and Russell D. Wolfinger (2009). A Multivariate Variance Components Model for Analysis of Covariance in Designed Experiments. *Statistical Science*, 24, 223-237.

Examples

```
data(pearce.apple)
dat <- pearce.apple

require(lattice)
xyplot(yield~prev|block, dat, main="pearce.apple", xlab="previous yield")

# Univariate fixed-effects model of Booth et al, using previous
```



```

# yield as a covariate.
m1 <- lm(yield ~ trt + block + prev, data=dat)
# Predict values, holding the covariate at its overall mean of 8.3
newdat <- expand.grid(trt=c('A','B','C','D','E','S'),
                     block=c('B1','B2','B3','B4'), prev=8.308333)
newdat$pred <- predict(m1, newdata=newdat)
# Average across blocks to get the adjusted mean, Booth et al. Table 1
tapply(newdat$pred, newdat$trt, mean)
#      A      B      C      D      E      S
# 280.4765 266.5666 274.0666 281.1370 300.9175 251.3357

# Same thing, but with blocks random
if(require(lme4)){
m2 <- lmer(yield ~ trt + (1|block) + prev, data=dat)
newdat$pred2 <- predict(m2, newdata=newdat)
tapply(newdat$pred2, newdat$trt, mean)
#      A      B      C      D      E      S
# 280.4041 266.5453 274.0453 281.3329 301.3432 250.8291
}

```

pearl.kernels

Counts of yellow/white and sweet/starchy maize kernels by 15 observers

Description

Counts of yellow/white and sweet/starchy kernels on each of 4 maize ears by 15 observers.

Format

A data frame with 59 observations on the following 6 variables.

ear ear, 8-11

obs observer, 1-15

ys number of yellow starchy kernels

yt yellow sweet

ws white starchy

wt white sweet

Details

An ear of white sweet corn was crossed with an ear of yellow starchy corn. The F1 kernels of the cross were grown and a sample of four ears was harvested. The F2 kernels of these ears were classified by each of 15 observers into white/yellow and sweet/starchy.

By Mendelian genetics, the kernels should occur in the ratio 9 yellow starch, 3 white starch, 3 yellow sweet, 1 white sweet.

The observers had the following positions:

- 1 Plant pathologist
- 2 Asst plant pathologist
- 3 Prof agronomy
- 4 Asst prof agronomy
- 5 Prof philosophy
- 6 Biologist
- 7 Biologist
- 8 Asst biologist
- 9 Computer
- 10 Farmer
- 11 Prof plant physiology
- 12 Instructor plant physiology
- 13 Asst plant physiology
- 14 Asst plant physiology
- 15 Prof biology

Source

Raymond Pearl, 1911. The Personal Equation In Breeding Experiments Involving Certain Characters of Maize, *Biol. Bull.*, 21, 339-366. <http://www.biobull.org/cgi/reprint/21/6/339.pdf>

Examples

```
data(pearl.kernels)
dat <- pearl.kernels

require(lattice)
xyplot(ys+yt+ws+wt~obs|ear, dat, type='l', as.table=TRUE,
       auto.key=list(columns=4),
       main="pearl.kernels", xlab="observer",ylab="kernels",
       layout=c(4,1), scales=list(x=list(rot=90)))

# Test hypothesis that distribution is 'Mendelian' 9:3:3:1
dat$pval <- apply(dat[, 3:6], 1, function(x)
                 chisq.test(x, p=c(9,3,3,1)/16)$p.val)
dotplot(pval~obs|ear, dat, layout=c(1,4), main="pearl.kernels",
        ylab="P-value for test of 9:3:3:1 distribution")
```

perry.springwheat

Multi-environment trial of wheat cultivars introduced 1860-1982.

Description

Yields of wheat cultivars introduced 1860-1982. Grown in 20 environments.

Usage

```
data("perry.springwheat")
```

Format

A data frame with 560 observations on the following 6 variables.

```
yield yield, kg/ha
gen genotype/cultivar factor, 28 levels
env environment factor, 20 levels
site site factor
year year, 1979-1982
yor year of release, 1860-1982
```

Details

Twenty-eight of the most significant wheat cultivars of the past century in Western Australia, were grown in 20 field trials over 4 years in the Central and Eastern wheat-belt of Australia.

At the Wongan Hills site there were separate early and late sown trials in 1979 and 1980. Later sowing dates generally have lower yields.

Note: Although not indicated by the original paper, it may be that the Merredin site in 1979 also had early/late sowing dates.

Source

MW Perry and MF D'Antuono. (1989). Yield improvement and associated characteristics of some Australian spring wheat cultivars introduced between 1860 and 1982. *Australian Journal of Agricultural Research*, 40(3), 457–472. <http://www.publish.csiro.au/nid/43/issue/1237.htm>

Used with permission of Mario D'Antuono and CSIRO Publishing.

Examples

```
data(perry.springwheat)
dat <- perry.springwheat

library(lattice)
xyplot(yield~yor|env, dat, type=c('p','r'), xlab="year of release",
       main="perry.springwheat")

# Show a table of sites*year
# library(latticeExtra)
# useOuterStrips(xyplot(yield~yor|site*factor(year), dat,
#                       type=c('p','r'))))

# Perry reports a rate of gain of 5.8 kg/ha/year. No model is given.
# We fit a model with separate intercept/slope for each env
```

```

m1 <- lm(yield ~ env + yor + env:yorc, data=dat)
# Average slope across environments
mean(c(coef(m1)[21], coef(m1)[21]+coef(m1)[22:40]))
## [1] 5.496781

# -----

## Not run:
# Now a mixed-effects model. Fixed overall int/slope. Random env int/slope.

# First, re-scale response so we don't have huge variances
dat$y <- dat$yield / 100

require(lme4)
# Use || for uncorrelated int/slope. Bad model. See below.
# m2 <- lmer(y ~ 1 + yor + (1+yor||env), data=dat)
## Warning messages:
## 1: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.55842 (tol = 0.002, component 1)
## 2: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model is nearly unidentifiable: very large eigenvalue
## - Rescale variables?;Model is nearly unidentifiable: large eigenvalue ratio
## - Rescale variables?

# Looks like lme4 is having trouble with variance of intercepts
# There is nothing special about 1800 years, so change the
# intercept -- 'correct' yor by subtracting 1800 and try again.
dat$yorc <- dat$yor - 1800
m3 <- lmer(y ~ 1 + yorc + (1+yorc||env), data=dat)

# Now lme4 succeeds. Rate of gain is 100*0.0549 = 5.49
fixef(m3)
## (Intercept)      yorc
## 5.87492444  0.05494464

# asreml3
require(asreml)
m3a <- asreml(y ~ 1 + yorc, data=dat, random = ~ env + env:yorc)

require(lucid)
vc(m3)
##      grp      var1 var2      vcov      sdcov
##      env (Intercept) <NA> 11.61      3.407
##      env.1      yorc <NA> 0.00063 0.02511
## Residual      <NA> <NA> 3.551      1.884

vc(m3a)
##      effect component std.error z.ratio      con
##      env!env.var 11.61      4.385      2.6 Positive
##      env:yorc!env.var 0.00063 0.000236      2.7 Positive
##      R!variance 3.551      0.2231      16 Positive

```

```
## End(Not run)
```

```
# -----
```

```
piepho.cocksfoot      Heading dates for cock's foot on 25 varieties in 7 years
```

Description

Heading dates for cock's foot on 25 varieties in 7 years

Usage

```
data("piepho.cocksfoot")
```

Format

A data frame with 111 observations on the following 3 variables.

gen genotype factor, 25 levels

year year, numeric

date heading date (days from April 1)

Details

These data are heading dates (days from April 1 to heading) of 25 cock's foot *Dactylis glomerata* varieties in trials at Hannover, Germany, repeated over seven years. Values are means over replications.

Source

Hans-Pieter Piepho. (1999). Fitting a Regression Model for Genotype-by-Environment Data on Heading Dates in Grasses by Methods for Nonlinear Mixed Models. *Biometrics*, 55, 1120-1128. <http://doi.org/10.1111/j.0006-341X.1999.01120.x>

Examples

```
data(piepho.cocksfoot)
dat <- piepho.cocksfoot
dat$env <- factor(dat$year)

require(lattice)
densityplot(~date|env, data=dat, main="piepho.cocksfoot - heading date") # Gaussian, not gamma distn

## Not run:
require(lme4)
# A simple mixed model gives results similar to Piepho, Table 2.
```

```

m1 <- lmer(date ~ -1 + gen + (1|env), data=dat)
summary(m1)
## Fixed effects:
##           Estimate Std. Error t value
## genG01    48.951      1.744   28.07
## genG02    42.270      1.755   24.09
## genG03    45.293      1.744   25.98
## genG04    45.467      1.744   26.08
## genG05    44.215      1.755   25.20

## End(Not run)

```

```

polson.safflower.uniformity
      Uniformity trial of safflower

```

Description

Uniformity trial of safflower at Farmington, Utah, 1962.

Usage

```
data("polson.safflower.uniformity")
```

Format

A data frame with 1716 observations on the following 3 variables.

```

row row
col column
yield yield (grams)

```

Details

A uniformity trial of safflower at the Utah State University field station in Farmington, Utah, in 1962. The field was approximately 0.5 acres in size, 110 x 189 feet. A four-row planter was used, 22 inches between rows. Four rows on either side and 12 feet on both ends were removed before harvesting.

Yield of threshed grain was recorded in grams.

Field width: (52 rows + 8 border rows) * 22 in = 110 ft

Field length: 33 sections * 5ft + 2 borders * 12 ft = 189 ft

Source

David Polson. 1964. Estimation of Optimum Size, Shape, and Replicate Number of Safflower Plots for Yield Trials. Table 6. <http://digitalcommons.usu.edu/etd/2979>

References

None.

Examples

```

data(polson.safflower.uniformity)
dat <- polson.safflower.uniformity

if(require("desplot")){
  desplot(yield ~ col*row, data=dat,
          flip=TRUE, aspect=189/110, # true aspect
          main="polson.safflower.uniformity")
}

if(require(agricolae) & require(reshape2)){
  dmat <- acast(dat, row~col, value.var="yield")
  # Similar to Polson fig 4.
  tab <- index.smith(dmat, col="red",
                    main="polson.safflower.uniformity - Smith Index",
                    xlab="Plot size in number of basic plots")

  # Polson p. 25 said CV decreased from 14.3 to 4.5
  # for increase from 1 unit to 90 units. Close match.
  tab <- data.frame(tab$uniformity)

  # Polson only uses log(Size) < 2 in his Fig 5, obtained slope -0.63
  coef(lm(log(Vx) ~ log(Size), subset(tab, Size <= 6))) # -0.70

  # Polson table 2 reported labor for
  # K1, number of plots, 133 hours 75%
  # K2, size of plot, 43.5 hours 24%
  # Optimum plot size
  # X = b K1 / ((1-b) K2)
  # Polson suggests optimum plot size 2.75 to 11 basic plots
}

```

 ratkowsky.onions

Onion yields for different densities at two locations

Description

Onion yields for different densities at two locations

Format

This data frame contains the following columns:

density planting density (plants per square meter)

yield yield (g / plant)
loc location, Purnong Landing or Virginia

Details

Spanish white onions.

Source

Ratkowsky, D. A. (1983). *Nonlinear Regression Modeling: A Unified Practical Approach*. New York: Marcel Dekker.

References

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003). *Semiparametric Regression*. Cambridge University Press. <http://stat.tamu.edu/~carroll/semiregbook/>

Examples

```
data(ratkowsky.onions)
dat <- ratkowsky.onions

# Model inverse yield as a quadratic. Could be better...
require(lattice)
dat <- transform(dat, iyield = 1/yield)
m1 <- lm(iyield ~ I(density^2)*loc, dat)
dat$pred <- predict(m1)

if(require(latticeExtra)) {
  foo <- xyplot(iyield ~ density, data=dat, group=loc, auto.key=TRUE,
               main="ratkowski.onions",ylab="Inverse yield")
  foo + xyplot(pred ~ density, data=dat, group=loc, type='l')
}
```

reid.grasses

Yields of four grasses for a wide range of nitrogen fertilizer

Description

Yields of four grasses for a wide range of nitrogen fertilizer, conducted over 3 years.

Usage

```
data("reid.grasses")
```


Format

A data frame with 210 observations on the following 5 variables.

nitro a numeric vector
 year a character vector
 gen a character vector
 drymatter a numeric vector
 protein a numeric vector

Details

Experiment at the Hannah Research Institute, Ayr.

Single plots were planted to 4 different kinds of grasses. Within each plot, 21 nitrogen treatments were randomized.

Reid modeled the dry matter yield with four-parameter logistic curves of the form $y = a - b \exp(-cx^d)$.

Source

D. Reid (1985). A comparison of the yield responses of four grasses to a wide range of nitrogen application rates. *J. Agric. Sci.*, 105, 381-387. Table 1 & 3. <https://doi.org/10.1017/S0021859600056434>

References

None

Examples

```
data(reid.grasses)
dat <- reid.grasses

if(require(latticeExtra)){
  foo <- xyplot(drymatter + protein ~ nitro|year, dat, group=gen,
               as.table=TRUE, type=c('p','l'),
               main="reid.grasses",
               scales=list(y=list(relation="free")))
  combineLimits(foo)
}

## Not run:

library(nlme)
dat2 <- dat
dat2$indiv <- paste(dat$year, dat$gen) # individual year+genotype curves

# use all data to get initial values
inits <- getInitial(drymatter ~ SSfpl(nitro, A, B, xmid, scal), data = dat2)
inits
##          A          B          xmid          scal
```

```

## -4.167902 12.139796 68.764796 128.313106
xvals <- 0:800
y1 <- with(as.list(inits), SSfpl(xvals, A, B, xmid, scal))
plot(drymatter ~ nitro, dat2)
lines(xvals,y1)

# must have groupedData object to use augPred
dat2 <- groupedData(drymatter ~ nitro|indiv, data=dat2)
plot(dat2)

# without 'random', all effects are included in 'random'
m1 <- nlme(drymatter ~ SSfpl(nitro, A, B, xmid,scale),
          data= dat3,
          fixed= A + B + xmid + scale ~ 1,
          # random = B ~ 1|indiv, # to make only B random
          random = A + B + xmid + scale ~ 1|indiv,
          start=inits)

fixef(m1)
summary(m1)
plot(augPred(m1, level=0:1)) # only works with groupedData object

## End(Not run)

```

ridout.appleshoots *Root counts for propagated columnar apple shoots.*

Description

Root counts for propagated columnar apple shoots.

Usage

```
data("ridout.appleshoots")
```

Format

A data frame with 270 observations on the following 4 variables.

roots number of roots per shoot

trtn number of shoots per treatment combination

photo photoperiod, 8 or 16

bap BAP concentration, numeric

Details

There were 270 micropropagated shoots from the columnar apple cultivar Trajan. During the rooting period, shoot tips of length 1.0-1.5 cm were cultured on media with different concentrations of the cytokinin BAP in two growth chambers with 8 or 16 hour photoperiod.

The response variable is the number of roots after 4 weeks at 22 degrees C.

Almost all of the shoots in the 8 hour photoperiod rooted. Under the 16 hour photoperiod only about half rooted.

High BAP concentrations often inhibit root formation of apples, but perhaps not for columnar varieties.

Source

Ridout, M. S., Hinde, J. P., and Demetrio, C. G. B. (1998). Models for Count Data with Many Zeros. *Proceedings of the 19th International Biometric Conference*, 179-192.

Used with permission of Martin Ridout.

References

SAS. Fitting Zero-Inflated Count Data Models by Using PROC GENMOD. support.sas.com/rnd/app/examples/stat/GENMOD

Examples

```
data(ridout.appleshoots)
dat <- ridout.appleshoots

# Change photo and bap to factors
dat <- transform(dat, photo=factor(photo), bap=factor(bap))

library(lattice)
# histogram(~roots, dat, breaks=0:18-0.5)

# For photo=8, Poisson distribution looks reasonable.
# For photo=16, half of the shoots had no roots
# Also, photo=8 has very roughly 1/45 as many zeros as photo=8,
# so we anticipate prob(zero) is about 1/45=0.22 for photo=8.
histogram(~roots|photo, dat, breaks=0:18-0.5, main="ridout.appleshoots")

foo.obs <- histogram(~roots|photo*bap, dat, breaks=0:18-0.5, type="density",
                    xlab="Number of roots for photoperiod 8, 16",
                    ylab="Density for BAP levels",
                    main="ridout.appleshoots")
if(require(latticeExtra)) useOuterStrips(foo.obs) else print(foo.obs)

# Ordinary (non-ZIP) Poisson GLM
m1 <- glm(roots ~ bap + photo + bap:photo, data=dat,
          family="poisson")
summary(m1) # Appears to have overdispersion
```

```

# ----- Fit a Zero-Inflated Poisson model -----

if(require(pscl)){

# Use SAS contrasts to match SAS output
oo <- options(contrasts=c('contr.SAS','contr.poly'))

# There are unequal counts for each trt combination, which obviously affects
# the distribution of counts, so use log(trtn) as an offset.
dat$ltrtn <- log(dat$trtn)

# Ordinary Poisson GLM: 1 + bap*photo.
# Zero inflated probability depends only on photoperiod: 1 + photo

m2 <- zeroinfl(roots ~ 1 + bap*photo | 1 + photo, data=dat,
              dist="poisson", offset=ltrtn)
logLik(m2)      # -622.2283 matches SAS Output 1
-2 * logLik(m2) # 1244.457 Matches Ridout Table 2, ZIP, H*P, P
summary(m2)     # Coefficients match SAS Output 3.

exp(coef(m2, "zero")) # Photo=8 has .015 times as many zeros as photo=16

# Get predicted _probabilities_

# Prediction data
newdat <- expand.grid(photo=c(8,16), bap=c(2.2, 4.4, 8.8, 17.6))
newdat <- aggregate(trtn~bap+photo, dat, FUN=mean)
newdat$ltrtn <- log(newdat$trtn)

# The predicted (Poisson + Zero) probabilities
d2 <- cbind(newdat[,c('bap','photo')], predict(m2, newdata=newdat, type="prob"))
if(require(reshape2)){
d2 <- melt(d2, id.var = c('bap','photo')) # wide to tall
d2$xpos <- as.numeric(as.character(d2$variable))
foo.poi <- xyplot(value~xpos|photo*bap, d2, col="black", pch=20, cex=1.5)

# Plot data and model
foo.obs <- update(foo.obs, main="ridout.appleshoots: observed (bars) & predicted (dots)")
useOuterStrips(foo.obs + foo.poi)
}

# Restore contrasts
options(oo)

}

```

Description

Yield of brussels sprouts, RCB with 9 fertilizer treatments

Format

A data frame with 48 observations on the following 5 variables.

row row

col column

yield yield of saleable sprouts, pounds

trt treatment, 9 levels

block block, 4 levels

Details

The block numbers are arbitrary, and may not match the original source.

Plots were 10 yards x 14 yards. Plot orientation is not clear.

Source

Rothamsted Experimental Station Report 1934-36. Brussels sprouts: effect of sulphate of ammonia, poultry manure, soot and rape dust, pp. 191-192. Harpenden: Lawes Agricultural Trust.

References

McCullagh, P. and Clifford, D., (2006). Evidence for conformal invariance of crop yields, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 462, 2119–2143.

Examples

```
data(rothamsted.brussels)
dat <- rothamsted.brussels

require(lattice)
bwplot(yield~trt, dat, main="rothamsted.brussels")

if(require(desplot)){
  desplot(yield~col*row, data=dat,
          num=trt, out1=block, cex=1, # aspect unknown
          main="rothamsted.brussels")
}
```

| | |
|-----------------|-------------------------------------|
| ryder.groundnut | <i>Wet and dry groundnut yields</i> |
|-----------------|-------------------------------------|

Description

Wet and dry groundnut yields with a possible data error.

Format

A data frame with 24 observations on the following 6 variables.

block block

row row

col column

gen genotype factor

wet wet yield, kg/plot

dry dry yield, kg/plot

Details

Ryder (1981) uses this data to discuss the importance of looking at the field plan for an experiment. Based on analysis of the residuals, he suggests that varieties A and B in block 3 may have had their data swapped.

Source

K. Ryder (1981). Field plans: why the biometrician finds them useful, *Experimental Agriculture*, 17, 243–256.

<http://doi.org/10.1017/S0014479700011601>

Examples

```
data(ryder.groundnut)
dat <- ryder.groundnut

# RCB model
m1 <- lm(dry~block+gen,dat)
dat$res1 <- resid(m1)

# Table 3 of Ryder. Scale up from kg/plot to kg/ha
round(dat$res1 * 596.6,0)

# Visually. Note largest positive/negative residuals are adjacent
if(require(desplot)){
  desplot(res1 ~ col + row, dat,
          text=gen, # aspect unknown
```

```

        main="ryder.groundnut - residuals")
    }

## Not run:
require(desplot)
# Swap the dry yields for two plots and re-analyze
dat[dat$block=="B3" & dat$gen=="A", "dry"] <- 2.8
dat[dat$block=="B3" & dat$gen=="B", "dry"] <- 1.4
m2 <- lm(dry~block+gen, dat)
dat$res2 <- resid(m2)
desplot(res2 ~ col+row, dat,
        # aspect unknown
        text=gen, main="ryder.groundnut")

## End(Not run)

```

salmon.bunt

Fungus infection in varieties of wheat

Description

Fungus infection in varieties of wheat

Format

A data frame with 400 observations on the following 4 variables.

bunt bunt factor, 20 levels
pct percent infected
rep rep factor, 2 levels
gen genotype factor, 10 levels

Details

Note: Salmon (1938) gives results for all 69 types of bunt, not just the 20 shown in the paper.

H. A. Rodenhiser and C. S. Holton (1937) say that races from two different species of bunt were used, *Tilletia tritici* and *T. levis*.

This data gives the results with 20 types of bunt (fungus) for winter wheat varieties at Kearneysville, W. Va., in 1935. Altogether there were 69 types of bunt included in the experiment, of which the 20 in this data are representative. Each type of wheat was grown in a short row (5 to 8 feet), the seed of which had been inoculated with the spores of bunt. The entire seeding was then repeated in the same order.

Infection was recorded as a percentage of the total number of heads counted at or near harvest. The number counted was seldom less than 200 and sometimes more than 400 per row.

Source

S.C. Salmon, 1938. Generalized standard errors for evaluating bunt experiments with wheat. *Agronomy Journal*, 30, 647–663. Table 1. <http://doi.org/10.2134/agronj1938.00021962003000080003x>

References

Salmon says the data came from:

H. A. Rodenhiser and C. S. Holton (1937). Physiologic races of *Tilletia tritici* and *T. levis*. *Journal of Agricultural Research*, 55, 483-496. naldc.nal.usda.gov/download/IND43969050/PDF

Examples

```
data(salmon.bunt)
dat <- salmon.bunt

d2 <- aggregate(pct~bunt+gen, dat, FUN=mean) # average reps
d2$gen <- reorder(d2$gen, d2$pct)
d2$bunt <- reorder(d2$bunt, d2$pct)
# Some wheat varieties (Hohenheimer) are resistant to all bunts, and some (Hybrid128)
# are susceptible to all bunts. Note the groups of bunt races that are similar,
# such as the first 4 rows of this plot. Also note the strong wheat*bunt interaction.
require(lattice)
redblue <- colorRampPalette(c("firebrick", "lightgray", "#375997"))
levelplot(pct~gen+bunt,d2, col.regions=redblue,
          main="salmon.bunt percent of heads infected",
          xlab="Wheat variety", ylab="bunt line")

# We don't have individual counts, so use beta regression
if(require(betareg) & require(latticeExtra)){
  dat$y <- dat$pct/100 + .001 # Beta regression does not allow 0
  dat$gen <- reorder(dat$gen, dat$pct) # For a prettier dot plot

  m1 <- betareg(y ~ gen + bunt + gen:bunt, data=dat)

  # Construct 95 percent confidence intervals
  p1 <- cbind(dat,
             lo = predict(m1, type='quantile', at=.025),
             est = predict(m1, type='quantile', at=.5),
             up = predict(m1, type='quantile', at=.975))
  p1 <- subset(p1, rep=="R1")

  # Plot the model intervals over the original data
  # require(latticeExtra)
  dotplot(bunt~y|gen, data=dat, pch='x', col='red',
          main="Observed data and 95 pct intervals for bunt infection") +
  segplot(bunt~lo+up|gen, data=p1, centers=est, draw.bands=FALSE)
}

# To evaluate wheat, we probably want to include bunt as a random effect...
```

sawyer.multi.uniformity

Uniformity trials of wheat, swedes, oats, 3 years on the same land.

Description

Uniformity trials of wheat, swedes, oats at Rothamsted, England, 1925-1927.

Usage

```
data("sawyer.multi.uniformity")
```

Format

A data frame with 48 observations on the following 7 variables.

year year

crop crop

row row

col column

grain wheat/oats grain weight, pounds

straw wheat/oats straw weight, pounds

leafwt swedes leaf weight, pounds

rootwt swedes root weight, pounds

rootct swedes root count

Details

An experiment conducted at Rothamsted, England, in 1925-1927, in Sawyers Field.

Row 6, column 1 was not planted in any year.

Wheat was harvested in 1925. Row 1, column 1 had partially missing data for the wheat values in 1925 and was not used in the Rothamsted summary statistics on page 155.

Swedes were harvested in 1926.

Oats were harvested in 1927.

Note the summaries statistics at the bottom of the page in each report are calibrated to ACRES.

Field width: 8 plots * 1 chain each Field length: 6 plots * 1 chain each

The field is 8 plots wide, 6 plots long. The plots are drawn in the source documents as squares .098 acres each.

Eden & Maskell (page 165) say the field was clover, and ploughed in the autumn of 1924. The field was laid out uniformly in lands of one chain width and each plot width made to coincide with the land width from ridge to ridge. The length of each plot was also one chain and from the point of view of yield data the trial comprised 47 plots in 8x6 except that the run of the hedge only allowed a rank of five plots at one of the ends.

Source

Rothamsted Experimental Station, Report 1925-26. Lawes Agricultural Trust, p. 154-155. <http://www.era.rothamsted.ac.uk/e>

Rothamsted Experimental Station, Report 1927-1928. Lawes Agricultural Trust, p. 153. <http://www.era.rothamsted.ac.uk/era/28-131-175>

References

Eden, T. and E. J. Maskell. (1928). The influence of soil heterogeneity on the growth and yield of successive crops. *Jour of Agricultural Science*, 18, 163-185. <https://archive.org/stream/in.ernet.dli.2015.25895/2015.25895.J. Of-Agricultural-Science-Vol-xviii-1928#page/n175>

McCullagh, P. and Clifford, D., (2006). Evidence for conformal invariance of crop yields, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 462, 2119–2143. <http://doi.org/10.1098/rspa.2006.1667>

Winifred A. Mackenzie. (1926) Note on a remarkable correlation between grain and straw, obtained at Rothamsted. *Journal of Agricultural Science*, 16, 275-279. <https://doi.org/10.1017/S0021859600018256>

Examples

```
data("sawyer.multi.uniformity")
dat <- sawyer.multi.uniformity

if(require(desplot)){
  # The field plan shows square plots
  desplot(grain~col*row, subset(dat, year==1925),
          main="sawyer.multi.uniformity - 1925 wheat grain yield",
          aspect=(6)/(8)) # true aspect
}
if(require(desplot)){
  desplot(rootwt~col*row, subset(dat, year==1926),
          main="sawyer.multi.uniformity - 1926 root weight of swedes",
          aspect=(6)/(8))
}
if(require(desplot)){
  desplot(grain~col*row, subset(dat, year==1927),
          main="sawyer.multi.uniformity - 1927 oats grain yield",
          aspect=(6)/(8))
}

## Not run:
# This plot shows the "outlier" in the wheat data reported by Mackenzie.
require(lattice)
xyplot(grain ~ straw, data=subset(dat, year==1925))

round(cor(dat[,7:9], use="pair"),2) # Matches McCullagh p 2121
##      leafwt rootwt rootct
## leafwt  1.00  0.66  0.47
## rootwt  0.66  1.00  0.43
## rootct  0.47  0.43  1.00
```

```
## pairs(dat[,7:9],  
##       main="sawyer.multi.uniformity")  
  
## End(Not run)
```

senshu.rice

Multi-environment trial of rice, with solar radiation and temperature

Description

Response of rice to solar radiation and temperature

Format

A data frame with 40 observations on the following 7 variables.

country country

loc location

year year of planting, last two digits

month month of planting

rad solar radiation

mint minimum temperature

yield yield t/ha

Details

Minimum temperature is the average across 30 days post flowering.

Opinion: Fitting a quadratic model to this data makes no sense.

Source

Seshu, D. V. and Cady, F. B. 1984. Response of rice to solar radiation and temperature estimated from international yield trials. *Crop Science*, 24, 649-654. <http://doi.org/10.2135/cropsci1984.0011183X002400040006x>

References

Walter W. Piegorisch, A. John Bailer. (2005) *Analyzing Environmental Data*, Wiley.

Examples

```

data(senshu.rice)
dat <- senshu.rice

# Model 1 of Senshu & Cady
m1 <- lm(yield ~ 1 + rad + mint + I(mint^2), dat)
coef(m1)

# Use Fieller to calculate conf int around optimum minimum temp
# See: Piegorsch & Bailer, p. 31.
# Calculation derived from vegan::fieller.MOSTest
m2 <- lm(yield ~ 1 + mint + I(mint^2), dat)
b1 <- coef(m2)[2]
b2 <- coef(m2)[3]
vc <- vcov(m2)
sig11 <- vc[2,2]
sig12 <- vc[2,3]
sig22 <- vc[3,3]
u <- -b1/2/b2
tval <- qt(1-.05/2, nrow(dat)-3)
gam <- tval^2 * sig22 / b2^2
x <- u + gam * sig12 / (2 * sig22)
f <- tval / (-2*b2)
sq <- sqrt(sig11 + 4*u*sig12 + 4*u^2*sig22 - gam * (sig11 - sig12^2 / sig22) )
ci <- (x + c(1,-1)*f*sq) / (1-gam)
plot(yield ~ mint, dat, xlim=c(17, 32),
     main="senshu.rice: Quadratic fit and Fieller confidence interval",
     xlab="Minimum temperature", ylab="Yield")
lines(17:32, predict(m2, new=data.frame(mint=17:32)))
abline(v=ci, col="blue")

```

shafii.rapeseed

*Multi-environment trial of rapeseed***Description**

Rapeseed yield multi-environment trial, 3 years

Format

A data frame with 648 observations on the following 5 variables.

year year, numeric: 87, 88, 89

loc location, 14 levels

rep rep, 3 levels

gen genotype, 6 levels

yield yield, kg/ha

Details

SAS codes for the analysis can be found at <http://www.uiweb.uidaho.edu/ag/statprog/ammi/>

The data are from the U.S. National Winter Rapeseed trials conducted in 1986, 1987, and 1988. Trial locations included Georgia (GGA, TGA), Idaho (ID), Kansas (KS), Mississippi (MS), Montana (MT), New York (NY), North Carolina (NC), Oregon (OR), South Carolina (SC), Tennessee (TN), Texas (TX), Virginia (VA), and Washington (WA).

Source

Bahman Shafii and William J Price, 1998. Analysis of Genotype-by-Environment Interaction Using the Additive Main Effects and Multiplicative Interaction Model and Stability Estimates, *Journal of Agricultural Biological Environmental Statistics*, 3, 335–345. <http://doi.org/10.2307/1400587>

Electronic version from: <http://www.uiweb.uidaho.edu/ag/statprog/ammi/yld.data>

Used with permission of Benjamin Price.

References

None.

Examples

```
data(shafii.rapeseed)
dat <- shafii.rapeseed

dat$gen <- with(dat, reorder(gen, yield, mean))
dat$loc <- with(dat, reorder(loc, yield, mean))
dat$yield <- dat$yield/1000

dat <- transform(dat, rep=factor(rep), year=as.factor(as.character(year)))
dat$locyr = paste(dat$loc, dat$year, sep="")

# The 'means' of reps
datm <- aggregate(yield~gen+year+loc+locyr, data=dat, FUN=mean)
datm <- datm[order(datm$gen),]
datm$gen <- as.character(datm$gen)
datm$gen <- factor(datm$gen,
                  levels=c("Bienvenu", "Bridger", "Cascade",
                           "Dwarf", "Glacier", "Jet"))
dat$locyr <- reorder(dat$locyr, dat$yield, mean)

require(lattice)
# This picture tells most of the story
# Now change symbols
op <- tpg <- trellis.par.get()
tpg$superpose.symbol$pch <- c('7', '8', '9')
trellis.par.set(tpg)
dotplot(loc~yield|gen, group=year, data=dat,
        auto.key=list(columns=3),
        main="shafii.rapeseed", ylab="Location")
```

```

#dotplot(loc~yield|gen,group=year,data=datm,auto.key=TRUE,
#      main="shafii.rapeseed")

# AMMI biplot. Remove gen and locyr effects.
m1.lm <- lm(yield ~ gen + locyr, data=datm)
datm$res <- resid(m1.lm)
# Convert to a matrix
if(require(reshape2)){
dm <- melt(datm, measure.var='res', id.var=c('gen', 'locyr'))
dmat <- acast(dm, gen~locyr)
# AMMI biplot. Figure 1 of Shafii (1998)
biplot(prcomp(dmat), main="shafii.rapeseed - AMMI biplot")
}

trellis.par.set(op) # Unset graphics changes

```

silva.cotton

Number of cotton bolls for different levels of defoliation.

Description

Number of cotton bolls, nodes, plant height, and plant weight for different levels of defoliation.

Usage

```
data("silva.cotton")
```

Format

A data frame with 125 observations on the following 4 variables.

stage growth stage

defoliation level of defoliation, 0, 25, 50, 75, 100

plant plant number

rep replicate

reproductive number of reproductive structures

bolls number of bolls

height plant height

nodes number of nodes

weight weight of bolls

Details

Data come from a greenhouse experiment with cotton plants. Completely randomized design with 5 replicates, 2 plants per pot.

Artificial defoliation was used at levels 0, 25, 50, 75, 100 percent.

Data was collected per plant at five growth stages: vegetative, flower-bud, blossom, fig and cotton boll.

The primary response variable is the number of bolls. The data are counts, underdispersed, correlated.

Zeviana et al. used this data to compared Poisson, Gamma-count, and quasi-Poisson GLMs.

Bonat & Zeviani used this data to fit multivariate correlated generalized linear model.

Source

Silva, Anderson Miguel da; Degrande, Paulo Eduardo; Suekane, Renato; Fernandes, Marcos Gino; & Zeviani, Walmes Marques. (2012). Impacto de diferentes niveis de desfolha artificial nos estadios fenologicos do algodoeiro. *Revista de Ciencias Agrarias*, 35(1), 163-172. http://www.scielo.mec.pt/scielo.php?script=sci_arttext&pid=SciArtText&lng=pt&lng=pt

Used with permission of Walmes Zeviani.

Electronic version from: http://www.leg.ufpr.br/~walmes/data/desfolha_algodao.txt

References

Zeviani, W. M., Ribeiro, P. J., Bonat, W. H., Shimakura, S. E., Muniz, J. A. (2014). The Gamma-count distribution in the analysis of experimental underdispersed data. *Journal of Applied Statistics*, 41(12), 1-11. <http://doi.org/10.1080/02664763.2014.922168> Online supplement: <http://leg.ufpr.br/doku.php/publications:paper:jas2014>

Regression Models for Count Data. <http://cursos.leg.ufpr.br/rmcd/applications.html#cotton-bolls>

Wagner Hugo Bonat & Walmes Marques Zeviani (2017). Multivariate Covariance Generalized Linear Models for the Analysis of Experimental Data. Short-course at: 62nd RBras and 17th SEAGRO meeting/ <https://github.com/leg-ufpr/mcglm4aed>

Examples

```
data(silva.cotton)
dat <- silva.cotton
dat$stage <- ordered(dat$stage,
                    levels=c("vegetative", "flowerbud", "blossom", "boll", "bollopen"))
# make stage a numeric factors
dat <- transform(dat,
                stage = factor(stage, levels = unique(stage),
                              labels = 1:nlevels(stage)))

# sum data across plants, 1 pot = 2 plants
dat <- aggregate(cbind(weight,height,bolls,nodes) ~
                stage+defoliation+rep, data=dat, FUN=sum)
```

```

# all traits, plant-level data
if(require(latticeExtra)){
  foo <- xyplot(weight + height + bolls + nodes ~ defoliation | stage,
               data = dat, outer=TRUE,
               xlab="Defoliation percent", ylab="", main="silva.cotton",
               as.table = TRUE, jitter.x = TRUE, type = c("p", "smooth"),
               scales = list(y = "free"))
  combineLimits(useOuterStrips(foo))
}

## Not run:

# poisson glm with quadratic effect for defoliation
m0 <- glm(bolls ~ 1, data=dat, family=poisson)
m1 <- glm(bolls ~ defoliation+I(defoliation^2), data=dat, family=poisson)
m2 <- glm(bolls ~ stage:defoliation+I(defoliation^2), data=dat, family=poisson)
m3 <- glm(bolls ~ stage:(defoliation+I(defoliation^2)), data=dat, family=poisson)
par(mfrow=c(2,2)); plot(m3); layout(1)
anova(m0, m1, m2, m3, test="Chisq")

# predicted values
preddat <- expand.grid(stage=levels(dat$stage),
                      defoliation=seq(0,100,length=20))
preddat$pred <- predict(m3, newdata=preddat, type="response")

# Zeviani figure 3
require(latticeExtra)
xyplot(bolls ~ jitter(defoliation)|stage, dat,
       as.table=TRUE,
       main="silva.cotton - observed and model predictions",
       xlab="Defoliation percent",
       ylab="Number of bolls") +
xyplot(pred ~ defoliation|stage, data=preddat,
       as.table=TRUE,
       type='smooth', col="black", lwd=2)

## End(Not run)

## Not run:
# ----- mcglm -----
dat <- transform(dat, deffac=factor(defoliation))

library(car)

vars <- c("weight", "height", "bolls", "nodes")
splom(~dat[vars], data=dat,
      groups = stage,
      auto.key = list(title = "Growth stage",
                      cex.title = 1,
                      columns = 3),
      par.settings = list(superpose.symbol = list(pch = 4)),
      as.matrix = TRUE)

```



```

splom(~dat[vars], data=dat,
      groups = defoliation,
      auto.key = list(title = "Artificial defoliation",
                     cex.title = 1,
                     columns = 3),
      as.matrix = TRUE)

# multivariate linear model.
m1 <- lm(cbind(weight, height, bolls, nodes) ~ stage * deffac,
        data = dat)
anova(m1)

summary.aov(m1)

r0 <- residuals(m1)

# Checking the models assumptions on the residuals.
car::scatterplotMatrix(r0,
                      gap = 0, smooth = FALSE, reg.line = FALSE, ellipse = TRUE,
                      diagonal = "qqplot")

## End(Not run)

```

sinclair.clover

Clover yields in a factorial fertilizer experiment

Description

Clover yields in a factorial fertilizer experiment

Usage

```
data("sinclair.clover")
```

Format

A data frame with 25 observations on the following 3 variables.

yield yield t/ha

P phosphorous fertilizer kg/ha

S sulfur fertilizer kg/ha

Details

A phosphorous by sulfur factorial experiment at Dipton in Southland, New Zealand. There were 3 reps. Plots were harvested repeatedly from Dec 1992 to Mar 1994. Yields reported are the total dry matter across all cuttings.

Source

Sinclair AG, Risk WH, Smith LC, Morrison JD & Dodds KG (1994) Sulphur and phosphorus in balanced pasture nutrition. *Proc N Z Grass Assoc*, 56, 13-16.

References

Dodds, KG and Sinclair, AG and Morrison, JD. (1995). A bivariate response surface for growth data. *Fertilizer research*, 45, 117-122. <http://doi.org/10.1007/BF00790661>

Examples

```
data(sinclair.clover)
dat <- sinclair.clover

require(lattice)
xyplot(yield~P|factor(S), dat, layout=c(5,1),
       main="sinclair.clover - Yield by sulfur levels",
       xlab="Phosphorous")

# Dodds fits a two-dimensional Mitscherlich-like model:
# z = a*(1+b*{(s+t*x)/(x+1)}^y) * (1+d*{(th+r*y)/(y+1)}^x)

# First, re-scale the problem to a more stable part of the parameter space
dat <- transform(dat, x=P/10, y=S/10)

# Response value for (x=0, y=maximal), (x=maximal, y=0), (x=max, y=max)
z0m <- 5
zm0 <- 5
zmm <- 10.5

# The parameters are somewhat sensitive to starting values.
# I had to try a couple different initial values to match the paper by Dodds
m1 <- nls(yield ~ alpha*(1 + beta*{(sig+tau*x)/(x+1)}^y) * (1 + del*{(th+rho*y)/(y+1)}^x),
         data=dat, # trace=TRUE,
         start=list(alpha=zmm, beta=(zm0/zmm)-1, del=(z0m/zmm)-1,
                   sig=.51, tau=.6, th=.5, rho=.7))

summary(m1) # Match Dodds Table 2
## Parameters:
##      Estimate Std. Error t value Pr(>|t|)
## alpha 11.15148    0.66484  16.773 1.96e-12 ***
## beta  -0.61223    0.03759 -16.286 3.23e-12 ***
## del   -0.48781    0.04046 -12.057 4.68e-10 ***
## sig    0.26783    0.16985   1.577 0.13224
## tau    0.68030    0.06333  10.741 2.94e-09 ***
## th     0.59656    0.16716   3.569 0.00219 **
## rho    0.83273    0.06204  13.421 8.16e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 0.5298 on 18 degrees of freedom

## Not run:
```

```

pred <- expand.grid(x=0:17, y=0:9)
pred$z <- predict(m1, pred)

# 3D plot of data with fitted surface. Matches Dodds figure 2.
require(rgl)
bg3d(color = "white")
clear3d()
spheres3d(dat$x, dat$y, dat$yield,
           radius=.2, col = rep("navy", nrow(dat)))
surface3d(seq(0, 17, by = 1), seq(0, 9, by = 1), pred$z,
           alpha=0.9, col="wheat",
           front="fill", back="fill")
axes3d()
title3d("sinclair.clover - yield", "", xlab="Phosphorous/10",
        ylab="Sulfur/10", zlab="", line=3, cex=1.5)
view3d(userMatrix=matrix(c(.7,.2,-.7,0, -.7,.2,-.6,0, 0,.9,.3,0, 0,0,0,1),ncol=4))
# snapshot3d(file, "png")
rgl.close()

## End(Not run)

```

smith.beans.uniformity

Uniformity trials of beans, 2 species in 2 years

Description

Uniformity trials of beans at California, 1954-1955, 2 species in 2 years

Usage

```
data("smith.beans.uniformity")
```

Format

A data frame with 912 observations on the following 4 variables.

expt experiment

row row

col column

yield yield, kg

Details

Trials were conducted in California. In 1955 plots were twice as wide and twice as long as in 1954. Red Kidney is a bush variety bean, Standard Pink is a viny variety.

Smith randomly assigned A,B,C,D to plots and used these as 'varieties' for calculating ANOVA tables. Plots were combined side-by-side and end-to-end to make larger plots. Decreasing LSDs were observed for increases in plot sizes. LSDs were seldom above 200, which was considered to be a noticeable difference for the farmers.

There are four datasets:

1954

Experiment 1: Red Kidney. Experiment 2: Standard Pink

Field width: 18 plots * 30 inches = 45 ft

Field height: 12 plots * 15 ft = 180 ft

1955

Experiment 3: Red Kidney. Experiment 4: Standard Pink

Field width: 16 plots * 2 rows * 30 in = 80 ft

Field height: 15 plots * 30 ft = 450 ft

Source

Francis L. Smith, 1958. Effects of plot size, plot shape, and number of replications on the efficacy of bean yield trials. *Hilgardia*, 28, 43-63. <http://doi.org/10.3733/hilg.v28n02p043>

References

None.

Examples

```
data(smith.beans.uniformity)

dat1 <- subset(smith.beans.uniformity, expt=="E1")
dat2 <- subset(smith.beans.uniformity, expt=="E2")
dat3 <- subset(smith.beans.uniformity, expt=="E3")
dat4 <- subset(smith.beans.uniformity, expt=="E4")

cv <- function(x) { sd(x)/mean(x) }
cv(dat1$yield)
cv(dat2$yield) # Does not match Smith. Checked all values by hand.
cv(dat3$yield)
cv(dat4$yield)

if(require("desplot")){
  desplot(yield ~ col*row, dat1,
          aspect=180/45, flip=TRUE, # true aspect
          main="smith.beans.uniformity, expt 1 (true aspect)")
}
if(require("desplot")){
```

```

desplot(yield ~ col*row, dat2,
        aspect=180/45, flip=TRUE, # true aspect
        main="smith.beans.uniformity, expt 2 (true aspect)")
}
if(require("desplot")){
  desplot(yield ~ col*row, dat3,
          aspect=450/80, flip=TRUE, # true aspect
          main="smith.beans.uniformity, expt 3 (true aspect)")
}
if(require("desplot")){
  desplot(yield ~ col*row, dat4,
          aspect=450/80, flip=TRUE, # true aspect
          main="smith.beans.uniformity expt 4, (true aspect)")
}

```

smith.corn.uniformity *Uniformity trial of corn, 3 years on same ground*

Description

Uniformity trial of corn, 3 years on same ground, 1895-1897

Format

A data frame with 360 observations on the following 5 variables.

row row

col column

plot plot number, consistent across years

year year. Last two digits of 1895, 1896, 1897

yield yield, bushels / acre

Details

The data values are from Smith (1910) and the field map is from Harris (1920). Each plot was 1/10 acre, but the dimensions are not given.

The following text is abridged from Smith (1910).

How much variability may we reasonably expect in land that is apparently uniform? Some data among the records of the soil plots at the Illinois Experiment station furnish interesting material for study in this connection.

A field that had lain sixteen years in pasture was broken up in 1895 and laid out into plots to be subsequently used for soil experiments. The land is slightly rolling but otherwise quite uniform in appearance. There are in the series to be considered in this connection 120 one-tenth acre plots. These plots were all planted to corn for three consecutive years without any soil treatment, so that the records offer a rather exceptional opportunity for a study of this kind.

The yields of all these plots in terms of bushels of shelled corn per acre for the three years are given in the accompanying table.

A study of this data reveals some very striking variations. It will be noticed in the first place that there is a tremendous difference in production in the different years. The first year, 1895, was an extremely unfavorable one for corn and the yields are exceptionally low. The weather records show that the season was not only unusually dry, but also cool in the early part. The following year we have an exceptionally favorable corn season, and the yields run unusually high. The third year was also a good one, and the yields are perhaps somewhat above the normal for this locality.

It will be observed that certain plots appear to be very abnormal. Thus plots 117, 118, 119, and 120 give an abnormally high yield in the first season and an abnormally low one in the two following years. This is to be accounted for in the topography of the land. These plots lie in a low spot which was favorable in the dry year of 1895, but unfavorable in 1896 and 1897. For this reason these four plots were rejected from further consideration in this study, as were also plots 616, 617, 618, 619, and 620. This leaves 111 plots whose variations are apparently unaccounted for and which furnish the data from which the following results are taken.

It is noticeable that the variability as measured by the standard deviation becomes less in each succeeding year. This suggests the question as to whether continued cropping might not tend to induce uniformity. The records of a few of these plots which were continued in corn for three years longer, however, do not support such a conclusion.

It seems reasonable to expect greater variability in seasons very unfavorable for production, such as that of 1895, because so much may depend upon certain critical factors of production coming into play and this suggestion may be the explanation of the high standard deviation in this first year. Results extending over a longer series of years would be extremely interesting in this connection.

If we consider the total range of variation in any single year, we find differences as follows: Plots lying adjoining have shown the following maximum variations: 18 bushels in 1895; 11 bushels in 1896; 8 bushels in 1897.

The above results give us a conception of the unaccountable plot variations which we have to deal with in field tests. The possibility remains that a still closer study might detect some abnormal factors at play to account for these variations in certain cases, but the study certainly suggests the importance of conservatism in arriving at conclusions based upon plot tests.

The particular value that the writer has derived from this study is the strengthening of his conviction that the only dependence to be placed upon variety tests and other field experiments is from records involving the average of liberal numbers and extending over long periods of time.

Source

Smith, L.H. 1910. Plot arrangement for variety experiments with corn. *Agronomy Journal*, 1, 84–89. Table 1. <http://books.google.com/books?id=mQT0AAAAMAAJ&pg=PA84>

Harris, J.A. 1920. Practical universality of field heterogeneity as a factor influencing plot yields. *Journal of Agricultural Research*, 19, 279–314. Page 296-297. <http://books.google.com/books?id=jyEXAAAAYAAJ&pg=PA296>

Examples

```
data(smith.corn.uniformity)
dat <- smith.corn.uniformity
```

```

dat = transform(dat, year=factor(year))
if(require(desplot)){
  desplot(yield~col*row|year, dat, layout=c(1,3),
          col.regions=colorRampPalette(gray(1:100/100)),
          main="smith.corn.uniformity: yield across years 1895-1987")
}

## # Outliers are obvious
## require(lattice)
## xyplot(yield~row|factor(col), dat, groups=year,
##        auto.key=list(columns=3), main="smith.corn.uniformity")

## Not run:
require(rgl)
# A few odd pairs of outliers in column 6
plot3d(dat$col, dat$row, dat$yield, col=dat$year,
        xlab="col", ylab="row", zlab="yield")
rgl.close()

## End(Not run)

```

snedecor.asparagus *Asparagus yields for different cutting treatments*

Description

Asparagus yields for different cutting treatments, in 4 years.

Format

A data frame with 64 observations on the following 4 variables.

block block factor, 4 levels

year year, numeric

trt treatment factor of final cutting date

yield yield, ounces

Details

Planted in 1927. Cutting began in 1929. Yield is the weight of asparagus cuttings up to Jun 1 in each plot. Some plots received continued cuttings until Jun 15, Jul 1, and Jul 15.

In the past, repeated-measurement experiments like this were sometimes analyzed as if they were a split-plot experiment. This violates some independence assumptions.

Source

Snedecor and Cochran, 1989. *Statistical Methods*.

References

Mick O'Neill, 2010. *A Guide To Linear Mixed Models In An Experimental Design Context*. Statistical Advisory & Training Service Pty Ltd.

Examples

```

data(snedecor.asparagus)
dat <- snedecor.asparagus

dat <- transform(dat, year=factor(year))
dat$trt <- factor(dat$trt,
                 levels=c("Jun-01", "Jun-15", "Jul-01", "Jul-15"))
# Continued cutting reduces plant vigor and yield
require(lattice)
dotplot(yield ~ trt|year, data=dat,
        xlab="Cutting treatment", main="snedecor.asparagus")

# Split-plot
## Not run:
require(lme4)
m1 <- lmer(yield ~ trt + year + trt:year + (1|block) + (1|block:trt), data=dat)

## End(Not run)

# -----
## Not run:
# Split-plot with asreml
# asreml3
require(asreml)
m2 <- asreml(yield ~ trt + year + trt:year, data=dat,
            random = ~ block + block:trt)

require(lucid)
vc(m2)
##           effect component std.error z.ratio constr
##   block!block.var    476.7    518.2    0.92   pos
## block:trt!block.var    499.7    287.4    1.7    pos
##           R!variance    430.2    101.4    4.2    pos

# Antedependence with asreml. See O'Neill (2010).
dat <- dat[order(dat$block, dat$trt), ]
m3 <- asreml(yield ~ year * trt, data=dat,
            random = ~ block,
            rcov = ~ block:trt:ante(year,1))

# Extract the covariance matrix for years and convert to correlation
covmat <- diag(4)
covmat[upper.tri(covmat,diag=TRUE)] <- m3$R.param$R$year$initial
covmat[lower.tri(covmat)] <- t(covmat)[lower.tri(covmat)]
round(cov2cor(covmat),2) # correlation among the 4 years

```



```

#      [,1] [,2] [,3] [,4]
# [1,] 1.00 0.45 0.39 0.31
# [2,] 0.45 1.00 0.86 0.69
# [3,] 0.39 0.86 1.00 0.80
# [4,] 0.31 0.69 0.80 1.00

# We can also build the covariance Sigma by hand from the estimated
# variance components via: Sigma^-1 = U D^-1 U'
vv <- vc(m3)
print(vv)
##          effect component std.error z.ratio constr
## block!block.var  86.56   156.9      0.55   pos
##      R!variance    1           NA      NA    fix
## R!year.1930:1930  0.00233  0.00106    2.2 uncon
## R!year.1931:1930 -0.7169   0.4528   -1.6 uncon
## R!year.1931:1931  0.00116  0.00048    2.4 uncon
## R!year.1932:1931 -1.139    0.1962   -5.8 uncon
## R!year.1932:1932  0.00208  0.00085    2.4 uncon
## R!year.1933:1932 -0.6782   0.1555   -4.4 uncon
## R!year.1933:1933  0.00201  0.00083    2.4 uncon

U <- diag(4)
U[1,2] <- vv[4,2] ; U[2,3] <- vv[6,2] ; U[3,4] <- vv[8,2]
Dinv <- diag(c(vv[3,2], vv[5,2], vv[7,2], vv[9,2]))
# solve(U
solve(crossprod(t(U), tcrossprod(Dinv, U)) )
##      [,1]      [,2]      [,3]      [,4]
## [1,] 428.4310  307.1478  349.8152  237.2453
## [2,] 307.1478 1083.9717 1234.5516  837.2751
## [3,] 349.8152 1234.5516 1886.5150 1279.4378
## [4,] 237.2453  837.2751 1279.4378 1364.8446

## End(Not run)

# -----

## Not run:
# Split-plot with asreml
## require(asreml4)
## m2 <- asreml(yield ~ trt + year + trt:year, data=dat,
##             random = ~ block + block:trt)

## require(lucid)
## vc(m2)
## ##      effect component std.error z.ratio bound
## ##      block      476.7    518.2    0.92    P    0
## ## block:trt     499.7    287.4    1.7     P    0
## ## units(R)     430.2    101.4    4.2     P    0

## # Antedependence with asreml. See O'Neill (2010).
## dat <- dat[order(dat$block, dat$trt), ]
## m3 <- asreml(yield ~ year * trt, data=dat,
##             random = ~ block,

```

```

##          residual = ~ block:trt:ante(year,1),
##          max=50)

## # Extract the covariance matrix for years and convert to correlation
## covmat <- diag(4)
## covmat[upper.tri(covmat,diag=TRUE)] <- m3$R.param$`block:trt:year`$year$initial
## covmat[lower.tri(covmat)] <- t(covmat)[lower.tri(covmat)]
## round(cov2cor(covmat),2) # correlation among the 4 years
## #      [,1] [,2] [,3] [,4]
## # [1,] 1.00 0.45 0.39 0.31
## # [2,] 0.45 1.00 0.86 0.69
## # [3,] 0.39 0.86 1.00 0.80
## # [4,] 0.31 0.69 0.80 1.00

## # We can also build the covariance Sigma by hand from the estimated
## # variance components via: Sigma^-1 = U D^-1 U'
## vv <- vc(m3)
## print(vv)
## ##          effect component std.error z.ratio constr
## ## block!block.var 86.56    156.9      0.55    pos
## ##          R!variance 1          NA      NA      fix
## ## R!year.1930:1930 0.00233 0.00106 2.2    uncon
## ## R!year.1931:1930 -0.7169 0.4528  -1.6    uncon
## ## R!year.1931:1931 0.00116 0.00048 2.4    uncon
## ## R!year.1932:1931 -1.139 0.1962  -5.8    uncon
## ## R!year.1932:1932 0.00208 0.00085 2.4    uncon
## ## R!year.1933:1932 -0.6782 0.1555  -4.4    uncon
## ## R!year.1933:1933 0.00201 0.00083 2.4    uncon

## U <- diag(4)
## U[1,2] <- vv[4,2] ; U[2,3] <- vv[6,2] ; U[3,4] <- vv[8,2]
## Dinv <- diag(c(vv[3,2], vv[5,2], vv[7,2], vv[9,2]))
## # solve(U
## solve(crossprod(t(U), tcrossprod(Dinv, U)) )
## ##          [,1] [,2] [,3] [,4]
## ## [1,] 428.4310 307.1478 349.8152 237.2453
## ## [2,] 307.1478 1083.9717 1234.5516 837.2751
## ## [3,] 349.8152 1234.5516 1886.5150 1279.4378
## ## [4,] 237.2453 837.2751 1279.4378 1364.8446

## End(Not run)

# -----

```

Description

Infection in wheat by different strains of Fusarium.

Format

A data frame with 204 observations on the following 4 variables.

```
gen wheat genotype
strain fusarium strain
year year
y Percent infected
```

Details

The data are the percent of leaf area affected by Fusarium head blight, averaged over 4-5 reps, for 17 winter wheat genotypes.

VanEeuwijk fit a generalized ammi-2 model to this data. It is a generalized model in the sense that a link function is used, and is a non-linear AMMI model in that there are main effects for variety and year-strain, but additional multiplicative effects for the interactions.

Note, the value for strain F348 in 1988, gen SVP75059-32 should be 28.3 (as shown in VanEeuwijk 1995) and not 38.3 (as shown in Snijders 1991).

Source

Snijders, CHA and Van Eeuwijk, FA. 1991. Genotype x strain interactions for resistance to Fusarium head blight caused by Fusarium culmorum in winter wheat. *Theoretical and Applied Genetics*, 81, 239–244. Table 1. <http://doi.org/10.1007/BF00215729>

Used with permission of Fred van Eeuwijk.

References

Fred A van Eeuwijk. 1995. Multiplicative interaction in generalized linear models. *Biometrics*, 51, 1017-1032. <http://doi.org/10.2307/2533001>

Examples

```
data(snijders.fusarium)
dat <- snijders.fusarium

aggregate(y ~ strain + year, dat, FUN=mean) # Match means in Snijders table 1

dat <- transform(dat, y=y/100, year=factor(year), yrstr=factor(paste0(year,"-",strain)))

# Strain F329 shows little variation across years. F39 shows a lot.
require(lattice)
dotplot(gen~y|strain, data=dat, group=year,
        main="snijders.fusarium : infection by strain",
        xlab="Fraction infected", ylab="variety",
```

```

auto.key=list(columns=3))

# Logit transform
dat <- transform(dat, logit=log(y/(1-y)))
m1 <- aov(logit ~ yrstr + gen, data=dat) # Match SS in VanEeuwijk table 4
anova(m1) # Match SS in VanEeuwijk table 4
m2 <- aov(logit ~ year*strain + gen + gen:year + gen:strain, data=dat)
anova(m2) # Match to VanEeuwijk table 5

## Not run:

# GLM on untransformed data using logit link, variance  $\mu^2(1-\mu)^2$ 
require(gnm) # for 'wedderburn' family
m2 <- glm(y ~ yrstr + gen, data=dat, family="wedderburn")
anova(m2) # Main effects match VanEeuwijk table 6

# Generalized AMMI-2 model. Matches VanEeuwijk table 6
bilin2 <- gnm(y ~ yrstr + gen + instances(Mult(yrstr, gen), 2),
             data=dat, family = wedderburn)
# plot(bilin2,1) # Resid vs fitted plot matches VanEeuwijk figure 3c
## anova(bilin2)
##
##              Df Deviance Resid. Df Resid. Dev
## NULL                    203      369.44
## yrstr                   11  150.847     192     218.60
## gen                      16  145.266     176      73.33
## Mult(yrstr, gen, inst = 1) 26   26.128     150      47.20
## Mult(yrstr, gen, inst = 2) 24   19.485     126      27.72

# Manually extract coordinates for biplot
cof <- coef(bilin2)
y1 <- cof[29:40]
g1 <- cof[41:57]
y2 <- cof[58:69]
g2 <- cof[70:86]
g12 <- cbind(g1,g2)
rownames(g12) <- substring(rownames(g12), 29)
y12 <- cbind(y1,y2)
rownames(y12) <- substring(rownames(y12), 31)
g12[,1] <- -1 * g12[,1]
y12[,1] <- -1 * y12[,1]
# GAMMI biplot. Inner-products of points projected onto
# arrows match VanEeuwijk figure 4. Slight rotation of graph is ignorable.
biplot(y12, g12, cex=.75, main="snijders.fusarium") # Arrows to genotypes.

## End(Not run)

```

Description

Uniformity trial of sorghum silage at Chillicothe, Texas, 1915.

Format

A data frame with 2000 observations on the following 3 variables.

```
row row
col column / rod
yield yield, ounces
```

Details

Grown near Chillicothe, TX in 1915. Rows 40 inches apart. Each row harvested in 1-rod (16.5 ft) lengths. East side higher yielding than west side. Yields are weight (ounces) of green forage each rod-row. Total area harvested: $100 \times 40 / 12 = 333.33$ feet by $20 \times 16.5 = 330$ feet.

Field width: $20 \text{ plots} \times 16.5 \text{ ft (1 rod)} = 330 \text{ feet}$.

Field length: $100 \text{ plots} \times 40 \text{ in} = 333 \text{ feet}$

Source

Stephens, Joseph C. 1928. Experimental methods and the probable error in field experiments with sorghum. *Journal of Agricultural Research*, 37, 629–646. <http://naldc.nal.usda.gov/catalog/IND43967516>

Examples

```
data(stephens.sorghum.uniformity)
dat <- stephens.sorghum.uniformity

dat <- subset(dat, row>2 & row<99) # omit outer two rows
# mean(dat$yield) # 180.27
# range(dat$yield) # 75,302 matches Stephens

# densityplot(~dat$yield) # Stephens figure 3

# Aggregate 4 side-by-side rows.
d4 <- dat
d4$row2 <- ceiling((d4$row-2)/4)
d4 <- aggregate(yield ~ row2+col, data=d4, FUN=sum)
d4$row2 <- 25-d4$row2 # flip horizontally

if(require(desplot)){
  grays <- colorRampPalette(c("#d9d9d9", "#252525"))
  desplot(yield ~ row2*col, data=d4,
    aspect=333/330, flip=TRUE, # true aspect
    main="stephens.sorghum.uniformity",
    col.regions=grays(3),
    at=c(500,680,780,1000))
  # Similar to Stephens Figure 7. North at top. East at right.
```

}

| | |
|---------------------|---|
| steptoe.morex.pheno | <i>Multi-environment trial of barley, phenotypic and genotypic data for a population of Steptoe x Morex</i> |
|---------------------|---|

Description

Phenotypic and genotypic data for a barley population of Steptoe x Morex. There were 150 doubled haploid crosses, evaluated at 223 markers. Phenotypic data was collected on 8 traits at 16 environments.

Usage

```
data("steptoe.morex.pheno")
```

Format

steptoe.morex.pheno is a data.frame of phenotypic data with 2432 observations on 10 variables:

gen genotype factor with parents Steptoe and Morex, and 150 crosses SM1, SM2, ..., SM200. Not all 200 numbers were used.

env environment, 16 levels

amylase alpha amylase (20 Deg Units)

diapow diastatic power (degree units)

hddate heading date (julian days)

lodging lodging (percent)

malt malt extract (percent)

height plant height (centimeters)

protein grain protein (percent)

yield grain yield (Mt/Ha)

steptoe.morex.geno is a cross object from the qt1 package with genotypic data of the 223 markers for the 150 crosses of Steptoe x Morex.

Details

As described by Hayes et al (1993), a population of 150 barley doubled haploid (DH) lines was developed by the Oregon State University Barley Breeding Program for the North American Barley Genome Mapping Project. The parentage of the population is Steptoe / Morex.

Steptoe is the dominant feed barley in the northwestern U.S.

Morex is the spring U.S. malting quality standard.

Seed from a single head of each parent was used to create the F1, from which a set of 150 lines was developed.

Phenotypic values for the parents Steptoe and Morex are here: http://wheat.pw.usda.gov/ggpages/SxM/parental_values.html

There are 16 locations, The average across locations is in column 17. Not all traits were collected at every location. At each location, all 150 lines were included in block 1, a random subset of 50 lines was used in block 2.

The traits are: Alpha Amylase (20 Deg Units), Diastatic Power (Deg Units), Heading Date (Julian Days), Lodging (percent), Malt Extract (percent), Grain Protein (percent), Grain Yield (Mt/Ha).

Phenotypic values of the 150 lines in the F1 population are here: <http://wheat.pw.usda.gov/ggpages/SxM/phenotypes.html>

Each trait is in a different file, in which each block of numbers represents one location.

The 223-markers Steptoe/Morex base map is here: <http://wheat.pw.usda.gov/ggpages/SxM/smbasev2.map>

. The data for these markers on the 150 lines is <http://wheat.pw.usda.gov/ggpages/SxM/smbasev2.mrk>

These were hand-assembled (e.g. marker distances were cumulated to marker positions) into a .csv file which was then imported into R using `qt1::read.cross`. The class was manually changed from `c('bc','cross')` to `c('dh','cross')`.

The marker data is coded as A = Steptoe, B = Morex, - = missing.

The pedigrees for the 150 lines are found here: <http://wheat.pw.usda.gov/ggpages/SxM/pedigrees.html>

Source

The phenotypic, genotypic, pedigree, etc information can be found from various pages found at the following site:

The Steptoe x Morex Barley Mapping Population. Map: Version 2, August 1, 1995 <http://wheat.pw.usda.gov/ggpages/SxM>. Accessed Jan 2015.

Data provided by the United States Department of Agriculture.

References

P.M. Hayes, B.H. Liu, S.J. Knapp, F. Chen, B. Jones, T. Blake, J. Franckowiak, D. Rasmusson, M. Sorrells, S.E. Ullrich, and others. 1993. Quantitative trait locus effects and environmental interaction in a sample of North American barley germplasm. *Theoretical and Applied Genetics*, 87, 392–401. <http://doi.org/10.1007/BF01184929>

Ignacio Romagosa, Steven E. Ullrich, Feng Han, Patrick M. Hayes. 1996. Use of the additive main effects and multiplicative interaction model in QTL mapping for adaptation in barley. *Theor Appl Genet*, 93, 30-37. <http://doi.org/10.1007/BF00225723>

Piepho, Hans-Peter. 2000. A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. *Genetics*, 156, 2043-2050.

M. Malosetti, J. Voltas, I. Romagosa, S.E. Ullrich, F.A. van Eeuwijk. (2004). Mixed models including environmental covariables for studying QTL by environment interaction. *Euphytica*, 137, 139-145. <http://doi.org/10.1023/B:EUPH.0000040511.4638>

Examples

```
data(steptoe.morex.pheno)
dat <- steptoe.morex.pheno
```

```

## Not run:
# Visualize GxE of traits
require(lattice)
levelplot(amylase~env*gen, data=dat, scales=list(x=list(rot=90)), main="amylase")
levelplot(diapow~env*gen, data=dat, scales=list(x=list(rot=90)), main="diapow")
levelplot(hddate~env*gen, data=dat, scales=list(x=list(rot=90)), main="hddate")
levelplot(lodging~env*gen, data=dat, scales=list(x=list(rot=90)), main="lodging")
levelplot(malt~env*gen, data=dat, scales=list(x=list(rot=90)), main="malt")
levelplot(height~env*gen, data=dat, scales=list(x=list(rot=90)), main="height")
levelplot(protein~env*gen, data=dat, scales=list(x=list(rot=90)), main="protein")
levelplot(yield~env*gen, data=dat, scales=list(x=list(rot=90)), main="yield")

## End(Not run)

# -----

# Calculate avg yield for each loc as in Romagosa 1996, table 3
t(t(round(tapply(dat$yield, dat$env, FUN=mean),2)))
# SKo92,SKg92 means in table 3 are switched. Who is right, him or me?

# -----

# Draw marker map
if(require(qtl)){
  data(steptoe.morex.geno)
  datg <- steptoe.morex.geno
  plot.map(datg, main="steptoe.morex.pheno") # or just use plot()
}

# -----

## Not run:
# This is a very rudimentary example.
# Fit a simple multi-environment mixed model
# asreml3
require(asreml)
m1 <- asreml(yield ~ env, data=dat, random=~gen)

require(wgaim)
qtl::plotMissing(datg)
link.map(datg)
# Create an interval object for wgaim
class(datg)[1] <- "bc"
dati <- cross2int(datg, id="gen")

# Whole genome qtl
assign("dat", dat, 1) # Needed for wgaim in script mode
q1 <- wgaim(m1, dat, dati, merge.by="gen", na.method.X='include')
link.map(q1, dati) # Visualize
out.stat(q1, dati) # outlier statistic
summary(q1, dati) # Table of important intervals
# Chrom Left Marker dist(cM) Right Marker dist(cM) Size Pvalue
# 3 ABG399 52.6 BCD828 56.1 0.254 0.000 45.0

```



```

#      5      MWG912      148      ABG387A      151.2  0.092  0.001  5.9
#      6      ABC169B      64.8      CD0497      67.5 -0.089  0.001  5.6

## End(Not run)

# -----

## Not run:
# This is a very rudimentary example.

# Fit a simple multi-environment mixed model
## require(asreml4)
## m1 <- asreml(yield ~ env, data=dat, random=~gen)

## require(wgaim)
## plotMissing(datg)
## link.map(datg)
## # Create an interval object for wgaim
## class(datg)[1] <- "bc"
## dati <- cross2int(datg, id="gen")

## # Whole genome qtl
## assign("dat", dat, 1) # Needed for wgaim in script mode
## q1 <- wgaim(m1, dat, dati, merge.by="gen", na.method.X='include')
## link.map(q1, dati) # Visualize
## out.stat(q1, dati) # outlier statistic
## summary(q1, dati) # Table of important intervals
## # Chrom Left Marker dist(cM) Right Marker dist(cM) Size Pvalue
## #      3      ABG399      52.6      BCD828      56.1  0.254  0.000  45.0
## #      5      MWG912      148      ABG387A      151.2  0.092  0.001  5.9
## #      6      ABC169B      64.8      CD0497      67.5 -0.089  0.001  5.6

## End(Not run)

# -----

```

```
stickler.sorghum.uniformity
```

Uniformity trial of sorghum

Description

Uniformity trial of sorghum in at Manhattan, Kansas, 1958

Usage

```
data("stickler.sorghum.uniformity")
```

Format

A data frame with 1600 observations on the following 4 variables.

```
expt experiment
row row
col col
yield yield
```

Details

Four sorghum experiments at the Agronomy Farm at Manhattan, Kansas. Experiments E1,E2 grown in 1958. Expts E3,E5 grown in 1959.

Experiment E1.

Field width = 20 units * 14 inches = 23.3 ft.

Field length = 20 units * 10 feet = 200 feet.

Experiment E2-E3.

Field width = 20 units * 44 inches = 73 feet

Field length = 20 units * 5 ft = 100 feet.

Source

F. C. Stickler (1960). Estimates of Optimum Plot Size from Grain Sorghum Uniformity Trial Data. Technical bulletin, Kansas Agricultural Experiment Station.

References

None.

Examples

```
data(stickler.sorghum.uniformity)
dat <- stickler.sorghum.uniformity

dat1 <- subset(dat, expt=="E1")
dat2 <- subset(dat, expt!="E1")

if(require(desplot)){
  desplot(yield ~ col*row|expt, data=dat1,
          #cex=1,text=yield, shorten="none",
          xlab="row",ylab="range",
          flip=TRUE, tick=TRUE, aspect=(20*10)/(20*14/12), # true aspect
          main="stickler.sorghum.uniformity: expt E1")
}
if(require(desplot)){
  desplot(yield ~ col*row|expt, data=dat2,
          xlab="row",ylab="range",
          flip=TRUE, tick=TRUE, aspect=(20*5)/(20*44/12), # true aspect
          main="stickler.sorghum.uniformity: expt E2,E3,E4")
}
```

```

}

# Stickler, p. 10-11 has
#   E1   E2   E3   E4
# 34.81 11.53 11.97 14.10
cv <- function(x) 100*sd(x)/mean(x)
tapply(dat$yield, dat$expt, cv)
# 35.74653 11.55062 11.97011 14.11389

```

stirret.borers

Corn borer control by application of fungal spores.

Description

Corn borer control by application of fungal spores.

Format

A data frame with 60 observations on the following 4 variables.

block block, 15 levels

trt treatment, 4 levels

count1 count of borers on August 18

count2 count of borers on October 19

Details

Experiment conducted in 1935, Ottawa. European corn borer infestation was established by application of egg masses to plants. Treatments were applied on July 8 and July 19 at two levels, 0 and 40 grams per acre. The number of borers per plot were counted on Aug 18 and Oct 19.

Source

Stirrett, George M and Beall, Geoffrey and Timonin, M. (1937). A field experiment on the control of the European corn borer, *Pyrausta nubilalis* Hubn, by *Beauveria bassiana* Vuill. *Sci. Agric.*, 17, 587–591. Table 2.

Examples

```

data(stirret.borers)
dat <- stirret.borers

require(lattice)
xyplot(count2~count1|trt,dat, main="stirret.borers - by treatment",
       xlab="Early count of borers", ylab="Late count")

```

```
# Even though the data are counts, Normal distribution seems okay
# qqmath(~count1|trt, dat, main="stirret.borers")

m1 <- lm(count1 ~ trt + block, dat)
anova(m1)

if(require(effects)){
  e1 <- effect('trt',m1)
  as.data.frame(e1)
  plot(e1, main="stirret.borer")
}
```

streibig.competition *Competition experiment between barley and sinapis.*

Description

Competition experiment between barley and sinapis, at different planting rates.

Format

A data frame with 135 observations on the following 8 variables.

pot pot number
bseeds barley seeds sown
sseeds sinapis seeds sown
block block
bfmt barley fresh weight
sfwt sinapis fresh weight
bdwt barley dry weight
sdwt sinapis dry weight

Details

The source data (in McCullagh) also contains a count of plants harvested (not included here) that sometimes is greater than the number of seeds planted.

Source

Peter McCullagh, John A. Nelder. *Generalized Linear Models*, page 318-320.
Used with permission of Jens Streibig.

References

Oliver Schabenberger and Francis J Pierce. 2002. *Contemporary Statistical Models for the Plant and Soil Sciences*, CRC Press. Page 370-375.

Examples

```

data(streibig.competition)
dat <- streibig.competition

# See Schabberger and Pierce, pages 370+
# Consider only the mono-species barley data (no competition from sinapis)
d1 <- subset(dat, sseeds<1)
d1 <- transform(d1, x=bseeds, y=bdwt, block=factor(block))

# Inverse yield looks like it will be a good fit for Gamma's inverse link
require(lattice)
xyplot(1/y~x, data=d1, group=block, auto.key=list(columns=3),
       xlab="Seeding rate", ylab="Inverse yield of barley dry weight",
       main="streibig.competition")

# linear predictor is quadratic, with separate intercept and slope per block
m1 <- glm(y ~ block + block:x + x+I(x^2), data=d1,
         family=Gamma(link="inverse"))
# Predict and plot
newdf <- expand.grid(x=seq(0,120,length=50), block=factor(c('B1','B2','B3')))
newdf$pred <- predict(m1, new=newdf, type='response')
plot(y~x, data=d1, col=block, main="streibig.competition - by block",
     xlab="Barley seeds", ylab="Barley dry weight")
for(bb in 1:3){
  newbb <- subset(newdf, block==c('B1','B2','B3')[bb])
  lines(pred~x, data=newbb, col=bb)
}

```

stroup.nin

RCB experiment of wheat at the Nebraska Intrastate Nursery

Description

The yield data from an advanced Nebraska Intrastate Nursery (NIN) breeding trial conducted at Alliance, Nebraska, in 1988/89.

Format

gen genotype, 56 levels

rep replicate, 4 levels

yield yield, bu/ac

col column

row row

Details

Four replicates of 19 released cultivars, 35 experimental wheat lines and 2 additional triticale lines were laid out in a 22 row by 11 column rectangular array of plots. The varieties were allocated to the plots using a randomised complete block (RCB) design. Each plot was sown in four rows 4.3 m long and 0.3 m apart. Plots were trimmed down to 2.4 m in length before harvest. The orientation of the plots is not clear from the paper, but the data in Littell et al are given in meters and make the orientation clear.

Field width: 22 plots * 1.2 m = 26.4 m

Field length: 11 plots * 4.3 m = 47.3 m

All plots with missing data are coded as being gen = "Lancer". (For ASREML, missing plots need to be included for spatial analysis and the level of 'gen' needs to be one that is already in the data.)

These data were first analyzed by Stroup et al (1994) and subsequently by Littell et al (1996, page 321), Pinheiro and Bates (2000, page 260), and Butler et al (2004).

This version of the data give the yield in bushels per acre. The yield values published in Stroup et al (1994) are expressed in kg/ha. For wheat, 1 bu/ac = 67.25 kg/ha.

Some of the gen names are different in Stroup et al (1994). (Sometimes an experimental genotype is given a new name when it is released for commercial use.) At a minimum, the following differences in gen names should be noted:

| | |
|------------|--------------|
| stroup.nin | Stroup et al |
| NE83498 | Rawhide |
| KS831374 | Karl |

Some published versions of the data use long/lat instead of col/row. To obtain the correct value of 'long', multiply 'col' by 1.2. To obtain the correct value of 'lat', multiply 'row' by 4.3.

Relatively low yields were clustered in the northwest corner, which is explained by a low rise in this part of the field, causing increased exposure to winter kill from wind damage and thus depressed yield. The genotype 'Buckskin' is a known superior variety, but was disadvantaged by assignment to unfavorable locations within the blocks.

Note that the figures in Stroup 2002 claim to be based on this data, but the number of rows and columns are both off by 1 and the positions of Buckskin as shown in Stroup 2002 do not appear to be quite right.

Source

Stroup, Walter W., P Stephen Baenziger, Dieter K Mulitze (1994) Removing Spatial Variation from Wheat Yield Trials: A Comparison of Methods. *Crop Science*, 86:62–66. <http://doi.org/10.2135/cropsci1994.0011183X0034>

References

Littell, R.C. and Milliken, G.A. and Stroup, W.W. and Wolfinger, R.D. 1996. *SAS system for mixed models*, SAS Institute, Cary, NC.

Jose Pinheiro and Douglas Bates, 2000, *Mixed Effects Models in S and S-Plus*, Springer.

Butler, D., B R Cullis, A R Gilmour, B J Goegel. (2004) *Spatial Analysis Mixed Models for S language environments*

W. W. Stroup (2002). Power Analysis Based on Spatial Effects Mixed Models: A Tool for Comparing Design and Analysis Strategies in the Presence of Spatial Variability. *Journal of Agricultural, Biological, and Environmental Statistics*, 7(4), 491-511. <http://doi.org/10.1198/108571102780>

See Also

Identical data (except for the missing values) are available in the nlme package as [Wheat2](#).

Examples

```
data(stroup.nin)
dat <- stroup.nin

# Experiment layout. All "Buckskin" plots are near left side
if(require(desplot)){
  desplot(yield~col*row, dat,
          aspect=47.3/26.4, out1="rep", num=gen, cex=0.6, # true aspect
          main="stroup.nin - yield heatmap (true shape)")
}

# ----- nlme -----

require(nlme)
# Random RCB model
lme1 <- lme(yield ~ 0 + gen, random=~1|rep, data=dat, na.action=na.omit)

# Linear (Manhattan distance) correlation model
lme2 <- gls(yield ~ 0 + gen, data=dat,
            correlation = corLin(form = ~ col + row, nugget=TRUE),
            na.action=na.omit)

# Random block and spatial correlation.
# Note: corExp and corSpher give nearly identical results
lme3 <- lme(yield ~ 0 + gen, data=dat,
            random = ~ 1 | rep,
            correlation = corExp(form = ~ col + row),
            na.action=na.omit)

AIC(lme1,lme2,lme3) # lme2 is lowest
##      df      AIC
## lme1 58 1333.702
## lme2 59 1189.135
## lme3 59 1216.704

# Compare the estimates from the two methods
eff = data.frame(ranblock=fixef(lme1), corLin = coef(lme2),
                 corExp = fixef(lme3))
rownames(eff) <- gsub("gen", "", rownames(eff))
plot(eff$ranblock, eff$corLin, xlim=c(13,37), ylim=c(13,37),
     main="stroup.nin - Gen estimates (intercepts differ)",
     xlab="RCB (random block)", ylab="corLin",type='n')
```

```

text(eff$ranblock, eff$corLin, rownames(eff), cex=0.5)
abline(0,1)

# -----

# Note, the 'sommer' package has dropped support for ar1

# -----

## Not run:
if(require(SpATS)){

  dat <- transform(dat, yf = as.factor(row), xf = as.factor(col))

  sp1 <- SpATS(response = "yield",
              spatial = ~ SAP(col, row, nseg = c(10,20), degree = 3, pord = 2),
              genotype = "gen",
              #fixed = ~ colcode + rowcode,
              random = ~ yf + xf,
              data = dat,
              control = list(tolerance = 1e-03))

  plot(sp1)

  eff <- cbind(eff, spats=predict(sp1, which="gen")$predicted.values)
  plot(eff$ranblock, eff$spats, xlim=c(13,37), ylim=c(13,37),
      main="stroup.nin - Gen estimates (intercepts differ)",
      xlab="RCB (random block)", ylab="SpATS", type='n')
  text(eff$ranblock, eff$spats, rownames(eff), cex=0.5)
  abline(0,1)
}

## End(Not run)

# -----

## Not run:
# asreml3
require(asreml)

# RCB analysis
dat.rcb <- asreml(yield ~ gen, random = ~ rep, data=dat,
                 na.method.X="omit")
pred.rcb <- predict(dat.rcb, data=dat, classify="gen")$predictions

# Two-dimensional AR1xAR1 spatial model
dat <- transform(dat, xf=factor(col), yf=factor(row))
dat <- dat[order(dat$xf, dat$yf),]
dat.sp <- asreml(yield~gen, data=dat,
                rcov=~ar1(xf):ar1(yf),
                na.method.X='ignore')
pred.sp <- predict(dat.sp, data=dat, classify="gen")$predictions

```



```

require(lucid)
vc(dat.sp)
##      effect component std.error z.ratio constr
## R!variance  48.7      7.155      6.8   pos
## R!xf.cor    0.6555   0.05638   12    unc
## R!yf.cor    0.4375   0.0806    5.4    unc

# Compare the estimates from the two methods
plot(pred.rcb$pvals[,2], pred.sp$pvals[,2], xlim=c(13,37), ylim=c(13,37),
      xlab="RCB",ylab="AR1xAR1",type='n')
title("stroup.nin: Comparison of predicted values")
text(pred.rcb$pvals[,2],pred.sp$pvals[,2],
      as.character(pred.rcb$pvals[,1]),cex=0.5)
abline(0,1)

## End(Not run)

# -----

## Not run:

## require(asreml4)
## data(stroup.nin)
## dat <- stroup.nin

## # Note, ASREML4 stand-alone might not need a completely-filled rectangle
## # of plots to perform AR1xAR1 analysis. It might be able to fill in missing
## # combinations of rows/columns. The R version asreml4 still seems to require
## # rectangles.

## # RCB analysis
## dat.rcb <- asreml(yield ~ gen, random = ~ rep, data=dat,
##                  na.action=na.method(x="omit"))
## pred.rcb <- predict(dat.rcb, data=dat, classify="gen")

## # Two-dimensional AR1xAR1 spatial model
## dat <- transform(dat, xf=factor(col), yf=factor(row))
## #dat <- dat[order(dat$xf, dat$yf),]
## dat.sp <- asreml(yield~gen, data=dat,
##                 residual = ~ar1(xf):ar1(yf),
##                 na.action=na.method(x="omit"))
## pred.sp <- predict(dat.sp, data=dat, classify="gen")

## require(lucid)
## vc(dat.sp)
## ##      effect component std.error z.ratio constr
## ## R!variance  48.7      7.155      6.8   pos
## ## R!xf.cor    0.6555   0.05638   12    unc
## ## R!yf.cor    0.4375   0.0806    5.4    unc

## # Compare the estimates from the two methods
## plot(pred.rcb$pvals[,2], pred.sp$pvals[,2], xlim=c(13,37), ylim=c(13,37),
##       xlab="RCB", ylab="AR1xAR1", type='n')

```

```
## title("stroup.nin: Comparison of predicted values")
## text(pred.rcb$pvals[,2],pred.sp$pvals[,2],
##       as.character(pred.rcb$pvals[,1]),cex=0.5)
## abline(0,1)
```

```
## End(Not run)
```

```
# -----
```

stroup.splitplot *Simulated split-plot experiment*

Description

A simulated dataset of a very simple split-plot experiment, used to illustrate the details of calculating predictable functions (broad space, narrow space, etc.).

For example, the density of narrow, intermediate and broad-space predictable function for factor level A1 is shown below (html help only)

Format

y simulated response
 rep replicate, 4 levels
 b sub-plot, 2 levels
 a whole-plot, 3 levels

Source

Walter W. Stroup, 1989. Predictable functions and prediction space in the mixed model procedure. *Applications of Mixed Models in Agriculture and Related Disciplines*.

Used with permission of Walt Stroup.

References

Wolfinger, R.D. and Kass, R.E., 2000. Nonconjugate Bayesian analysis of variance component models, *Biometrics*, 56, 768–774. <http://doi.org/10.1111/j.0006-341X.2000.00768.x>

Examples

```
data(stroup.splitplot)
dat <- stroup.splitplot
```

```
# -----
```

```
## Not run:

# ---- lme4 -----
require(lme4)
m0 <- lmer(y~ -1 + a + b + a:b + (1|rep) + (1|a:rep), data=dat)
# No predict function

# ----- nlme -----
require(nlme)
m0 <- lme(y ~ -1 + a + b + a:b, data=dat, random = ~ 1|rep/a)

# ----- ASREML model -----
require(asreml)
m1 <- asreml(y~ -1 + a + b + a:b, random=~ rep + a:rep, data=dat)

library(lucid)
vc(m1) # Variance components match Stroup p. 41
##   effect component std.error z.ratio bound
##   rep      62.4      56.54      1.1    P    0
##   a:rep     15.38     11.79      1.3    P    0
##   units(R)   9.361     4.413     2.1    P    0

# Narrow space predictions
predict(m1, data=dat, classify="a", average=list(rep=NULL))$predictions$pvals
# a Predicted Std Err   Status
# a1      32.88   1.082 Estimable
# a2      34.12   1.082 Estimable
# a3      25.75   1.082 Estimable

# Intermediate space predictions
predict(m1, data=dat, classify="a", ignore=list("a:rep"),
        average=list(rep=NULL))$predictions$pvals
# a Predicted Std Err   Status
# a1      32.88   2.24 Estimable
# a2      34.12   2.24 Estimable
# a3      25.75   2.24 Estimable

# Broad space predictions
predict(m1, data=dat, classify="a")$predictions$pvals
# a Predicted Std Err   Status
# a1      32.88   4.54 Estimable
# a2      34.12   4.54 Estimable
# a3      25.75   4.54 Estimable

# ----- Mcmcglmm model -----
# Use the point estimates from REML with a prior distribution
require(MCMCglmm)
prior2 = list(
  G = list(G1=list(V=62.40, nu=1), G2=list(V=15.38, nu=1)),
  R = list(V = 9.4, nu=1)
)
m2 <- MCMCglmm(y~ -1 + a + b + a:b,
```

```

random=~ rep + a:rep, data=dat,
pr=TRUE, # save random effects as columns of 'Sol'
nitt=23000, # double the default 13000
prior=prior2, verbose=FALSE)

# Now create a matrix of coefficients for the prediction.
# Each column is for a different prediction. For example,
# the values in the column called 'a1a2n' are multiplied times
# the model coefficients (identified at the right side) to create
# the linear contrast for the the narrow-space predictions
# (also called adjusted mean) for the a1:a2 interaction.
#
#      a1n  a1i  a1b  a1a2n  a1a2ib
cm <- matrix(c( 1,  1,  1,  1,  1,  # a1
               0,  0,  0, -1, -1,  # a2
               0,  0,  0,  0,  0,  # a3
               1/2, 1/2, 1/2,  0,  0,  # b2
               0,  0,  0, -1/2, -1/2, # a2:b2
               0,  0,  0,  0,  0,  # a3:b2
               1/4, 1/4,  0,  0,  0,  # r1
               1/4, 1/4,  0,  0,  0,  # r2
               1/4, 1/4,  0,  0,  0,  # r3
               1/4, 1/4,  0,  0,  0,  # r4
               1/4,  0,  0,  1/4,  0,  # a1r1
               0,  0,  0, -1/4,  0,  # a2r1
               0,  0,  0,  0,  0,  # a3r1
               1/4,  0,  0,  1/4,  0,  # a1r2
               0,  0,  0, -1/4,  0,  # a2r2
               0,  0,  0,  0,  0,  # a3r2
               1/4,  0,  0,  1/4,  0,  # a1r3
               0,  0,  0, -1/4,  0,  # a2r3
               0,  0,  0,  0,  0,  # a3r3
               1/4,  0,  0,  1/4,  0,  # a1r4
               0,  0,  0, -1/4,  0,  # a2r4
               0,  0,  0,  0,  0), # a3r4
            ncol=5, byrow=TRUE)
rownames(cm) <- c("a1", "a2", "a3", "b2", "a2:b2", "a3:b2",
                "r1", "r2", "r3", "r4",
                "a1r1", "a1r2", "a1r3", "a1r4", "a2r1", "a2r2",
                "a2r3", "a2r4", "a3r1", "a3r2", "a3r3", "a3r4")
colnames(cm) <- c("A1n", "A1i", "A1b", "A1-A2n", "A1-A2ib")
print(cm)
# post2 <- as.mcmc(m2$Sol)
post2 <- as.mcmc(crossprod(t(m2$Sol), cm))

# Following table has columns for A1 estimate (narrow, intermediate, broad)
# A1-A2 estimate (narrow and intermediat/broad).
# The REML estimates are from Stroup 1989.
est <- rbind("REML est"=c(32.88, 32.88, 32.88, -1.25, -1.25),
            "REML stderr"=c(1.08, 2.24, 4.54, 1.53, 3.17),
            "MCMC mode"=posterior.mode(post2),
            "MCMC stderr"=apply(post2, 2, sd))
round(est,2)
#
#      A1n  A1i  A1b  A1-A2n  A1-A2ib

```

```

# REML est    32.88 32.88 32.88 -1.25 -1.25
# REML stderr 1.08  2.24  4.54  1.53  3.17
# MCMC mode   32.95 32.38 31.96 -1.07 -1.17
# MCMC stderr 1.23  2.64  5.93  1.72  3.73

post22 <- make.groups(Narrow=post2[,1], Intermediate=post2[,2],
                      Broad=post2[,3])
print(densityplot(~data|which, data=post22, groups=which,
                  cex=.25, lty=1, layout=c(1,3),
                  xlab="MCMC model value of predictable function for A1"))

## End(Not run)

```

student.barley

Multi-environment trial of barley

Description

Yield for two varieties of barley grown at 51 locations in the years 1901 to 1906.

Format

A data frame with 102 observations on the following 7 variables.

year year, 1901-1906

farmer farmer name

place place (nearest town)

district district, geographical area

gen genotype, Archer and Goldthorpe

yield yield, 'stones' per acre (1 stone = 14 pounds)

income income per acre in shillings, based on yield and quality

Details

Experiments were conducted for six years by the Department of Agriculture in Ireland. A total of seven varieties were tested, but only Archer and Goldthorpe were tested in all six years (others were dropped after being found inferior, or were added later). Plots were two acres in size. The value of the grain depended on the yield and quality. Quality varied much from farm to farm, but not so much within the same farm.

The phrase "analysis of variance" first appears in the abstract (only) of a 1918 paper by Fisher. The 1923 paper by Student contained the first analysis of variance table (but not for this data).

One stone is 14 pounds. To convert lb/ac to tonnes/ha, multiply by 0.00112085116

Note: The analysis of Student cannot be reproduced exactly. For example, Student states that the maximum income of Goldthorpe is 230 shillings. A quick glance at Table I of Student shows that the maximum income for Goldthorpe is 220 shillings (11 pounds, 0 shillings) in 1901 at Thurles. Also, the results of Kempton could not be reproduced exactly, perhaps due to rounding or the conversion factor that was used.

Source

Student. 1923. On Testing Varieties of Cereals. *Biometrika*, 15, 271–293. <http://doi.org/10.1093/biomet/15.3-4.271>

References

R A Kempton and P N Fox, 1997. *Statistical Methods for Plant Variety Evaluation*.

Examples

```

data(student.barley)
dat <- student.barley

require(lattice)
bwplot(yield ~ gen|district, dat, main="student.barley - yield")

dat$year <- factor(dat$year)
dat$income <- NULL

# convert to tons/ha
dat <- transform(dat, yield=yield*14 * 0.00112085116)

# Define 'loc' the way that Kempton does
dat$loc <- rep("",nrow(dat))
dat[is.element(dat$farmer, c("Allardyce","Roche","Quinn")), "loc"] <- "1"
dat[is.element(dat$farmer, c("Luttrell","Dooley")), "loc"] <- "2"
dat[is.element(dat$year, c("1904","1905","1906")) & dat$farmer=="Kearney", "loc"] <- "2"
dat[dat$farmer=="Mulhall", "loc"] <- "3"

dat <- transform(dat, loc=factor(paste(place,loc,sep="")))

if(require(reshape2)){
  datm <- melt(dat, measure.var='yield')

# Kempton Table 9.5
  round(acast(datm, loc+gen~year),2)

# Kempton Table 9.6
  d2 <- dcast(datm, year+loc~gen)
  mean(d2$Archer)
  mean(d2$Goldthorpe)
  mean(d2$Archer-d2$Goldthorpe)
  sqrt(var(d2$Archer-d2$Goldthorpe)/51)
  cor(d2$Archer,d2$Goldthorpe)
}

## Not run:
# Kempton Table 9.6b
require(lme4)
m2 <- lmer(yield~1 + (1|loc) + (1|year) +

```

```
(1|loc:year) + (1|gen:loc) + (1|gen:year), data=dat,  
control=lmerControl(check.nobs.vs.rankZ="ignore")  
  
## End(Not run)
```

tai.potato

Multi-environment trial of potato

Description

Multi-environment trial of potato tuber yields

Usage

```
data("tai.potato")
```

Format

A data frame with 48 observations on the following 6 variables.

yield yield, kg/plot

gen genotype code

variety variety name

env environment code

loc location

year year

Details

Mean tuber yield of 8 genotypes in 3 locations over two years. Katahdin and Sebago are check varieties. Each location was planted as a 4-rep RCB design.

In Tai's plot of the stability parameters, F5751 and Sebago were in the average stability area. The highest yielding genotype F6032 had an unstable performance.

Source

G.C.C. Tai, 1971. Genotypic stability analysis and its application to potato regional trials. *Crop Sci* 11, 184-190. Table 2, p. 187. <http://doi.org/10.2135/cropsci1971.0011183X001100020006x>

References

George Fernandez (1991). Analysis of Genotype x Environment Interaction by Stability Estimates. *Hort Science*, 26, 947-950.

Examples

```

data(tai.potato)
dat = tai.potato

require(lattice)
dotplot(variety ~ yield|env, dat, main="tai.potato")

# fixme - need to add tai() example

# note, st4gi::tai assumes there are replications in the data
# https://github.com/reyzaguirre/st4gi/blob/master/R/tai.R

```

| | |
|---------------|---|
| talbot.potato | <i>Multi-environment trial of potato in UK, yields and trait scores at 12 locations</i> |
|---------------|---|

Description

Yield and 14 trait scores for each of 9 potato varieties at 12 locations in UK.

Usage

```

data("talbot.potato.traits")
data("talbot.potato.yield")

```

Format

The talbot.potato.yield dataframe has 126 observations on the following 3 variables.

```

gen genotype/variety
trait trait
score trait score, 1-9

```

The talbot.potato.yield dataframe has 108 observations on the following 3 variables.

```

gen genotype/variety
loc location/center
yield yield, t/ha

```

Details

The talbot.potato.yield dataframe contains mean tuber yields (t/ha) of 9 varieties of potato at 12 centers in the United Kingdom over five years 1983-1987. The following abbreviations are used for the centers.

```

BU   Bush
CA   Cambridge

```


| | |
|----|-------------------|
| CB | Conon Bridge |
| CC | Crossacreevy |
| CP | Cockle Park |
| CR | Craibstone |
| GR | Greenmount |
| HA | Harper Adams |
| MO | Morley |
| RO | Rosemaund |
| SB | Sutton Bonnington |
| TE | Terrington |

Source

Mike Talbot and A V Wheelwright, 1989, The analysis of genotype x analysis interactions by partial least squares regression. *Biuletyn Oceny Odmian*, 21/22, 19–25.

Used with permission of Mike Talbot.

Examples

```
if(require(pls) & require(reshape2)){  
  
  data(talbot.potato.traits)  
  datt <- talbot.potato.traits  
  data(talbot.potato.yield)  
  daty <- talbot.potato.yield  
  
  datt <- acast(datt, gen ~ trait, value.var='score')  
  daty <- acast(daty, gen ~ loc, value.var='yield')  
  
  # Transform columns to zero mean and unit variance  
  datt <- scale(datt)  
  daty <- scale(daty)  
  
  m1 <- pls(daty ~ datt, ncomp=3)  
  summary(m1)  
  
  # Loadings factor 1  
  lo <- loadings(m1)[,1,drop=FALSE]  
  round(-1*lo[order(-1*lo),1,drop=FALSE],2)  
  
  biplot(m1, main="talbot.potato - biplot")  
  
}
```

 theobald.barley

Multi-environment trial of barley, multiple years & fertilizer levels

Description

Barley yields at multiple locs, years, fertilizer levels

Usage

```
data("theobald.barley")
```

Format

A data frame with 105 observations on the following 5 variables.

yield yield, tonnes/ha

gen genotype

loc location, 5 levels

nitro nitrogen kg/ha

year year, 2 levels

Details

Theobald and Talbot used BUGS to fit a fully Bayesian model for yield response curves.

Locations of the experiment were in north-east Scotland.

Assumed nitrogen cost 400 pounds per tonne. Grain prices used were 100, 110, and 107.50 pounds per tonne for Georgie, Midas and Sundance.

Source

Chris M. Theobald and Mike Talbot, (2002). The Bayesian choice of crop variety and fertilizer dose. *Appl Statistics*, 51, 23-36. <http://doi.org/10.1111/1467-9876.04863>

Data provided by Chris Theobald and Mike Talbot.

Examples

```
data(theobald.barley)
dat <- theobald.barley
dat <- transform(dat, env=paste(loc,year,sep="-"))
dat <- transform(dat, income=100*yield - 400*nitro/1000)

require(lattice)
xyplot(income~nitro|env, dat, groups=gen, type='b',
       auto.key=list(columns=3), main="theobald.barley")
```

| | |
|--------------------|--|
| theobald.covariate | <i>Multi-environment trial of corn silage, Year * Loc * Variety with covariate</i> |
|--------------------|--|

Description

Corn silage yields for maize in 5 years at 7 districts for 10 hybrids.

Format

A data frame with 256 observations on the following 5 variables.

year year, 1990-1994

env environment/district, 1-7

gen genotype, 1-10

yield dry-matter silage yield for corn

chu corn heat units, thousand degrees Celsius

Details

The trials were carried out in seven districts in the maritime provinces of Eastern Canada. Different fields were used in successive years. The covariate CHU (Corn Heat Units) is the accumulated average daily temperatures (thousands of degrees Celsius) during the growing season at each location.

Used with permission of Chris Theobald.

Source

Chris M. Theobald and Mike Talbot and Fabian Nabugoomu, 2002. A Bayesian Approach to Regional and Local-Area Prediction From Crop Variety Trials, *Journ Agric Biol Env Sciences*, 7, 403–419. <http://doi.org/10.1198/108571102230>

Examples

```
data(theobald.covariate)
dat <- theobald.covariate
require(lattice)
xyplot(yield ~ chu|gen, dat, type=c('p','smooth'),
       xlab = "chu = corn heat units",
       main="theobald.covariate - yield vs heat")

## Not run:
# REML estimates (Means) in table 3 of Theobald 2002
require(lme4)
dat <- transform(dat, year=factor(year))
m0 <- lmer(yield ~ -1 + gen + (1|year/env) + (1|gen:year), data=dat)
round(fixef(m0),2)
```

```

## End(Not run)

## Not run:

# Use JAGS to fit Theobald (2002) model 3.2 with 'Expert' prior

require(reshape2)
ymat <- acast(dat, year+env~gen, value.var='yield')
chu <- acast(dat, year+env~., mean, value.var='chu', na.rm=TRUE)
chu <- as.vector(chu - mean(chu)) # Center the covariate
dat$yr <- as.numeric(dat$year)
yridx <- as.vector(acast(dat, year+env~., mean, value.var='yr', na.rm=TRUE))
dat$loc <- as.numeric(dat$env)
locidx <- acast(dat, year+env~., mean, value.var='loc', na.rm=TRUE)
locidx <- as.vector(locidx)

jdat <- list(nVar = 10, nYear = 5, nLoc = 7, nYL = 29, yield = ymat,
            chu = chu, year = yridx, loc = locidx)

require(rjags)
m1 <- jags.model(file=system.file(package="agridat", "files/theobald.covariate.jag"),
                 data=jdat, n.chains=2)

# Table 3, Variety deviations from means (Expert prior)
c1 <- coda.samples(m1, variable.names=(c('alpha')),
                  n.iter=10000, thin=10)
s1 <- summary(c1)
effs <- s1$statistics[, 'Mean']
rev(sort(round(effs - mean(effs), 2))) # Perfect match (different order?)

## End(Not run)

```

thompson.cornsoy

Multi-environment trial of corn & soybean, 1930-1962, with temperature and precipitation

Description

Average yield of corn and soybeans in five U.S. states (IA, IL, IN, MO, OH) during the years 1930-1962. Pre-season precipitation and average temperature and precipitation during each month of the growing season is included.

Format

```

state state
year year, 1930-1962
rain0 pre-season precipitation in inches

```

temp5 may temperature, Fahrenheit
 rain6 june rain, inches
 temp6 june temp
 rain7 july rain
 temp7 july temp
 rain8 august rain
 temp8 august temp
 corn corn yield, bu/acre
 soy soybean yield, bu/acre

Details

Note: The Iowa corn data has sometimes been identified (in other sources) as the "Iowa wheat" data, but this is incorrect.

The 'year' variable affects yield through (1) improvements in plant genetics (2) changes in management techniques such as fertilizer, chemicals, tillage, planting date, and (3) climate, pest infestations, etc.

Double-cross corn hybrids were introduced in the 1920s. Single-cross hybrids became common around 1960.

During World War II, nitrogen was used in the production of TNT for bombs. After the war, these factories switched to producing ammonia for fertilizer. Nitrogen fertilizer use greatly increased after WWII and is a major reason for yield gains of corn. Soybeans gain little benefit from nitrogen fertilizer. The other major reason for increasing yields in both crops is due to improved plant genetics.

Crops are often planted in May, and harvest begins in September.

Yields in 1936 were very low due to July being one of the hottest and driest on record.

Some relevant maps of yield, heat, and precipitation can be found in *Atlas of crop yield and summer weather patterns, 1931-1975*, <http://www.isws.illinois.edu/pubdoc/C/ISWSC-150.pdf>

The following notes pertain to the Iowa data.

The 1947 June precipitation of 10.33 inches was the wettest June on record (a new Iowa June record of 10.34 inches was set in 2010). As quoted in *Monthly Weather Review* (Dec 1957, p. 396) "The dependence of Iowa agriculture upon the vagaries of the weather was closely demonstrated during the 1947 season. A cool wet spring delayed crop planting activity and plant growth; then, in addition, a hard freeze on May 29th ... further set back the corn. The heavy rains and subsequent floods during June caused appreciable crop acreage to be abandoned ... followed by a hot dry weather regime that persisted from mid-July through the first week of September."

In 1949 soybean yields were average while corn yields were low. From the same source above, "The year 1949 saw the greatest infestation of corn borer in the history of corn in Iowa".

1955 yields were reduced due to dry weather in late July and August.

Source

Thompson, L.M., 1963. *Weather and technology in the production of corn and soybeans*. CAED Report 17. The Center for Agriculture and Economic Development, Iowa State University, Ames, Iowa.

References

Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis*, second ed., Wiley, New York.

Examples

```

data(thompson.cornsoy)
dat <- thompson.cornsoy

# The droughts of 1934/36 were severe in IA/MO. Less so in OH.
require(lattice)
xyplot(corn+soy~year|state, dat,
       type=c('p','l','r'), auto.key=list(columns=2),
       main="thompson.cornsoy",
       layout=c(5,1),ylab='yield')

# In 1954, only Missouri suffered very hot, dry weather
## xyplot(corn~year, dat,
##       groups=state, type=c('p','l'),
##       main="thompson.cornsoy",
##       auto.key=list(columns=5), ylab='corn yield')

# Rain and temperature have negative correlation in each month.
# July is a critical month: temp and yield are negatively correlated,
# while rain and yield are positively correlated.
# splom(~dat[-1,-1], col=dat$state, cex=.5, main="thompson.cornsoy")

## Not run:
# Plots similar to those in Venables' Exegeses paper.

dat.ia <- subset(dat, state=="Iowa")

require(splines)
m2 <- aov(corn ~ ns(rain0, 3) + ns(rain7, 3) +
         ns(temp8, 3) + ns(year,3), dat.ia)
op <- par(mfrow=c(2,2))
termplot(m2, se=TRUE, rug=TRUE, partial=TRUE)
par(op)

# do NOT use gam package
require(mgcv)
m1 <- gam(corn ~ s(year, k=5) + s(rain0, k=5) +
         s(rain7, k=5) + s(temp8, k=5), data=dat.ia)
op <- par(mfrow=c(2,2))
plot.gam(m1, residuals=TRUE, se=TRUE, cex=2)
par(op)

## End(Not run)

```

turner.herbicide *Herbicide control of larkspur*

Description

Herbicide control of larkspur

Usage

```
data("turner.herbicide")
```

Format

A data frame with 12 observations on the following 4 variables.

rep rep factor

rate rate of herbicide

live number of live plants before application

dead number of plants killed by herbicide

Details

Effectiveness of the herbicide Picloram on larkspur plants at 4 doses (0, 1.1, 2.2, 4.5) in 3 reps. Experiment was done in 1986 at Manti, Utah.

Source

David L. Turner and Michael H. Ralphs and John O. Evans (1992). Logistic Analysis for Monitoring and Assessing Herbicide Efficacy. *Weed Technology*, 6, 424-430. <http://www.jstor.org/stable/3987312>

References

Christopher Bilder, Thomas Loughin. *Analysis of Categorical Data with R*.

Examples

```
data(turner.herbicide)
dat <- turner.herbicide

dat <- transform(dat, prop=dead/live)
# xyplot(prop~rate,dat, pch=20, main="turner.herbicide", ylab="Proportion killed")

m1 <- glm(prop~rate, data=dat, weights=live, family=binomial)
coef(m1) # -3.46, 2.6567 Same as Turner eqn 3

# Make conf int on link scale and back-transform
p1 <- expand.grid(rate=seq(0,to=5,length=50))
```

```

p1 <- cbind(p1, predict(m1, newdata=p1, type='link', se.fit=TRUE))
p1 <- transform(p1, lo = plogis(fit - 2*se.fit),
                fit = plogis(fit),
                up = plogis(fit + 2*se.fit))

# Figure 2 of Turner
if(require(latticeExtra)){
  foo1 <- xyplot(prop~rate,dat, cex=1.5,
                main="turner.herbicide (model with 2*S.E.)",
                xlab="Herbicide rate", ylab="Proportion killed")
  foo2 <- xyplot(fit~rate, p1, type='l')
  foo3 <- xyplot(lo+up~rate, p1, type='l', lty=1, col='gray')
  print(foo1 + foo2 + foo3)
}

# What dose gives a LD90 percent kill rate?
# require(MASS)
# dose.p(m1, p=.9)
##           Dose           SE
## p = 0.9: 2.12939 0.128418

# Alternative method
# require(car) # logit(.9) = 2.197225
# deltaMethod(m1, g="(log(.9/(1-.9))-b0)/(b1)", parameterNames=c('b0','b1'))
##           Estimate           SE
## (2.197225 - b0)/(b1) 2.12939 0.128418

# What is a 95 percent conf interval for LD90? Bilder & Loughin page 138
root <- function(x, prob=.9, alpha=0.05){
  co <- coef(m1) # b0,b1
  covs <- vcov(m1) # b00,b11,b01
  # .95 = b0 + b1*x
  # (b0+b1*x) + Z(alpha/2) * sqrt(b00 + x^2*b11 + 2*x*b01) > .95
  # (b0+b1*x) - Z(alpha/2) * sqrt(b00 + x^2*b11 + 2*x*b01) < .95
  f <- abs(co[1] + co[2]*x - log(prob/(1-prob))) /
    sqrt(covs[1,1] + x^2 * covs[2,2] + 2*x*covs[1,2])
  return( f - qnorm(1-alpha/2))
}
lower <- uniroot(f=root, c(0,2.13))
upper <- uniroot(f=root, c(2.12, 5))
c(lower$root, upper$root)
# 1.92 2.45

```

urquhart.feedlot

Weight gain calves in a feedlot

Description

Weight gain calves in a feedlot, given three different diets.

Usage

```
data("urquhart.feedlot")
```

Format

A data frame with 67 observations on the following 5 variables.

animal animal ID

herd herd ID

diet diet: Low, Medium, High

weight1 initial weight

weight2 slaughter weight

Details

Calves born in 1975 in 11 different herds entered a feedlot as yearlings. Each animal was fed one of three diets with low, medium, or high energy. The original sources explored the use of some contrasts for comparing breeds.

| Herd | Breed |
|------|---------------------------------|
| 9 | New Mexico Herefords |
| 16 | New Mexico Herefords |
| 3 | Utah State University Herefords |
| 32 | Angus |
| 24 | Angus x Hereford (cross) |
| 31 | Charolais x Hereford |
| 19 | Charolais x Hereford |
| 36 | Charolais x Hereford |
| 34 | Brangus |
| 35 | Brangus |
| 33 | Southern Select |

Source

N. Scott Urquhart (1982). Adjustment in Covariance when One Factor Affects the Covariate Biometrics, 38, 651-660. Table 4, p. 659. <http://doi.org/10.2307/2530046>

References

N. Scott Urquhart and David L. Weeks (1978). Linear Models in Messy Data: Some Problems and Alternatives Biometrics, 34, 696-705. <http://doi.org/10.2307/2530391>

Also available in the 'lsmeans' package as the 'feedlot' data.

Examples

```
data(urquhart.feedlot)
```

```

dat <- urquhart.feedlot

if(require(reshape2) & require(latticeExtra)){
  d2 <- melt(dat, id.vars=c('animal','herd','diet'))

  useOuterStrips(xyplot(value ~ variable|diet*herd, data=d2, group=animal,
                        type='l',
                        xlab="Initial & slaughter timepoint for each diet",
                        ylab="Weight for each herd",
                        main="urquhart.feedlot - weight gain by animal"))
}

# simple fixed-effects model
dat <- transform(dat, animal = factor(animal), herd=factor(herd))
m1 <- lm(weight2 ~ weight1 + herd*diet, data = dat)
coef(m1) # weight1 = 1.1373 match Urquhart table 5 common slope

# random-effects model might be better, for example
# require(lme4)
# m1 <- lmer(weight2 ~ -1 + diet + weight1 + (1|herd), data=dat)
# summary(m1) # weight1 = 1.2269

```

usgs.herbicides

Concentrations of herbicides in streams in the United States

Description

Concentrations of selected herbicides and degradation products determined by laboratory method analysis code GCS for water samples collected from 51 streams in nine Midwestern States,2002

Usage

```
data("usgs.herbicides")
```

Format

A data frame with 184 observations on the following 19 variables.

```

mapnum map number
usgsid USGS ID
long longitude
lat latitude
site site name
city city
sampltype sample type code
date date sample was collected

```

hour hour sample was collected
 acetochlor concentration as character
 alachlor concentration as character
 ametryn concentration as character
 atrazine concentration as character
 CIAT concentration as character
 CEAT concentration as character
 cyanazine concentration as character
 CAM concentration as character
 dimethenamid concentration as character
 flufenacet concentration as character

Details

Concentrations of selected herbicides and degradation products determined by laboratory method analysis code GCS for water samples collected from 51 streams in nine Midwestern States, 2002.

All concentrations are micrograms/liter, "<" means "less than". The data are in character format to allow for "<".

The original report contains data for more herbicides. This data is for illustrative purposes.

Sample types: CR = concurrent replicate sample, FB = field blank, LD = laboratory duplicate, S1 = sample from pre-emergence runoff, S2 = sample from post-emergence runoff, S3 = sample from harvest-season runoff.

Source

Scribner, E.A., Battaglin, W.A., Dietze, J.E., and Thurman, E.M., "Reconnaissance Data for Glyphosate, Other Selected Herbicides, their Degradation Products, and Antibiotics in 51 streams in Nine Midwestern States, 2002". U.S. Geological Survey Open File Report 03-217. Herbicide data from table 5, page 30-37. Site coordinates page 7-8. <https://ks.water.usgs.gov/pubs/reports/ofr.03-217.html>

References

None.

Examples

```
data(usgs.herbicides)
dat <- usgs.herbicides

## Not run:

require("NADA")
# create censored data for one trait
dat$y <- as.numeric(dat$atrazine)
dat$ycen <- is.na(dat$y)
dat$y[is.na(dat$y)] <- .05
```

```

# percent censored
with(dat, censummary(y, censored=ycen))
# median/mean
with(dat, cenmle(y, ycen, dist="lognormal"))
# boxplot
with(dat, cenboxplot(obs=y, cen=ycen, log=FALSE))
# with(dat, boxplot(y))
pp <- with(dat, ros(obs=y, censored=ycen, forwardT="log")) # default lognormal
plot(pp)

## End(Not run)

if(require("NADA")){
  plotfun <- function(vv){
    dat$y <- as.numeric(dat[[vv]])
    dat$ycen <- is.na(dat$y)
    dat$y[is.na(dat$y)] <- .01
    # qqnorm(log(dat$y), main=vv) # ordinary qq plot shows censored values
    pp <- with(dat, ros(obs=y, censored=ycen, forwardT="log"))
    plot(pp, main=vv) # omits censored values
  }
  op <- par(mfrow=c(3,3))
  vnames <- c("acetochlor", "alachlor", "ametryn", "atrazine", "CIAT", "CEAT", "cyanazine", "#CAM",
             "dimethenamid", "flufenacet")
  for(vv in vnames) plotfun(vv)
  par(op)
}

```

vaneeuwijk.drymatter *Dry matter content for Dutch maize varieties in multiple years/sites*

Description

Dry matter content for Dutch maize varieties in multiple years/sites

Usage

```
data("vaneeuwijk.drymatter")
```

Format

A data frame with 168 observations on the following 5 variables.

year year

site site, 4 levels

variety variety, 6 levels

y dry matter percent

Details

Percent dry matter is given.

Site codes are soil type classifications: SS=Southern Sand, CS=Central Sand, NS=Northern Sand, RC=River Clay.

These data are a balanced subset of the data analyzed in van Eeuwijk, Keizer, and Bakker (1995b) and Kroonenberg, Basford, and Ebskamp (1995).

Source

van Eeuwijk, Fred A. and Pieter M. Kroonenberg (1998). Multiplicative Models for Interaction in Three-Way ANOVA, with Applications to Plant Breeding *Biometrics*, 54, 1315-1333. <http://doi.org/10.2307/2533660>

Used with permission of Fred van Eeuwijk.

References

Kroonenberg, P.M., Basford, K.E. & Ebskamp, A.G.M. (1995). Three-way cluster and component analysis of maize variety trials. *Euphytica*, 84(1):31-42. <http://doi.org/10.1007/BF01677554>

van Eeuwijk, F.A., Keizer, L.C.P. & Bakker, J.J. Van Eeuwijk. (1995b). Linear and bilinear models for the analysis of multi-environment trials: II. An application to data from the Dutch Maize Variety Trials *Euphytica*, 84(1):9-22. <http://doi.org/10.1007/BF01677552>

Hardeo Sahai, Mario M. Ojeda. *Analysis of Variance for Random Models*, Volume 1. Page 261.

Examples

```
data(vaneeuwijk.drymatter)
dat <- vaneeuwijk.drymatter
dat <- transform(dat, year=factor(year))
dat <- transform(dat, env=factor(paste(year,site)))

if(require(HH)){
  HH::interaction2wt(y ~ year+site+variety,dat,rot=c(90,0),
                    x.between=0, y.between=0,
                    main="vaneeuwijk.drymatter")
}

## Not run:
# anova model
m1 <- aov(y ~ variety+env+variety:env, data=dat)
anova(m1) # Similar to VanEeuwijk table 2
m2 <- aov(y ~ year*site*variety, data=dat)
anova(m2) # matches Sahai table 5.5

# variance components model
require(lme4)
require(lucid)
m3 <- lmer(y ~ (1|year) + (1|site) + (1|variety) +
           (1|year:site) + (1|year:variety) + (1|site:variety),
           data=dat)
```

```
vc(m3) # matches Sahai page 266
##      grp      var1 var2   vcov  sdcov
## year:variety (Intercept) <NA> 0.3187 0.5645
##   year:site (Intercept) <NA> 7.735  2.781
## site:variety (Intercept) <NA> 0.03502 0.1871
##      year (Intercept) <NA> 6.272  2.504
##      variety (Intercept) <NA> 0.4867 0.6976
##      site (Intercept) <NA> 6.504  2.55
##      Residual      <NA> <NA> 0.8885 0.9426

## End(Not run)
```

vaneeuwijk.fusarium *Infection of wheat varieties by Fusarium strains from 1990 to 1993*

Description

Infection of wheat varieties by Fusarium strains from 1990 to 1993

Usage

```
data("vaneeuwijk.fusarium")
```

Format

A data frame with 560 observations on the following 4 variables.

```
year year, 1990-1993
strain strain of fusarium
gen genotype/variety
y
```

Details

Data come from Hungary. There were 20 wheat varieties infected with 7 strains of Fusarium in the years 1990-1993. The measured value is a rating of the severity of disease due to Fusarium head blight, expressed as a number 1-100.

Three-way interactions for varieties 21 and 23 were the only ones in 1992 suffering from strain infections. This was due to incorrect storage of the inoculum (strain) which rendered it incapable of infecting most other varieties.

The data is a subset of the data analyzed by VanEeuwijk et al. 1995.

Source

van Eeuwijk, Fred A. and Pieter M. Kroonenberg (1998). Multiplicative Models for Interaction in Three-Way ANOVA, with Applications to Plant Breeding *Biometrics*, 54, 1315-1333. <http://doi.org/10.2307/2533660>
Used with permission of Fred van Eeuwijk.

References

F. A. van Eeuwijk, A. Mesterhazy, Ch. I. Kling, P. Ruckenbauer, L. Saur, H. Burstmayr, M. Lemmens, L. C. P. Keizer, N. Maurin, C. H. A. Snijders. (1995). Assessing non-specificity of resistance in wheat to head blight caused by inoculation with European strains of *Fusarium culmorum*, *F. graminearum* and *F. nivale* using a multiplicative model for interaction. *Theor Appl Genet.* 90(2), 221-8. <http://doi.org/10.1007/BF00222205>

Examples

```
data(vaneeuwijk.fusarium)
dat <- vaneeuwijk.fusarium
dat <- transform(dat, year=factor(year))
dat <- transform(dat, logity=log((y/100)/(1-y/100)))

if(require(HH)){
  position(dat$year) <- c(3,9,14,19)
  position(dat$strain) <- c(2,5,8,11,14,17,20)
  HH::interaction2wt(logity ~ gen+year+strain,dat,rot=c(90,0),
                    x.between=0, y.between=0,
                    main="vaneeuwijk.fusarium")
}

## Not run:
# anova on logit scale. Near match to VanEeuwijk table 6
m1 <- aov(logity ~ gen*strain*year, data=dat)
anova(m1)
## Response: logity
##           Df Sum Sq Mean Sq F value Pr(>F)
## gen           19 157.55   8.292
## strain         6  91.54  15.256
## year           3 321.99 107.331
## gen:strain    114  34.03   0.299
## gen:year      57 140.94   2.473
## strain:year   18 236.95  13.164
## gen:strain:year 342  93.15   0.272

## End(Not run)
```

vaneeuwijk.nematodes *Number of cysts on 11 potato genotypes for 5 potato cyst nematode populations.*

Description

The number of cysts on 11 potato genotypes for 5 potato cyst nematode populations.

Usage

```
data("vaneeuwijk.nematodes")
```

Format

A data frame with 55 observations on the following 3 variables.

```
gen potato genotype
pop nematode population
y number of cysts
```

Details

The number of cysts on 11 potato genotypes for 5 potato cyst nematode populations belonging to the species *Globodera pallida*. This is part of a larger table in . The numbers are the means over four or five replicates.

Van Eeuwijk used this data to illustrate fitting a generalized linear model.

Source

Fred A. van Eeuwijk, (1995). Multiplicative Interaction in Generalized Linear Models. *Biometrics*, 51, 1017-1032. <http://doi.org/10.2307/2533001>

References

Arntzen, F.K. & van Eeuwijk (1992). Variation in resistance level of potato genotypes and virulence level of potato cyst nematode populations. *Euphytica*, 62, 135-143. <http://doi.org/10.1007/BF00037939>

Examples

```
data(vaneeuwijk.nematodes)
dat <- vaneeuwijk.nematodes

# show non-normality
op <- par(mfrow=c(2,1), mar=c(5,4,3,2))
boxplot(y ~ pop, data=dat, las=2,
        ylab="number of cysts")
title("vaneeuwijk.nematodes - cysts per nematode pop")
boxplot(y ~ gen, data=dat, las=2)
title("vaneeuwijk.nematodes - cysts per potato")
par(op)

## Not run:
# normal distribution
lm1 <- lm(y ~ gen + pop, data=dat)

# poisson distribution
glm1 <- glm(y ~ gen+pop, data=dat, family=quasipoisson(link=log))
anova(glm1)
```



```

require(gnm)

# main-effects non-interaction model
gnm0 <- gnm(y ~ pop + gen, data=dat,
            family=quasipoisson(link=log))
# one interaction
gnm1 <- gnm(y ~ pop + gen + Mult(pop,gen,inst=1), data=dat,
            family=quasipoisson(link=log))
# two interactions
gnm2 <- gnm(y ~ pop + gen + Mult(pop,gen,inst=1) + Mult(pop,gen,inst=2),
            data=dat,
            family=quasipoisson(link=log))

# anova(gnm0, gnm1, gnm2, test="F")
# only 2, not 3 axes needed

# match vaneeuwijk table 2
anova(gnm2)
##
##          Df Deviance Resid. Df Resid. Dev
## NULL                    54      8947.4
## pop                   4      690.6     50      8256.8
## gen                   10     7111.4     40     1145.4
## Mult(pop, gen, inst = 1) 13      716.0     27      429.4
## Mult(pop, gen, inst = 2) 11      351.1     16       78.3

# compare residual qq plots from models
op <- par(mfrow=c(2,2))
plot(lm1, which=2, main="LM")
plot(glm1, which=2, main="GLM")
plot(gnm0, which=2, main="GNM, no interaction")
plot(gnm2, which=2, main="GNM, 2 interactions")
par(op)

# extract interaction-term coefficients, make a biplot
pops <- pickCoef(gnm2, "[.]pop")
gens <- pickCoef(gnm2, "[.]gen")
coefs <- coef(gnm2)
A <- matrix(coefs[pops], nc = 2)
B <- matrix(coefs[gens], nc = 2)
A2=scale(A)
B2=scale(B)
rownames(A2) <- levels(dat$pop)
rownames(B2) <- levels(dat$gen)
# near-match with vaneeuwijk figure 1
biplot(A2,B2, expand=2.5,xlim=c(-2,2),ylim=c(-2,2),
       main="vaneeuwijk.nematodes - GAMMI biplot")

## End(Not run)

```

Description

Treatment x environment interaction in agronomy trials

Usage

```
data("vargas.txe.covs")  
data("vargas.txe.yield")
```

Format

The 'vargas.txe.covs' data has 10 years of measurements on 28 environmental covariates:

year year
MTD mean maximum temperature in December
MTJ mean maximum temperature in January
MTF mean maximum temperature in February
MTM mean maximum temperature in March
MTA mean maximum temperature in April
mTD mean minimum temperature in December
mTJ mean minimum temperature in January
mTF mean minimum temperature in February
mTM mean minimum temperature in March
mTA mean minimum temperature in April
mTUD mean minimum temperature in December
mTUJ mean minimum temperature in January
mTUF mean minimum temperature in February
mTUM mean minimum temperature in March
mTUA mean minimum temperature in April
PRD total monthly precipitation in December
PRJ total monthly precipitation in Jan
PRF total monthly precipitation in Feb
PRM total monthly precipitation in Mar
SHD sun hours per day in Dec
SHJ sun hours per day in Jan
SHF sun hours per day in Feb
EVD total monthly evaporation in Dec
EVJ total monthly evaporation in Jan
EVF total monthly evaporation in Feb
EVM total monthly evaporation in Mar
EVA total monthly evaporation in Apr

The 'vargas.txe.yield' dataframe contains 240 observations on three variables

year Year

trt Treatment. See details section

yield Grain yield, kg/ha

Details

The treatment names indicate:

| | |
|---|--------------------|
| T | deep knife |
| t | no deep knife |
| S | sesbania |
| s | soybean |
| M | chicken manure |
| m | no chicken manure |
| 0 | no nitrogen |
| n | 100 kg/ha nitrogen |
| N | 200 kg/ha nitrogen |

Source

Vargas, Mateo and Crossa, Jose and van Eeuwijk, Fred and Sayre, Kenneth D. and Reynolds, Matthew P. (2001). Interpreting Treatment x Environment Interaction in Agronomy Trials. *Agron. J.*, 93, 949-960. Table A1, A3. <https://doi.org/10.2134/agronj2001.934949x>

Used with permission of Jose Crossa.

Examples

```
data(vargas.txe.covs)
data(vargas.txe.yield)

if(require(reshape2)){

  require(lattice)
  redblue <- colorRampPalette(c("firebrick", "lightgray", "#375997"))
  Z <- vargas.txe.yield
  Z <- acast(Z, year ~ trt, value.var='yield')
  levelplot(Z, col.regions=redblue,
            main="vargas.txe.yield", xlab="year", ylab="treatment",
            scales=list(x=list(rot=90)))

  # Double-centered like AMMI
  Z <- sweep(Z, 1, rowMeans(Z))
  Z <- sweep(Z, 2, colMeans(Z))
  # Vargas figure 1
  biplot(prcomp(Z, scale.=FALSE), main="vargas.txe.yield")
}
```

```

# Now, PLS relating the two matrices
U <- vargas.txe.covs
U <- scale(U) # Standardized covariates
require(pls)
m1 <- pls(Z~U)
# Vargas Fig 2, flipped vertical/horizontal
biplot(m1, which="x", var.axes=TRUE)

}

```

vargas.wheat1

Wheat yields in 7 years with genetic and environment covariates

Description

Yield of Durum wheat, 7 genotypes, 6 years, with 16 genotypic variates and 16 environment variates.

Usage

```

data("vargas.wheat1.covs")
data("vargas.wheat1.traits")

```

Format

The vargas.wheat1.covs dataframe has 6 observations on the following 17 variables.

year year, 1990-1995

MTD Mean daily max temperature December, deg C

MTJ January

MTF February

MTM March

mTD Mean daily minimum temperature December, deg C

mTJ January

mTF February

mTM March

PRD Monthly precipitation in December, mm

PRJ January

PRF February

PRM March

SHD a numeric vector

SHJ January

SHF February

SHM March

The vargas.wheat1.traits dataframe has 126 observations on the following 19 variables.

year year, 1990-1995
rep replicate, 3 levels
gen genotype, 7 levels
yield yield, kg/ha
ANT anthesis, days after emergence
MAT maturity, days after emergence
GFI grainfill, MAT-ANT
PLH plant height, cm
BIO biomass above ground, kg/ha
HID harvest index
STW straw yield, kg/ha
NSM spikes / m²
NGM grains / m²
NGS grains per spike
TKW thousand kernel weight, g
WTI weight per tiller, g
SGW spike grain weight, g
VGR vegetative growth rate, kg/ha/day, STW/ANT
KGR kernel growth rate, mg/kernel/day

Details

Conducted in Ciudad Obregon, Mexico.

Source

Mateo Vargas and Jose Crossa and Ken Sayre and Matthew Renolds and Martha E Ramirez and Mike Talbot, 1998. Interpreting Genotype x Environment Interaction in Wheat by Partial Least Squares Regression, *Crop Science*, 38, 679–689. <http://doi.org/10.2135/cropsci1998.0011183X003800030010x>

Data provided by Jose Crossa.

Examples

```
## Not run:  
  
data(vargas.wheat1.covs)  
data(vargas.wheat1.traits)  
  
require(pls)  
require(reshape2)
```

```

# Yield as a function of non-yield traits
Y0 <- vargas.wheat1.traits[,c('gen','rep','year','yield')]
Y0 <- acast(Y0, gen ~ year, value.var='yield', fun=mean)
Y0 <- sweep(Y0, 1, rowMeans(Y0))
Y0 <- sweep(Y0, 2, colMeans(Y0)) # GxE residuals
Y1 <- scale(Y0) # scaled columns
X1 <- vargas.wheat1.traits[, -4] # omit yield
X1 <- aggregate(cbind(ANT,MAT,GFI,PLH,BIO,HID,STW,NSM,NGM,
                     NGS,TKW,WTI,SGW,VGR,KGR) ~ gen, data=X1, FUN=mean)
rownames(X1) <- X1$gen
X1$gen <- NULL
X1 <- scale(X1) # scaled columns
m1 <- pls(Y1~X1)
loadings(m1)[,1,drop=FALSE] # X loadings in Table 1 of Vargas

biplot(m1, cex=.5, which="x", var.axes=TRUE,
       main="vargas.wheat1 - gen ~ trait") # Vargas figure 2a

# Yield as a function of environment covariates
Y2 <- t(Y0)
X2 <- vargas.wheat1.covs
rownames(X2) <- X2$year
X2$year <- NULL
Y2 <- scale(Y2)
X2 <- scale(X2)

m2 <- pls(Y2~X2)
loadings(m2)[,1,drop=FALSE] # X loadings in Table 2 of Vargas

## End(Not run)

```

vargas.wheat2

Multi-environment trial of wheat with environmental covariates

Description

The yield of 8 wheat genotypes was measured in 21 low-humidity environments. Each environment had 13 covariates recorded.

Usage

```

data("vargas.wheat2.covs")
data("vargas.wheat2.yield")

```

Format

The 'vargas.wheat2.covs' data frame has 21 observations on the following 14 variables.

env environment

CYC length of growth cycle in days
 mTC mean daily minimum temperature in degrees Celsius
 MTC mean daily maximum temperature
 SHC sun hours per day
 mTV mean daily minimum temp during vegetative stage
 MTV mean daily maximum temp during vegetative stage
 SHV sun hours per day during vegetative stage
 mTS mean daily minimum temp during spike growth stage
 MTS mean daily maximum temp during spike growth stage
 SHS sun hours per day during spike growth stage
 mTG mean daily minimum temp during grainfill stage
 MTG mean daily maximum temp during grainfill stage
 SHG sun hours per day during grainfill stage

The 'vargas.wheat2.yield' data frame has 168 observations on the following 3 variables.

env environment
 gen genotype
 yield yield (kg/ha)

Details

Grain yields (kg/ha) for 8 wheat genotypes at 21 low-humidity environments grown during 1990-1994. The data is environment-centered and genotype-centered. The rows and columns of the GxE matrix have mean zero. The locations of the experiments were:

| | |
|-----|---|
| OBD | Ciudad Obregon, Mexico, planted in December |
| SUD | Wad Medani, Sudan |
| TLD | Tlaltizapan, Mexico, planted in December |
| TLF | Tlaltizapan, Mexico, planted in February |
| IND | Dharwar, India |
| SYR | Aleppo, Syria |
| NIG | Kadawa, Nigeria |

Source

Mateo Vargas and Jose Crossa and Ken Sayre and Matthew Renolds and Martha E Ramirez and Mike Talbot, 1998. Interpreting Genotype x Environment Interaction in Wheat by Partial Least Squares Regression, *Crop Science*, 38, 679–689. <http://doi.org/10.2135/cropsci1998.0011183X003800030010x>
 Data provided by Jose Crossa.

Examples

```

if(require(pls) & require(reshape2)){

  data(vargas.wheat2.covs)
  datc <- vargas.wheat2.covs

  data(vargas.wheat2.yield)
  daty <- vargas.wheat2.yield

  # Cast to matrix
  daty <- acast(daty, env ~ gen, value.var='yield')
  rownames(datc) <- datc$env
  datc$env <- NULL

  # The pls package centers, but does not (by default) use scaled covariates
  # Vargas says you should
  # daty <- scale(daty)
  datc <- scale(datc)

  m2 <- plsr(daty ~ datc)

  # Plot predicted vs observed for each genotype using all components
  plot(m2)

  # Loadings
  # plot(m2, "loadings", xaxt='n')
  # axis(1, at=1:ncol(datc), labels=colnames(datc), las=2)

  # Biplots
  biplot(m2, cex=.5, which="y", var.axes=TRUE,
         main="vargas.wheat2 - daty ~ datc") # Vargas figure 2a
  biplot(m2, cex=.5, which="x", var.axes=TRUE) # Vectors form figure 2 b
  # biplot(m2, cex=.5, which="scores", var.axes=TRUE)
  # biplot(m2, cex=.5, which="loadings", var.axes=TRUE)
}

```

 verbyla.lupin

Multi-environment trial of lupin, multiple varieties and densities

Description

Yield of 9 varieties of lupin at different planting densities across 2 years and multiple locations.

Format

gen genotype, 9 varieties
 site site, 11 levels
 rep rep, 2-3 levels
 rate seeding rate in plants/m²

row row
 col column
 serp factor of 4 levels for serpentine seeding effect
 linrow centered row position as a numeric variate (row-8.5)/10
 lincol centered column position as a numeric variate (col-3.5)
 linrate linear effect of seedrate, scaled (seedrate-41.92958)/10
 yield yield in tons/hectare
 year year, 1991-1992
 loc location

Details

Nine varieties of lupin were tested for yield response to plant density at 11 sites. The target density in 1991 was 10, 20, ..., 60 plants per m², and in 1992 was 20, 30, ..., 70 plants per m².

Plot dimensions are not given.

The variety Myallie was grown only in 1992.

Each site had 2 reps in 1991 and 3 reps in 1992. Each rep was laid out as a factorial RCB design; one randomization was used for all sites in 1991 and one (different) randomization was used for all sites in 1992. (This was confirmed with the principal investigator.)

In 1991 at the Mt. Barker location, the data for columns 5 and 6 was discarded due to problems with weeds.

Variety 'Myallie' was called '84L:439' in Verbyla 1997.

The year of release for the varieties is

| | |
|-----------|------|
| Unicrop | 1973 |
| Illyarrie | 1979 |
| Yandee | 1980 |
| Danja | 1986 |
| Gungurru | 1988 |
| Yorrel | 1989 |
| Warrah | 1989 |
| Merrit | 1991 |
| Myallie | 1995 |

Source

Arunas P. Verbyla and Brian R. Cullis and Michael G. Kenward and Sue J. Welham, (1999). The analysis of designed experiments and longitudinal data by using smoothing splines. *Appl. Statist.*, 48, 269–311. <http://doi.org/10.1111/1467-9876.00154>

Arunas P. Verbyla and Brian R. Cullis and Michael G. Kenward and Sue J. Welham, (1997). The analysis of designed experiments and longitudinal data by using smoothing splines. University of Adelaide, Department of Statistics, Research Report 97/4. <http://http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.56>

Data retrieved Oct 2010 from <http://www.blackwellpublishers.co.uk/rss>. (No longer available).

Used with permission of Blackwell Publishing.

Examples

```

data(verbyla.lupin)
dat <- verbyla.lupin

# The same RCB randomization was used at all sites in each year
if(require(desplot)){
  desplot(gen~col+row|site, data=dat,
          out1=rep, num=rate,
          # aspect unknown
          main="verbyla.lupin - experiment design")
}

# Figure 3 of Verbyla
require(lattice)
foo <- xyplot(yield ~ rate|loc*gen, data=dat, subset=year==92,
             type=c('p','smooth'), cex=.5,
             main="verbyla.lupin: 1992 yield response curves",
             xlab="Seed rate (plants/m^2)",
             ylab="Yield (tons/ha)",
             strip=strip.custom(par.strip.text=list(cex=.7)))

if (require(latticeExtra)) { # for useOuterStrips
  useOuterStrips(foo,
                 strip=strip.custom(par.strip.text=list(cex=.7)),
                 strip.left=strip.custom(par.strip.text=list(cex=.7)))
}

# -----

## Not run:

# We try to reproduce the analysis of Verbyla 1999.
# May not be exactly the same, but is pretty close.

# Check nlevels for size of random-coefficient structures
length(with(dat, table(gen)))      # 9 varieties for RC1
length(with(dat, table(gen,site))) # 99 site:gen combinations for RC2

# Make row and col into factors
dat <- transform(dat, colf=factor(col), rowf=factor(row))
# sort for asreml
dat <- dat[order(dat$site, dat$rowf, dat$colf),]

# Make site names more useful for plots
dat <- transform(dat, site=factor(paste0(year,".",substring(loc,1,4))))

# Initial model from top of Verbyla table 9.
# require(asreml4)
# FIXME asreml4 This is not working in asreml4

```

```

## m0 <- asreml(yield ~ 1
##           + site
##           + linrate
##           + site:linrate,
##           data = dat,
##           random = ~ spl(rate)
##           + dev(rate)
##           + site:spl(rate)
##           + site:dev(rate)
##           + str(~gen+gen:linrate, ~us(2):id(9)) # RC1
##           + gen:spl(rate)
##           + gen:dev(rate)
##           + str(~site+site:gen:linrate, ~us(2):id(99)) # RC2
##           + site:gen:spl(rate)
##           + site:gen:dev(rate),
##           residual = ~ dsum( ~ ar1(rowf):ar1(colf)|site) # Spatial AR1 x AR1
##           )
## m0 <- update(m0)
## m0 <- update(m0) # to get convergence

## # Variograms match Verbyla 1999 figure 7 (scale slightly different)
## plot(varioGram(m0), xlim=c(1:19), zlim=c(0,2),
##       main="verbyla.lupin - variogram by site")

## # Sequence of models in Verbyla 1999 table 10
## m1 <- update(m0, fixed= ~ .
##           + at(site, c(2,5,6,8,9,10)):lincol
##           + at(site, c(3,5,7,8)):linrow
##           + at(site, c(2,3,5,7,8,9,11)):serp
##           , random = ~ .
##           + at(site, c(3,6,7,9)):rowf
##           + at(site, c(1,2,3,9,10)):colf
##           + at(site, c(5,7,8,10)):units)
## m1 <- update(m1)

## m2 <- update(m1,
##           random = ~ .
##           - site:gen:spl(rate) - site:gen:dev(rate))

## m3 <- update(m2,
##           random = ~ .
##           - site:dev(rate) - gen:dev(rate))

## m4 <- update(m3,
##           random = ~ .
##           - dev(rate))

## m5 <- update(m4,
##           random = ~ .
##           - at(site, c(5,7,8,10)):units + at(site, c(5,7,8)):units)

## # Variance components are a pretty good match to Verbyla 1997, table 15

```

```
## require(lucid)
## vc(m5)
## .001004/sqrt(.005446*.0003662) # .711 correlation for RC1
## .00175/sqrt(.01881*.000167) # .987 correlation for RC2

## # Matches Verbyla 1999 figure 5
## plot(varioGram(m5),
##       main="verbyla.lupin - final model variograms",
##       xlim=c(1:19), zlim=c(0,1.5))

## End(Not run)
```

vold.longterm

Long-term barley yields at different fertilizer levels

Description

Long-term barley yields at different fertilizer levels

Usage

```
data("vold.longterm")
```

Format

A data frame with 76 observations on the following 3 variables.

year year

nitro nitrogen fertilizer, grams/m²

yield yield, grams/m²

Details

Trials conducted at Osaker, Norway. Nitrogen fertilizer amounts were increased by twenty percent in 1978.

Vold (1998) fit a Michaelis-Menten type equation with a different maximum in each year and a decreasing covariate for non-fertilizer nitrogen.

Miguez used a non-linear mixed effects model with asymptotic curve.

Source

Arild Vold (1998). A generalization of ordinary yield response functions. *Ecological modelling*, 108, 227-236. [http://doi.org/10.1016/S0304-3800\(98\)00031-3](http://doi.org/10.1016/S0304-3800(98)00031-3)

References

Fernando E. Miguez (2008). Using Non-Linear Mixed Models for Agricultural Data.

Examples

```

data(vold.longterm)
dat <- vold.longterm

require(lattice)
foo1 <- xyplot(yield ~ nitro | factor(year), data = dat,
  as.table=TRUE, type = "o",
  main=list("vold.longterm", cex=1.5),
  xlab = list("N fertilizer", cex=1.5, font=4),
  ylab = list("Yield", cex=1.5))
# print(foo1)

## Not run:
# Long term trend shows decreasing yields
xyplot(yield ~ year , data = dat, group=nitro, type='o',
  main="vold.longterm - yield level by nitrogen",
  auto.key=list(columns=4))

## End(Not run)

## Not run:
# Global model
m1.nls <- nls(yield ~ SSasymp(nitro, max, int, lograte), data=dat)
summary(m1.nls)
require(MASS) # for 'confint'
confint(m1.nls)

# Raw data plus global model. Year variation not modeled.
pdat <- data.frame(nitro=seq(0,14,0.5))
pdat$pred <- predict(m1.nls, newdata=pdat)
if(require(latticeExtra)) { # for layers
  foo1 + xyplot(pred ~ nitro , data = pdat,
    as.table=TRUE, type='l', col='red', lwd=2)
}

# Separate fit for each year. Overfitting with 3x19=57 params.
require(nlme)
m2.lis <- nlsList(yield ~ SSasymp(nitro,max,int,lograte) | year, data=dat)
plot(intervals(m2.lis), layout = c(3,1)) # lograte might be same for each year

## End(Not run)

# Fixed overall asymptotic model, plus random deviations for each year
# Simpler code, but less clear about what model is fit: m3.lme <- nlme(m2.lis)
require(nlme)
m3.lme <- nlme(yield ~ SSasymp(nitro, max, int, lograte), data=dat,

```

```

      groups = ~ year,
      fixed = list(max~1, int~1, lograte~1),
      random= max + int + lograte ~ 1,
      start= c(max=300, int=100, rate=-2))

## # Fixed effects are similar for the nls/lme models
## coef(m1.nls)
## fixef(m3.lme)
## # Random effects are normally distributed
## qqnorm(m3.lme, ~ ranef(.),col="black")
## # Note the trend in intercept effects over time
## plot(ranef(m3.lme),layout=c(3,1))

## # Correlation between int,lograte int,max may not be needed
## intervals(m3.lme,which="var-cov")
## pairs(m3.lme,pch=19,col="black")

## # Model with int uncorrelated with max,lograte. AIC is worse.
## # fit4.lm3 <- update(m3.lme, random=pdBlocked(list(max+lograte~1,int ~ 1)))
## # intervals(fit4.lm3, which="var-cov")
## # anova(m3.lme, fit4.lm3)

# Plot the random-effect model. Excellent fit with few parameters.
pdat2 <- expand.grid(year=1970:1988, nitro=seq(0,15,length=50))
pdat2$pred <- predict(m3.lme, new=pdat2)
pdat2$predf <- predict(m3.lme, new=pdat2, level=0)
foo1 <- update(foo1, type='p', key=simpleKey(c("Observed","Fixed","Random"),
      col=c("blue","red","darkgreen"), points=FALSE, columns=3))
if(require(latticeExtra)){
  foo2 <- xyplot(pred~nitro|year, data=pdat2, type='l', col="darkgreen", lwd=2)
  foo3 <- xyplot(predf~nitro|year, data=pdat2, type='l', col="red",lwd=1)
  foo1 + foo2 + foo3
}

## # Income is maximized at about 15
## pdat2 <- transform(pdat2, income = predf*2 - 7*nitro)
## with(pdat2, xyplot(income~nitro))

```

vsn.lupin3

Multi-environment trial of lupin, early generation trial

Description

Early generation lupin trial with 3 sites, 330 test lines, 6 check lines.

Format

A data frame with 1236 observations on the following 5 variables.

site site, levels S1 S2 S3

col column
 row row
 gen genotype
 yield yield

Details

An early-stage multi-environment trial, with 6 check lines and 300 test lines. The 6 check lines were replicated in each environment.

Source

Multi-Environment Trials - Lupins. <http://www.vsn.co.uk/software/asreml/htmlhelp/asreml/xlupin.htm>
 Used with permission of Arthur Gilmour, Brian Cullis, Robin Thompson.

Examples

```
data(vsn.lupin3)
dat <- vsn.lupin3

# Split gen into check/test, make factors
dat <- within(dat, {
  check <- ifelse(gen>336, 0, gen)
  check <- ifelse(check<7, check, 7)
  check <- factor(check)
  test <- factor(ifelse(gen>6 & gen<337, gen, 0))
  gen=factor(gen)
})

if(require(desplot)){
  desplot(yield~ col*row|site, dat,
          # aspect unknown
          main="vsn.lupin3 - yield")
  desplot(check~ col*row|site, dat,
          main="vsn.lupin3: check plot placement") # Site 1 & 2 used same randomization
}

## Not run:
# asreml3
require(asreml)

# Single-site analyses suggested random row term for site 3, random column terms
# for all sites, AR1 was unnecessary for the col dimension of site 3
dat <- transform(dat, colf=factor(col), rowf=factor(row))
dat <- dat[order(dat$site, dat$colf, dat$rowf),] # Sort for asreml
m1 <- asreml(yield ~ site + check:site, data=dat,
            random = ~ at(site):colf + at(site,3):rowf + test,
            rcov = ~ at(site,c(1,2)):ar1(colf):ar1(rowf)
            + at(site,3):id(colf):ar1(rowf))

m1$loglik
```

```

## [1] -314.2616

require(lucid)
vc(m1)
##
##          effect component std.error z.ratio constr
## at(site, S1):colf!colf.var  0.6228  0.4284    1.5   pos
## at(site, S2):colf!colf.var  0.159   0.1139    1.4   pos
## at(site, S3):colf!colf.var  0.04832 0.02618    1.8   pos
## at(site, S3):rowf!rowf.var  0.0235  0.008483   2.8   pos
##          test!test.var  0.1031  0.01468    7    pos
##          site_S1!variance  2.771   0.314    8.8   pos
##          site_S1!colf.cor  0.1959  0.05375   3.6   uncon
##          site_S1!rowf.cor  0.6503  0.03873  17    uncon
##          site_S2!variance  0.9926  0.1079   9.2   pos
##          site_S2!colf.cor  0.2868  0.05246   5.5   uncon
##          site_S2!rowf.cor  0.5744  0.0421   14    uncon
##          site_S3!variance  0.1205  0.01875   6.4   pos
##          site_S3!rowf.cor  0.6394  0.06323  10    uncon

# Add site:test
m2 <- update(m1, random=~. + site:test)
m2$loglik
## [1] -310.8794

# CORUH structure on the site component of site:test
m3 <- asreml(yield ~ site + check:site, data=dat,
             random = ~ at(site):colf + at(site,3):rowf + corh(site):test,
             rcov = ~ at(site,c(1,2)):ar1(colf):ar1(rowf)
             + at(site,3):id(colf):ar1(rowf))
m3$loglik
## [1] -288.4837

# Unstructured genetic variance matrix
m4 <- asreml(yield ~ site + check:site, data=dat,
             random = ~ at(site):colf + at(site,3):rowf + us(site):test,
             rcov = ~ at(site,c(1,2)):ar1(colf):ar1(rowf)
             + at(site,3):id(colf):ar1(rowf))
m4$loglik
## [1] -286.8239

# Note that a 3x3 unstructured matrix can be written LL'+Psi with 1 factor L
# Explicitly fit the factor analytic model
m5 <- asreml(yield ~ site + check:site, data=dat,
             random = ~ at(site):colf + at(site,3):rowf
             + fa(site,1, init=c(.7,.1,.1,.5,.3,.2)):test,
             rcov = ~ at(site,c(1,2)):ar1(colf):ar1(rowf)
             + at(site,3):id(colf):ar1(rowf))
m5$loglik # Same as m4
## [1] -286.8484

# Model 4, Unstructured (symmetric) genetic variance matrix
un <- diag(3)
un[upper.tri(un,TRUE)] <- m4$gammas[5:10]

```



```

round(un+t(un)-diag(diag(un)),3)
##      [,1] [,2] [,3]
## [1,] 0.992 0.158 0.132
## [2,] 0.158 0.073 0.078
## [3,] 0.132 0.078 0.122

# Model 5, FA matrix = LL'+Psi. Not quite the same as unstructured,
# since the FA model fixes site 2 variance at 0.
psi <- diag(m5$gammas[5:7])
lam <- matrix(m5$gammas[8:10], ncol=1)
round(tcrossprod(lam,lam)+psi,3)
##      [,1] [,2] [,3]
## [1,] 0.991 0.156 0.133
## [2,] 0.156 0.092 0.078
## [3,] 0.133 0.078 0.122

## End(Not run)

```

wallace.iowaland *Iowa farmland values by county in 1925*

Description

Iowa farmland values by county in 1925

Usage

```
data("wallace.iowaland")
```

Format

A data frame with 99 observations on the following 10 variables.

county county factor, 99 levels

fips FIPS code (state+county)

lat latitude

long longitude

yield average corn yield per acre (bu)

corn percent of land in corn

grain percent of land in small grains

untillable percent of land untillable

fedval land value (excluding buildings) per acre, 1925 federal census

stval land value (excluding buildings) per acre, 1925 state census

Details

None.

Source

H.A. Wallace (1926). Comparative Farm-Land Values in Iowa. *The Journal of Land & Public Utility Economics*, 2, 385-392. Page 387-388. <http://doi.org/10.2307/3138610>

References

Larry Winner. Spatial Data Analysis. <http://www.stat.ufl.edu/~winner/data/iowaland.txt>

Examples

```
data(wallace.iowaland)
dat <- wallace.iowaland

# Interesting trends involving latitude
require(lattice)
splom(~dat[, -c(1:2)], type=c('p', 'smooth'), lwd=2, main="wallace.iowaland")

# Means. Similar to Wallace table 1
apply(dat[, c('yield', 'corn', 'grain', 'untillable', 'fedval')], 2, mean)

# Correlations. Similar to Wallace table 2
round(cor(dat[, c('yield', 'corn', 'grain', 'untillable', 'fedval')]), 2)

m1 <- lm(fedval ~ yield + corn + grain + untillable, dat)
summary(m1) # estimates similar to Wallace, top of p. 389

# Choropleth map
require(maps)
data(county.fips)
dat <- transform(dat, polnm = paste0('iowa,', county)) # polnm example: iowa, adair

if(require("latticeExtra")){ # for mapplot
  redblue <- colorRampPalette(c("firebrick", "lightgray", "#375997"))
  mapplot(polnm~fedval, data=dat, colramp=redblue,
          main="wallace.iowaland - Federal land values",
          xlab="Land value, dollars per acre",
          scales=list(draw=FALSE),
          map=map('county', 'iowa', plot=FALSE,
                 fill=TRUE, projection="mercator"))
}
```

walsh.cottonprice *Acres and price of cotton 1910-1943*

Description

Acres and price of cotton 1910-1943

Format

A data frame with 34 observations on the following 9 variables.

year year, numeric 1910-1943
 acres acres of cototn (1000s)
 cotton price per pound (cents) in previous year
 cottonseed price per ton (dollars) in previous year
 combined cotton price/pound + 1.857 x cottonseed price/pound (cents)
 index price index, 1911-1914=100
 adjcotton adjusted cotton price per pound (cents) in previous year
 adjcottonseed adjusted cottonseed price per ton (dollars) in previous year
 adjcombined adjusted combined price/pound (cents)

Details

The 'index' is a price index for all farm commodities.

Source

R.M. Walsh (1944). Response to Price in Production of Cotton and Cottonseed, *Journal of Farm Economics*, 26, 359-372. <http://doi.org/10.2307/1232237>

Examples

```
data(walsh.cottonprice)
dat <- walsh.cottonprice

dat <- transform(dat, acres=acres/1000) # convert to million acres

percentchg <- function(x){ # percent change from previous to current
  ix <- 2:(nrow(dat))
  c(NA, (x[ix]-x[ix-1])/x[ix-1])
}

# Compare percent change in acres with percent change in previous price
# using constant dollars
dat <- transform(dat, chga = percentchg(acres), chgp = percentchg(adjcombined))
```

```
with(dat, cor(chga, chgp, use='pair')) # .501 correlation
require(lattice)
xyplot(chga~chgp, dat, type=c('p','r'),
       main="walsh.cottonprice",
       xlab="Percent change in previous price", ylab="Percent change in acres")
```

wassom.brome.uniformity

Uniformity trials of bromegrass

Description

Uniformity trials of bromegrass at Ames, Iowa, 1950-1951.

Usage

```
data("wassom.brome.uniformity")
```

Format

A data frame with 1296 observations on the following 3 variables.

```
expt experiment
row row
col column
yield forage yield, pounds
```

Details

Experiments were conducted at Ames, Iowa. The response variable is forage yield in pounds of green weight.

Optimum plot size was estimated to be about 3.5 x 7.5 feet.

Wassom and Kalton used two different methods to estimate optimum plot size. 1. Relative efficiency of different plot sizes. 2. Regression of the log variance of yield vs log plot size.

There are three Experiments:

Experiment E1 was broadcast seeded, harvested in 1950.

Experiment E2 was row planted, harvested in 1950.

Experiment E3 was broadcast seeded, harvested in 1951. This field contained a mixture of alfalfa and brome in about equal proportions.

Each plot was 3.5 ft x 4 ft, but the orientation of the plot is not clear.

Field width: 36 plots

Field length: 36 plots

Source

Wassom and R.R. Kalton. (1953). Estimations of Optimum Plot Size Using Data from Bromegrass Uniformity Trials. Agricultural Experiment Station, Iowa State College, Bulletin 396. http://lib.dr.iastate.edu/ag_researchbull

Examples

```
data(wassom.brome.uniformity)
dat <- wassom.brome.uniformity

if(require(desplot)){
  desplot(yield~col*row|expt, dat,
          flip=TRUE, aspect=1, # approximate aspect
          main="wassom.brome.uniformity")
}
```

waynick.soil

Soil nitrogen and carbon in two fields

Description

Soil nitrogen and carbon in two fields

Format

A data frame with 200 observations on the following 6 variables.

field field name, 2 levels

sample sample number

x x ordinate

y y ordinate

nitro nitrogen content, percent

carbon carbon content, percent

Details

Two fields were studied, one at University Farm in Davis, the other near Oakley. The Davis field is silty clay loam, the Oakley field is blow sand.

Source

Waynick, Dean, and Sharp, Leslie. (1918). Variability in soils and its significance to past and future soil investigations, I-II. University of California press. <http://archive.org/details/variabilityinsoi45wayn>

Examples

```

data(waynick.soil)
dat <- waynick.soil

# Strong relationship between N,C
require(lattice)
xyplot(nitro~carbon|field, data=dat, main="waynick.soil")

# Spatial plot
if(require(sp) & require(gstat)){
d1 <- subset(dat, field=="Davis")
d2 <- subset(dat, field=="Oakley")
coordinates(d1) <- data.frame(x=d1$x, y=d1$y)
coordinates(d2) <- data.frame(x=d2$x, y=d2$y)
sppplot(d1, zcol = "nitro", cuts=8, cex = 1.6,
        main = "waynick.soil - Davis field - nitrogen",
        col.regions = bpy.colors(8), key.space = "right")

# Variogram
v1 <- gstat::variogram(nitro~1, data=d1)
plot(v1, main="waynick.soil - Davis field - nitrogen") # Maybe hasn't reached sill
}

```

| | |
|-------------------|---|
| wedderburn.barley | <i>Multi-environment trial of barley, percent of leaves affected by leaf blotch</i> |
|-------------------|---|

Description

Percent of leaf area affected by leaf blotch on 10 varieties of barley at 9 sites.

Format

A data frame with 90 observations on the following 3 variables.

y Percent of leaf area affected, 0-100.

site Site factor, 9 levels

gen Variety factor, 10 levels

Details

Incidence of *Rhynchosporium secalis* (leaf blotch) on the leaves of 10 varieties of barley grown at 9 sites in 1965.

Source

Wedderburn, R W M (1974). Quasilikelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika*, 61, 439–47. <http://doi.org/10.2307/2334725>

Wedderburn credits the original data to an unpublished thesis by J. F. Jenkyn.

References

McCullagh, P and Nelder, J A (1989). *Generalized Linear Models* (2nd ed).

R. B. Millar. *Maximum Likelihood Estimation and Inference: With Examples in R, SAS and ADMB*. Chapter 8.

Examples

```
data(wedderburn.barley)
dat <- wedderburn.barley
dat$y <- dat$y/100

require(lattice)
dotplot(gen~y|site, dat, main="wedderburn.barley")

# Use the variance function mu(1-mu). McCullagh page 330
# Note, 'binomial' gives same results as 'quasibinomial', but also a warning
m1 <- glm(y ~ gen + site, data=dat, family="quasibinomial")
summary(m1)

# Same shape (different scale) as McCullagh fig 9.1a
plot(m1, which=1, main="wedderburn.barley")

# Compare data and model
dat$pbm <- predict(m1, type="response")
dotplot(gen~pbm+y|site, dat, main="wedderburn.barley: observed/predicted")

# Wedderburn suggested variance function: mu^2 * (1-mu)^2
# Millar shows how to do this explicitly.
wedder <- list(varfun=function(mu) (mu*(1-mu))^2,
              validmu=function(mu) all(mu>0) && all(mu<1),
              dev.resids=function(y,mu,wt) wt * ((y-mu)^2)/(mu*(1-mu))^2,
              initialize=expression({
                n <- rep.int(1, nobs)
                mustart <- pmax(0.001, pmin(0.99,y)) }),
              name="(mu(1-mu))^2")
m2 <- glm(y ~ gen + site, data=dat, family=quasi(link="logit", variance=wedder))
#plot(m2)

## Not run:
# Alternatively, the 'gnm' package has the 'wedderburn' family.
require(gnm)
m3 <- glm(y ~ gen + site, data=dat, family="wedderburn")
summary(m3)
```

```
# Similar to McCullagh fig 9.2
plot(m3, which=1)

# Compare data and model
dat$pwed <- predict(m3, type="response")
dotplot(gen~pwed+y|site, dat)

## End(Not run)
```

weiss.incblock

Soybean balanced incomplete block experiment

Description

Soybean balanced incomplete block experiment

Usage

```
data("weiss.incblock")
```

Format

A data frame with 186 observations on the following 5 variables.

block block factor

gen genotype (variety) factor

yield yield (bu/ac)

row row

col column

Details

Grown at Ames, Iowa in 1937. Each plot was 6 feet by 16 feet (2 rows, 3 feet apart). Including space between plots, the entire experiment was 252 ft x 96 feet (7 block * 6 plots * 6 feet = 252, 16*5 plots plus 4 gaps of 4 feet). Weiss shows a figure of the field (that was later doubled in dize via using two rows per plot).

Note that only 30 varieties were tested. Varieties 7 and 14 are the same variety (Mukden). Although total yields of these varieties were not equal, the correction for blocks adjusted their means to identical values. Such accuracy is not, however, claimed to be a constant characteristic of the design.

Field length: 252 feet

Field width: 96 feet

Source

Weiss, Martin G. and Cox, Gertrude M. (1939). Balanced Incomplete Block and Lattice Square Designs for Testing Yield Differences Among Large Numbers of Soybean Varieties. *Agricultural Research Bulletins, Nos. 251-259*. http://lib.dr.iastate.edu/ag_researchbulletins/24/

Examples

```
data(weiss.incblock)
dat <- weiss.incblock

# True aspect as shown in Weiss and Cox
if(require(desplot)){
  desplot(yield~col*row, dat,
          text=gen, shorten='none', cex=.6, out1=block,
          aspect=252/96, # true aspect
          main="weiss.incblock")
}

## Not run:
# Standard inc block analysis used by Weiss and Cox
require(asreml)
m1 <- asreml(yield ~ gen + block , data=dat)
predict(m1, data=dat, classify="gen")$predictions$pvals
## gen pred.value std.error est.stat
## G01      24.59    0.8312 Estimable
## G02      26.92    0.8312 Estimable
## G03      32.62    0.8312 Estimable
## G04      26.97    0.8312 Estimable
## G05      26.02    0.8312 Estimable

## End(Not run)
```

weiss.lattice

Lattice experiment in soybeans.

Description

Lattice experiment in soybeans.

Usage

```
data("weiss.lattice")
```

Format

A data frame with 196 observations on the following 5 variables.

```
yield yield (bu/ac)
gen genotype factor, 49 levels
rep rep factor, 4 levels
col column
row row
```

Details

Yield test of 49 soybean varieties, grown at Ames, IA, in 1938. Plot dimensions were 3x16 feet. The varieties are compared to variety 26 (Mukden).

It is not clear how the reps were positioned in the field. On the one hand, the middle three columns of each rep/square are higher yielding, giving the appearance of the reps being stacked on top of each other. On the other hand, the analysis by Weiss uses 24 degrees of freedom $4*(7-1)$ to fit a separate effect for each column in each rep (instead of across reps).

Source

Weiss, Martin G. and Cox, Gertrude M. (1939). Balanced Incomplete Block and Lattice Square Designs for Testing Yield Differences Among Large Numbers of Soybean Varieties. Table 5. *Agricultural Research Bulletins*, Nos. 251-259. http://lib.dr.iastate.edu/ag_researchbulletins/24/

Examples

```
data(weiss.lattice)
dat <- weiss.lattice

if(require(desplot)){
  desplot(yield~col*row|rep,dat,
          text=gen, shorten="none", cex=.8, aspect=3/16, # true aspect
          main="weiss.lattice (layout uncertain)", xlab="Soybean yields")
}

dat <- transform(dat, xf=factor(col), yf=factor(row))
m1 <- lm(terms(yield ~ rep + rep:xf + rep:yf + gen, keep.order=TRUE), data=dat)
anova(m1) # Matches Weiss table 7
## Response: yield
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## rep         3   91.57   30.525  4.7414 0.0039709 **
## rep:xf      24 2913.43  121.393 18.8557 < 2.2e-16 ***
## rep:yf      24  390.21   16.259  2.5254 0.0007734 ***
## gen        48 1029.87   21.456  3.3327 2.652e-07 ***
## Residuals 96  618.05    6.438

# -----
```

```
## Not run:
require(asreml)
m2 <- asreml(yield ~ rep + rep:xf + rep:yf + gen, data=dat)
anova(m2)
predict(m2, data=dat, classify="gen")$predictions$pvals # Weiss table 6 means
##   gen pred.value std.error est.stat
##   G01      27.74    1.461 Estimable
##   G02      24.95    1.461 Estimable
##   G03      24.38    1.461 Estimable
##   G04      28.05    1.461 Estimable
##   G05      19.6    1.461 Estimable
##   G06      23.79    1.461 Estimable

## End(Not run)
```

welch.bermudagrass *Yield of bermuda grass with N, P, K fertilizers*

Description

Yield of bermuda grass with N, P, K fertilizers

Format

A data frame with 64 observations on the following 4 variables.

n nitrogen fertilizer, pounds/acre

p phosphorus, lb/ac

k potassium, lb/ac

yield yield of grass, tons/ac

Details

The experiment was conducted 1955, 1956, and 1957. There were 3 blocks. The harvests were oven-dried. Each value is the mean for 3 years and 3 replications. In most cases, the yield increased with additions of the fertilizer nutrients.

Source

Welch, Louis Frederick and Adams, William Eugenius and Carmon, JL. 1963. Yield response surfaces, isoquants, and economic fertilizer optima for Coastal Bermudagrass. *Agronomy Journal*, 55, 63-67. Table 1. <https://doi.org/10.2134/agronj1963.00021962005500010023x>

References

Jim Albert. *Bayesian Computation with R*. Page 256.

Peter Congdon. *Bayesian Statistical Modeling*. Page 124-125.

P. McCullagh, John A. Nelder. *Generalized Linear Models*, Second Edition. Page 382.

Examples

```

data(welch.bermudagrass)
dat <- welch.bermudagrass
# Welch uses 100-pound units of n,p,k.
dat <- transform(dat, n=n/100, p=p/100, k=k/100)

if(require(latticeExtra)){
  useOuterStrips(xyplot(yield~n|factor(p)*factor(k), data=dat, type='b',
    main="welch.bermudagrass: yield for each P*K",
    xlab="Nitro for each Phosphorous level",
    ylab="Yield for each Potassim level"))
}

# Fit a quadratic model
m1 <- lm(yield ~ n + p + k + I(n^2) + I(p^2) + I(k^2) + n:p + n:k + p:k + n:p:k, data=dat)
signif(coef(m1),4) # These match the 3-yr coefficients of Welch, Table 2
## (Intercept)          n          p          k      I(n^2)      I(p^2)
##   1.94300    2.00700    1.47100    0.61880   -0.33150   -1.29500
##      I(k^2)      n:p      n:k      p:k      n:p:k
##   -0.37430    0.20780    0.18740    0.23480    0.02789

# Welch Fig 4. Modeled response curves
d1 <- expand.grid(n=seq(0, 4, length=50), p=0, k=0)
d1$pred <- predict(m1, d1)
d2 <- expand.grid(n=0, p=0, k=seq(0, 1.68, length=50))
d2$pred <- predict(m1, d2)
d3 <- expand.grid(n=0, p=seq(0, .88, length=50), k=0)
d3$pred <- predict(m1, d3)

op <- par(mfrow=c(1,3), mar=c(5,3,4,1))
plot(pred~n, data=d1, type='l', ylim=c(0,6), xlab="N 100 lb/ac", ylab="")
plot(pred~k, data=d2, type='l', ylim=c(0,6), xlab="K 100 lb/ac", ylab="")
title("welch.bermudagrass - Predicted yield vs fertilizer", outer=TRUE, line= -3)
plot(pred~p, data=d3, type='l', ylim=c(0,6), xlab="P 100 lb/ac",
ylab="")
par(op)

# Brute-force grid-search optimization of fertilizer quantities, using
# $25/ton for grass, $.12/lb for N, $.18/lb for P, $.07/lb for K
# Similar to Example 5 in Table 4 of Welch
d4 <- expand.grid(n=seq(3,4,length=20), p=seq(.5, 1.5, length=20),
  k=seq(.8, 1.8, length=20))
d4$pred <- predict(m1, newdata=d4)
d4 <- transform(d4, income = 25*pred - .12*n*100 + -.18*p*100 - .07*k*100)
d4[which.max(d4$income),] # Optimum at 300 lb N, 71 lb P, 148 lb K

# ----- JAGS -----
## Not run:
# Congdon (2007) p. 124, provides a Bayesian model based on a GLM
# by McCullagh & Nelder. We use JAGS and simplify the code.

```

```

# y ~ gamma with shape = nu, scale = nu * eps_i
# 1/eps = b0 + b1/(N+a1) + b2/(P+a2) + b3/(K+a3)
# N,P,K are added fertilizer amounts, a1,a2,a3 are background
# nutrient levels and b1,b2,b3 are growth parameters.

require(rjags)

mod.bug =
"model {
for(i in 1:nobs) {
  yield[i] ~ dgamma(nu, mu[i])
  mu[i] <- nu * eta[i]
  eta[i] <- b0 + b1 / (N[i]+a1) + b2 / (P[i]+a2) + b3 / (K[i]+a3)
  yhat[i] <- 1 / eta[i]
}

# Hyperparameters
nu ~ dgamma(0.01, 0.01)
a1 ~ dnorm(40, 0.01) # Informative priors
a2 ~ dnorm(22, 0.01)
a3 ~ dnorm(32, 0.01)
b0 ~ dnorm(0, 0.0001)
b1 ~ dnorm(0, 0.0001) I(0,) # Keep b1 non-negative
b2 ~ dnorm(0, 0.0001) I(0,)
b3 ~ dnorm(0, 0.0001) I(0,)
}"

jdat <- with(welch.bermudagrass,
             list(yield=yield, N=n, P=p, K=k, nobs=64))
jinit = list(a1=40, a2=22, a3=32, b0=.1, b1=10, b2=1, b3=1)

oo <- textConnection(mod.bug)
j1 <- jags.model(oo, data=jdat, inits=jinit, n.chains=3)
close(oo)

c1 <- coda.samples(j1, c("b0","b1","b2","b3", "a1","a2","a3"),
                  n.iter=10000)

# Results nearly identical go Congdon
print(summary(c1)$statistics[,1:2],dig=1)
# require(lucid)
# print(vc(c1),3)
##      Mean    SD
## a1  44.85  4.123
## a2  23.63  7.37
## a3  35.42  8.57
## b0   0.092 0.0076
## b1  13.23  1.34
## b2   1.186 0.47
## b3   1.50  0.48

d2 <- coda.samples(j1, "yhat", n.iter=10000)
dat$yhat <- summary(d2)$statistics[,1]

```

```
with(dat, plot(yield, yield-yhat))  
## End(Not run)
```

wheatley.carrot *Insecticide treatments for carrot fly larvae*

Description

Insecticide treatments for carrot fly larvae. Two insecticides with five depths.

Usage

```
data("wheatley.carrot")
```

Format

A data frame with 36 observations on the following 6 variables.

```
treatment treatment factor, 11 levels  
insecticide insecticide factor  
depth depth  
rep block  
damaged number of damaged plants  
total total number of plants
```

Details

In 1964 an experiment was conducted with microplots to evaluate the effectiveness of treatments against carrot fly larvae. The treatment factor is a combination of insecticide and depth.

Hardin & Hilbe used this data to fit a generalized binomial model.

Famoye (1995) used the same data to fit a generalized binomial regression model. Results for Famoye are not shown.

Source

G A Wheatley & H Freeman. (1982). A method of using the proportions of undamaged carrots or parsnips to estimate the relative population densities of carrot fly (*Psila rosae*) larvae, and its practical applications. *Annals of Applied Biology*, 100, 229-244. Table 2.

<http://doi.org/10.1111/j.1744-7348.1982.tb01935.x>

References

James William Hardin, Joseph M. Hilbe. *Generalized Linear Models and Extensions*, 2nd ed.

F Famoye (1995). Generalized Binomial Regression. *Biom J*, 37, 581-594.

Examples

```

data(wheatley.carrot)
dat <- wheatley.carrot

# Observed proportions of damage
dat <- transform(dat, prop=damaged/total)
require(lattice)
xyplot(prop~depth|insecticide, data=dat, subset=treatment!="T11",
       cex=1.5, main="wheatley.carrot", ylab="proportion damaged")

# Model for Wheatley. Deviance for treatment matches Wheatley, but other
# deviances do not. Why?
# treatment:rep is the residual
m1 <- glm(cbind(damaged,total-damaged) ~ rep + treatment + treatment:rep,
         data=dat, family=binomial("cloglog"))
anova(m1)

# GLM of Hardin & Hilbe p. 161. By default, R uses T01 as the base,
# but Hardin uses T11. Results match.
m2 <- glm(cbind(damaged,total-damaged) ~ rep + C(treatment, base=11),
         data=dat, family=binomial("cloglog"))
summary(m2)

```

wiebe.wheat.uniformity

Uniformity trial of wheat

Description

Uniformity trial of wheat at Aberdeen, Idaho, 1927.

Format

A data frame with 1500 observations on the following 3 variables.

row row

col column (series)

yield yield in grams per plot

Details

Yield trial conducted in 1927 near Aberdeen, Idaho. The crop was Federation wheat (C.I. no 4734). Plots were seeded on April 18 with a drill that sowed eight rows at a time. Individual rows were harvested in August and threshed with a small nursery thresher. Some authors recommend analyzing the square root of the yields.

Rows were 15 feet long, 1 foot apart.

Field width: 12 cols * 15 feet = 180 feet wide.

Field length: 125 rows * 12 in = 125 feet

Source

Wiebe, G.A. 1935. Variation and Correlation in Grain Yield among 1,500 Wheat Nursery Plots, *Journal of Agricultural Research*, 50, 331-357. <http://naldc.nal.usda.gov/download/IND43968632/PDF>

References

D.A. Preece, 1981, Distributions of final digits in data, *The Statistician*, 30, 31–60. <http://doi.org/10.2307/2987702>

Wilkinson et al. (1983). Nearest Neighbour (NN) Analysis of Field Experiments. *J. R. Statist. Soc. B*, 45, 151-211.

Examples

```
data(wiebe.wheat.uniformity)
dat <- wiebe.wheat.uniformity

if(require(desplot)){
  desplot(yield~col+row, dat,
          aspect=125/180, flip=TRUE, # true aspect
          main="wiebe.wheat.uniformity: yield") # row 1 is at south
}

# Preece (1981) found the last digits have an interesting distribution
# with 0 and 5 much more common than other digits.
dig <- substring(dat$yield, nchar(dat$yield))
dig <- as.numeric(dig)
hist(dig, breaks=0:10-.5, xlab="Last digit",
     main="wiebe.wheat.uniformity - histogram of last digit")
table(dat$col, dig) # Table 3 of Preece

# Loess
# m3 <- loess(yield~row+col, dat)
# plot(dat$row, resid(m3), ylim=c(-300,300),
#      main="wiebe.wheat.uniformity")

# Wilkinson (1983, p. 152) noted that an 8-row planter was used which
# produced a recurring pattern of row effects on yield. This can be seen
# in the high autocorrelations of row means at lag 8 and lag 16
rowm <- tapply(dat$yield, dat$row, mean)
acf(rowm, main="wiebe.wheat.uniformity row means")
# Plot the row mean against the planter row unit 1-8
require("lattice")
xyplot(rowm~rep(1:8, length=125),
       main="wiebe.wheat.uniformity",
       xlab="Planter row unit", ylab="Row mean yield")
```

wiedemann.safflower.uniformity
Uniformity trial of safflower

Description

Uniformity trial of safflower at Farmington, Utah, 1960.

Usage

```
data("wiedemann.safflower.uniformity")
```

Format

A data frame with 1782 observations on the following 3 variables.

row row

col column

yield yield, grams

Details

This trial was planted at University Field Station, Farmington, Utah, in 1960, on a plot of land about one half acre in size. The soil was not too uniform...the northern third of the field was clay and the rest was gravelly. Rows were planted 22 inches apart, 62 rows total, each row running the length of the field. Before harvest, 4 rows were removed from each side, and 12 feet was removed from each end. Each row was harvested in five-foot lengths, threshed, and the seed weighed to the nearest gram.

The northern third of the field had yields twice as high as the remaining part of the field because the soil had better moisture retention. The remaining part of the field had yields that were more uniform.

Wiedemann determined the optimum plot size to be about 8 basic plots. The shape of the plot was not very important. But, two-row plots were recommended for simplicity of harvest, so 3.33 feet by 20 feet.

Based on operational costs, $K_1=74$ percent and $K_2=26$ percent.

Field width: 33 plots/ranges * 5ft = 165 feet

Field length: 54 rows * 22 in/row = 99 feet

For this R package, the tables in Wiedemann were converted by OCR to digital format, and all values were checked by hand.

The original source document has columns labeled 33, 32, ... 1. Here the columns are labeled 1:33 so that plotting tools work normally. See Wiedemann figure 8.

Wiedemann notes the statistical analysis of the data required 100 hours of labor. Today the analysis takes only a second.

Source

Wiedemann, Alfred Max. 1962. "Estimation of Optimum Plot Size and Shape for Use in Safflower Yield Trails". Table 5. All Graduate Theses and Dissertations. Paper 3600. Table 5. <http://digitalcommons.usu.edu/etd/3600>

References

None.

Examples

```
data(wiedemann.safflower.uniformity)
dat <- wiedemann.safflower.uniformity

# CV of entire field = 39%, matches Wiedemann p. 16
sd(dat$yield)/mean(dat$yield)

if(require(desplot)){
  desplot(yield~col*row, dat,
          flip=TRUE, tick=TRUE, aspect =99/165, # true aspect
          main="wiedemann.safflower.uniformity (true shape)")
}

## Not run:
require(agricolae) ; require(reshape2)
dmat <- acast(dat, row~col, value.var='yield')
index.smith(dmat,
            main="wiedemann.safflower.uniformity",
            col="red")

## End(Not run)
```

williams.barley.uniformity

Uniformity trial of barley

Description

Uniformity trial of barley at Narrabri, New South Wales, 1984.

Format

A data frame with 720 observations on the following 3 variables.

row row

col column

yield grain yield kg/ha divided by 10

Details

Grown at Roseworthy Agricultural College. Plots were 5 m long (4 m sown, 3.3 m harvested) by 0.75 m wide.

A three-plot seeder was used, planting in a serpentine fashion. Williams noted that it appears that the middle plot of each pass has a lower yield, possibly due to soil compaction from the tractor.

Field width: 48 plots * .75 m = 36 m

Field length: 15 plots * 5 m = 75 m

Source

Williams, ER and Luckett, DJ. 1988. The use of uniformity data in the design and analysis of cotton and barley variety trials. *Australian Journal of Agricultural Research*, 39, 339-350. <http://doi.org/10.1071/AR9880339>

Examples

```
data(williams.barley.uniformity)
dat <- williams.barley.uniformity

if(require(desplot)){
  desplot(yield ~ col*row, dat,
          aspect= 75/36, # true aspect
          main="williams.barley.uniformity")
}

# Smoothed contour/persp plot like Williams Fig 1b, 2b
require("lattice")
dat$fit <- fitted(loess(yield~col*row, dat, span=.1))
contourplot(fit~col*row, data=dat,
            aspect=75/36, region=TRUE, col.regions=RedGrayBlue,
            main="williams.barley.uniformity")
wireframe(fit~col*row, data=dat, zlim=c(100, 350),
          main="williams.barley.uniformity")

# Williams table 1
anova(aov(yield ~ factor(row) + factor(col), dat))
```

williams.cotton.uniformity

Uniformity trial of cotton

Description

Uniformity trial of cotton at Narrabri, New South Wales, 1984.

Format

A data frame with 288 observations on the following 3 variables.

```
row row
col column
yield lint yield, kg/ha divided by 10
```

Details

Cotton uniformity trial grown at Narrabri, New South Wales, 1984-1985. Plots were 12m long, 1m apart, 12 rows by 24 columns, with an irrigation furrow between columns.

Field width: 24 plots * 1 m = 24 m

Field length: 12 plots * 12 m = 144 m

Source

Williams, ER and Luckett, DJ. 1988. The use of uniformity data in the design and analysis of cotton and barley variety trials. *Australian Journal of Agricultural Research*, 39, 339-350. <http://doi.org/10.1071/AR9880339>

Examples

```
data(williams.cotton.uniformity)
dat <- williams.cotton.uniformity

if(require(desplot)){
  desplot(yield ~ col*row, dat,
          aspect=144/24, # true aspect
          main="williams.cotton.uniformity")
}

# Smoothed contour/persp plot like Williams 1988 Fig 1a, 2a
dat$fit <- fitted(loess(yield~col*row, dat, span=.5))
require("lattice")
contourplot(fit~col*row, data=dat,
            aspect=144/24,
            region=TRUE, cuts=6, col.regions=RedGrayBlue,
            main="williams.cotton.uniformity")
# wireframe(fit~col*row, data=dat, zlim=c(100, 250),
#           main="williams.cotton.uniformity")

# Williams table 1
anova(aov(yield ~ factor(row) + factor(col), dat))
```

`williams.trees`*Height / Survival of 37 tree species at six sites in Thailand*

Description

Height / Survival of 37 tree species at six sites in Thailand

Format

A data frame with 222 observations on the following 4 variables.

`env` Environment factor, 6 levels

`gen` Genetic factor, 37 levels

`height` Height (cm)

`survival` Survival percentage

Details

Planted in 1985 at six sites in Thailand. RCB with 3 reps. The data here is the mean of the three reps. Plots were 5 meters square with spacing 2m x 2m. Measurements collected at 24 months. The `gen` column in the data is actually *seedlot*, as some tree species have multiple seed lots. The trees are mostly acacia and eucalyptus.

Source

Williams, ER and Luangviriyasaeng, V. 1989. Statistical analysis of tree species trial and seedlot:site interaction in Thailand. Chapter 14 of *Trees for the Tropics: Growing Australian Multipurpose Trees and Shrubs in Developing Countries*. Pages 145–152. <http://aciarc.gov.au/publication/MN010>

Used with permission of Emlyn Williams.

References

E. R. Williams and A. C. Matheson and C. E Harwood, *Experimental Design and Analysis for Tree Improvement*, CSIRO Publishing, 2002.

Examples

```
data(williams.trees)
dat <- williams.trees
```

```
require(lattice)
xyplot(survival~height|env,dat, main="williams.trees", xlab="Height",
ylab="Percent surviving")
```

yan.winterwheat

Multi-environment trial of winter wheat in Ontario

Description

Yield of 18 varieties of winter wheat grown at 9 environments in Ontario in 1993.

Format

A data frame with 162 observations on the following 3 variables.

gen genotype

env environment

yield yield in metric tons per hectare

Details

The yield is the mean of several reps, measured in metric tons per hectare.

This data has often been used to illustrate GGE biplots.

Source

Weikai Yan and M.S. Kang. *GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists*, 2002. CRC. Page 59.

Weikai Yan and Nicholas A. Tinker. 2006. Biplot analysis of multi-environment trial data: Principles and applications. Table 1.

Used with permission of Weikai Yan.

References

Weikai Yan and Manjit S. Kang and Baoluo Ma and Sheila Woods, 2007, GGE Biplot vs. AMMI Analysis of Genotype-by-Environment Data, *Crop Science*, 2007, 47, 641–653. <http://doi.org/10.2135/cropsci2006.06.0374>

Examples

```
data(yan.winterwheat)
dat <- yan.winterwheat

if(require(gge)){
m1 <- gge(yield ~ gen*env, data=dat)
biplot(m1, flip=c(1,1), hull=TRUE,
       main="yan.winterwheat - GGE biplot")
}
```

| | |
|-------------|--|
| yang.barley | <i>Multi-environment trial of barley in Alberta, 6 varieties at 18 locations in Alberta.</i> |
|-------------|--|

Description

Yield of 6 barley varieties at 18 locations in Alberta.

Usage

```
data("yang.barley")
```

Format

A data frame with 108 observations on the following 3 variables.

site site factor, 18 levels

gen genotype factor, 6 levels

yield yield, Mg/ha

Details

From an experiment in 2003. Yang (2013) uses this data to illustrate a procedure for bootstrapping biplots.

| site | long | lat |
|-----------------|--------|-------|
| Beaverlodge | 119.43 | 55.21 |
| BigLakes | 113.70 | 53.61 |
| Calmar | 113.85 | 53.26 |
| CdcNorth | 113.33 | 53.63 |
| DawsonCreek | 120.23 | 55.76 |
| FtKent | 110.61 | 54.31 |
| FtStJohn | 120.85 | 56.25 |
| Irricana | 113.60 | 51.32 |
| Killam | 111.85 | 52.78 |
| Lacombe | 113.73 | 52.46 |
| LethbridgeDry | 112.81 | 49.70 |
| LethbridgeIrr | 112.81 | 49.70 |
| Lomond | 112.65 | 50.35 |
| Neapolis | 113.86 | 51.65 |
| NorthernSunrise | NA | NA |
| Olds | 114.09 | 51.78 |
| StPaul | 111.28 | 53.98 |
| Stettler | 112.71 | 52.31 |

Source

Rong-Cai Yang (2007). Mixed-Model Analysis of Crossover Genotype-Environment Interactions. *Crop Science*, 47, 1051-1062. <http://doi.org/10.2135/cropsci2006.09.0611>

Used with permission of Rong-Cai Yang.

References

Zhiqiu Hu and Rong-Cai Yang, (2013). Improved Statistical Inference for Graphical Description and Interpretation of Genotype x Environment Interaction. *Crop Science*, 53, 2400-2410. <http://doi.org/10.2135/cropsci2013.04.0218>

Examples

```
data(yang.barley)
dat <- yang.barley

if(require(reshape2)){
  dat <- acast(dat, gen~site, value.var='yield')
}

## Not run:

## For bootstrapping of a biplot, see the non-cran packages:
## 'bbplot' and 'distfree.cr'
## http://statgen.ualberta.ca/index.html?open=software.html
## install.packages("http://statgen.ualberta.ca/download/software/bbplot_1.0.zip")
## install.packages("http://statgen.ualberta.ca/download/software/distfree.cr_1.5.zip")

## require(SDMTools)
## require(distfree.cr)
## require(bbplot)

## d1 <- bbplot.boot(dat, nsample=2000) # bootstrap the data
## plot(d1) # plot distributions of principal components
## b1 <- bbplot(d1) # create data structures for the biplot
## plot(b1) # create the confidence regions on the biplot

## End(Not run)
```

yates.missing

Factorial experiment with missing values

Description

Potato factorial experiment with missing values

Format

A data frame with 80 observations on the following 3 variables.

trt treatment factor with levels 0 K N P NK KP NP NKP

block block, 10 levels

y infection intensity

Details

The response variable *y* is the intensity of infection of potato tubers inoculated with *Phytophthora Erythroseptica*.

Yates (1933) presents an iterative algorithm to estimate missing values in a matrix, using this data as an example.

Source

F. Yates, 1933. The analysis of replicated experiments when the field results are incomplete. *Emp. J. Exp. Agric.*, 1, 129–142.

References

Steel & Torrie, 1980, Principles and Procedures of Statistics, 2nd Edition, page 212.

Examples

```
data(yates.missing)
dat <- yates.missing

require(lattice)
bwplot(y ~ trt, data=dat,
       xlab="Treatment", ylab="Infection intensity",
       main="yates.missing")

if(require(reshape2)){
mat0 <- acast(dat[, c('trt','block','y')], trt~block,
             id.var=c('trt','block'), value.var='y')

# Use lm to estimate missing values. The estimated missing values
# are the same as in Yates (1933)
m1 <- lm(y~trt+block, dat)
dat$pred <- predict(m1, new=dat[, c('trt','block')])
dat$filled <- ifelse(is.na(dat$y), dat$pred, dat$y)
mat1 <- acast(dat[, c('trt','block','pred')], trt~block,
             id.var=c('trt','block'), value.var='pred')
}

## Not run:
# Another method to estimate missing values via PCA
require("pcaMethods") # From Bioconductor, not CRAN
```

```

m2 <- pca(mat0, method="nipals", center=FALSE, nPcs=3)
# mat2 <- m2@scores
mat2 <- tcrossprod(m2@scores, m2@loadings)

# Compare
ord <- c("0", "N", "K", "P", "NK", "NP", "KP", "NKP")
print(mat0[ord,], na.print=".")
round(mat1[ord,], 2)
round(mat2[ord,], 2)

# SVD with 3 components recovers original data better
sum((mat0-mat1)^2, na.rm=TRUE)
sum((mat0-mat2)^2, na.rm=TRUE) # Smaller SS => better fit

## End(Not run)

```

yates.oats

Yield of oats in a split-plot experiment

Description

The yield of oats from a split-plot field trial conducted at Rothamsted in 1931.

Varieties were applied to the main plots.

Manurial (nitrogen) treatments were applied to the sub-plots.

Each plot is 1/80 acre = 28.4 links * 44 links.

Field length: 18 rows * 28.4 links = 511 links

Field width: 4 plots * 44 links = 176 links.

The 'block' numbers in this data are as given in the Rothamsted Report. The 'grain' and 'straw' values are the actual pounds per sub-plot as shown in the Rothamsted Report. Each sub-plot is 1/80 acre, and a 'hundredweight (cwt)' is 112 pounds, so converting from sub-plot weight to hundredweight/acre needs a conversion factor of 80/112.

The 'yield' values are the values as they appeared in the paper by Yates, who used 1/4-pounds as the units (i.e. he multiplied the original weight by 4) for simpler calculations.

Format

row row

col column

yield yield in 1/4 lbs per sub-plot, each 1/80 acre

nitro nitrogen treatment in hundredweight per acre

gen genotype, 3 levels

block block, 6 levels

grain grain weight in lbs per sub-plot

straw straw weight in lbs per sub-plot

Source

Report for 1931. Rothamsted Experiment Station. Page 143. <http://www.era.rothamsted.ac.uk/eradoc/article/ResReport1931-141-159>

References

Yates, Frank (1935) Complex experiments, *Journal of the Royal Statistical Society Suppl 2*, 181-247. Figure 2. <http://doi.org/10.2307/2983638>

Examples

```
data(yates.oats)
dat <- yates.oats

## # Means match Rothamsted report p. 144
## require(dplyr)
## dat %>% group_by(nitro,gen) %>%
##   summarize(grain=mean(grain)*80/112,
##             straw=mean(straw)*80/112)

if(require(desplot)){
  # Experiment design & yield heatmap
  desplot(yield ~ col*row, dat,
          out1=block, num=nitro, col=gen,
          cex=1, aspect=511/176, # true aspect
          main="yates.oats")
}

# Roughly linear gradient across the field. The right-half of each
# block has lower yield. The blocking is inadequate!
require("lattice")
xyplot(yield ~ col|factor(nitro), dat,
       type = c('p', 'r'), xlab='col', as.table = TRUE,
       main="yates.oats")

## Not run:
require(lme4)
# Typical split-plot analysis. Non-significant gen differences
m3 <- lmer(yield ~ nitro * gen + (1|block/gen), data=dat)
# Residuals still show structure
xyplot(resid(m3) ~ dat$col, xlab='col', type=c('p','smooth'),
       main="yates.oats")

# Add a linear trend for column
m4 <- lmer(yield ~ col + nitro * gen + (1|block/gen), data=dat)
# xyplot(resid(m4) ~ dat$col, type=c('p','smooth'), xlab='col')

## Compare fits
AIC(m3,m4)
##   df   AIC
```

```

## m3 9 581.2372
## m4 10 557.9424 # Substantially better

## End(Not run)

# -----

## Not run:

# Marginal predictions from lsmeans package and asreml::predict

# --- nlme ---
require(nlme)
require(lsmeans)
# create unbalance
dat2 <- yates.oats[-c(1,2,3,5,8,13,21,34,55),]
m5l <- lme(yield ~ factor(nitro) + gen, random = ~1 | block/gen,
          data = dat2)

# --- asreml ---
require(asreml)
m5a <- asreml(yield ~ factor(nitro) + gen,
             random = ~ block + block:gen, data=dat2)

require(lucid)
vc(m5l)
vc(m5a)

lsmeans::lsmeans(m5l, "gen")
predict(m5a, data=dat, classify="gen")$predictions$pvals

## End(Not run)

# -----

## Not run:

# Demonstrate use of regress package, compare to lme

library(regress)
dat$nitrof <- factor(dat$nitro)
m6 <- regress(yield ~ nitrof + gen, ~block + I(block:gen), identity=TRUE,
             verbose=1, data=dat)
summary(m6)
## Variance Coefficients:
##           Estimate Std. Error
##  block           214.468    168.794
##  I(block:gen)    109.700     67.741
##  In              162.558     32.189

# ordinal causes clash with VarCorr
if(is.element("package:ordinal", search())) detach(package:ordinal)

```

```
m7 <- lme(yield ~ nitrof + gen, random = ~ 1|block/gen, data=dat)
lme4::VarCorr(m7)
##           Variance      StdDev
## block =    pdLogChol(1)
## (Intercept) 214.4716    14.64485
## gen =       pdLogChol(1)
## (Intercept) 109.6929    10.47344
## Residual    162.5591    12.74987

## End(Not run)
```

zuidhof.broiler

Daily weight, feed, egg measurements for a broiler

Description

Daily weight, feed, egg measurements for a broiler

Format

A data frame with 59 observations on the following 6 variables.

bw Body weight, grams

targetbw Target body weight, grams

adfi Average daily feed intake, grams

adg Average daily gain, grams

eggwt Egg weight, grams

age Age, days

Details

Using graphs like the one in the examples section, the authors discovered that a drop in body weight commonly occurs around the time of first egg production.

Source

Martin J. Zuidhof and Robert A. Renema and Frank E. Robinson, (2008). Understanding Multiple, Repeated Animal Measurements with the Help of PROC GPLOT. SAS Global Forum 2008, Paper 250-2008.

Used with permission of Martin Zuidhof.

Examples

```

data(zuidhof.broiler)
dat <- zuidhof.broiler

dat <- transform(dat, age=age/7) # Change days into weeks

# Reproducing figure 1 of Zuidhof et al.

# Plot using left axis
op <- par(mar=c(5,4,4,4))
plot(bw~age, dat, xlab="Age (weeks)", ylab="Bodyweight (g)",
      main="zuidhof.broiler",
      xlim=c(20,32), ylim=c(0,4000), pch=20)
lines(targetbw~age, subset(dat, !is.na(targetbw)), col="black")

# Now plot using the right axis
par(new=TRUE)
plot(adfi~age, subset(dat, !is.na(adfi)),
      xlab="", ylab="", xlim=c(20,32), xaxt="n", yaxt="n",
      ylim=c(-50,175), type="s", lty=2)
axis(4, at=c(-50,-25,0,25,50,75,100,125,150,175), col="red", col.axis="red")
mtext("Weight (g)", side=4, line=2, col="red")
lines(adg~age, subset(dat, !is.na(adg)), col="red", type="s", lty=1, lwd=2)
abline(h=c(0,52), col="red")
with(dat, segments(age, 0, age, eggwt, col="red"))

legend(20, -40, c("Body weight", "Target BW", "Feed/day", "Gain/day", "Egg wt"),
      bty="n", cex=.5, ncol=5,
      col=c("black","black","red","red","red"),
      lty=c(-1,1,2,1,1), lwd=c(1,1,1,2,1), pch=c(20,-1,-1,-1,-1))
par(op)

```

Index

*Topic **datasets**

acorsi.grayleafspot, 10
adugna.sorghum, 13
archbold.apple, 23
ars.earlywhitecorn96, 25
baker.barley.uniformity, 28
battese.survey, 33
beall.webworms, 35
beaven.barley, 37
becker.chicken, 38
besag.bayesian, 39
besag.beans, 41
besag.endive, 45
besag.triticale, 53
blackman.wheat, 55
bliss.borers, 57
bond.diallel, 58
box.cork, 59
bradley.multi.uniformity, 62
brandle.rape, 64
brandt.switchback, 66
broadbalk.wheat, 70
burgueno.alpha, 72
burgueno.rowcol, 76
burgueno.unreplicated, 78
butron.maize, 81
byers.apple, 84
caribbean.maize, 85
carlson.germination, 87
carmer.density, 88
cate.potassium, 90
chinloy.fractionalfactorial, 91
christidis.competition, 93
christidis.wheat.uniformity, 95
cleveland.soil, 96
cochran.beets, 97
cochran.crd, 99
cochran.eelworms, 100
cochran.factorial, 102
cochran.latin, 103
cochran.lattice, 104
cochran.wireworms, 105
connolly.potato, 106
cornelius.maize, 109
cramer.cucumber, 114
crampton.pig, 116
crossa.wheat, 118
cullis.earlygen, 124
darwin.maize, 128
day.wheat.uniformity, 130
denis.ryegrass, 132
depalluel.sheep, 133
digby.jointregression, 135
diggle.cow, 137
draper.safflower.uniformity, 139
eden.nonnormal, 147
eden.potato, 150
engelstad.nitro, 152
fan.stability, 153
federer.diagcheck, 154
fisher.barley, 160
fisher.latin, 163
foulley.calving, 164
fox.wheat, 166
garber.multi.uniformity, 167
gartner.corn, 169
giles.wheat, 174
gilmour.slatehall, 179
gomez.heteroskedastic, 184
gomez.nitrogen, 186
gomez.nonnormal1, 188
gomez.nonnormal2, 189
gomez.nonnormal3, 190
gomez.seedrate, 193
gomez.splitplot.subsample, 194
gomez.wetdry, 200
gotway.hessianfly, 202
goulden.barley.uniformity, 204

- goulden.eggs, 205
 goulden.latin, 206
 goulden.splitsplit, 207
 gregory.cotton, 210
 gumpertz.pepper, 212
 hanover.whitepine, 216
 harris.multi.uniformity, 218
 harris.wateruse, 221
 harrison.priors, 225
 harvey.lsmmeans, 226
 harville.lamb, 228
 hayman.tobacco, 229
 hazell.vegetables, 232
 heady.fertilizer, 234
 henderson.milkfat, 236
 hernandez.nitrogen, 238
 hessling.argentina, 240
 holland.arthropods, 245
 holshouser.splitstrip, 246
 holtsmark.timothy.uniformity, 248
 huehn.wheat, 249
 hunter.corn, 253
 ilri.sheep, 254
 immer.sugarbeet.uniformity, 256
 ivins.herbs, 258
 iyer.wheat.uniformity, 259
 jansen.apple, 261
 jansen.carrot, 262
 jansen.strawberry, 263
 jenkyn.mildew, 264
 john.alpha, 266
 johnson.blight, 270
 kadam.millet.uniformity, 271
 kalamkar.potato.uniformity, 273
 kalamkar.wheat.uniformity, 274
 kang.maize, 275
 kang.peanut, 277
 karcher.turfgrass, 278
 keen.potatodamage, 279
 kempton.slatehall, 285
 kenward.cattle, 288
 khin.rice.uniformity, 291
 kiesselbach.oats.uniformity, 292
 kreusler.maize, 293
 kristensen.barley.uniformity, 296
 lambert.soiltemp, 298
 lasrosas.corn, 299
 lavoranti.eucalyptus, 301
 lee.potatobligh, 302
 lehner.soybeanmold, 305
 lessman.sorghum.uniformity, 307
 li.millet.uniformity, 309
 lillemo.wheat, 310
 lin.superiority, 312
 lin.unbalanced, 314
 little.splitblock, 315
 lonnquist.maize, 317
 love.cotton.uniformity, 319
 lu.stability, 320
 lucas.switchback, 322
 lyon.potato.uniformity, 324
 lyons.wheat, 325
 masood.rice.uniformity, 326
 mclelland.corn.uniformity, 327
 mcleod.barley, 330
 mead.cauliflower, 332
 mead.cowpeamaize, 333
 mead.germination, 335
 mead.lamb, 336
 mercer.mangold.uniformity, 339
 minnesota.barley.weather, 342
 minnesota.barley.yield, 345
 montgomery.wheat.uniformity, 347
 moore.uniformity, 349
 nass.corn, 351
 nebraska.farmincome, 353
 nonnecke.corn.uniformity, 354
 nonnecke.peas.uniformity, 356
 odland.soybean.uniformity, 357
 omer.sorghum, 358
 onofri.winterwheat, 360
 ortiz.tomato, 361
 pacheco.soybean, 363
 parker.orange.uniformity, 364
 patterson.switchback, 366
 pearce.apple, 368
 perry.springwheat, 370
 piepho.cocksfoot, 373
 polson.safflower.uniformity, 374
 ratkowsky.onions, 375
 reid.grasses, 376
 ridout.appleshoots, 378
 ryder.groundnut, 382
 salmon.bunt, 383
 sawyer.multi.uniformity, 385
 senshu.rice, 387

- silva.cotton, 390
 sinclair.clover, 393
 smith.beans.uniformity, 395
 snedecor.asparagus, 399
 snijders.fusarium, 402
 stephens.sorghum.uniformity, 404
 steptoe.morex.pheno, 406
 stickler.sorghum.uniformity, 409
 stirret.borers, 411
 streibig.competition, 412
 tai.potato, 423
 theobald.barley, 426
 turner.herbicide, 431
 urquhart.feedlot, 432
 usgs.herbicides, 434
 vaneeuwijk.drymatter, 436
 vaneeuwijk.fusarium, 438
 vaneeuwijk.nematodes, 439
 vargas.txe, 442
 vold.longterm, 452
 vsn.lupin3, 454
 wallace.iowaland, 457
 walsh.cottonprice, 459
 wassom.brome.uniformity, 460
 waynick.soil, 461
 weiss.incblock, 464
 weiss.lattice, 465
 welch.bermutagrass, 467
 wheatley.carrot, 470
 wiebe.wheat.uniformity, 471
 wiedemann.safflower.uniformity, 473
 williams.barley.uniformity, 474
 williams.cotton.uniformity, 475
 yang.barley, 479
 zuidhof.broiler, 485
- aastveit.barley, 7, 16
 acorsi.grayleafspot, 10, 16
 adugna.sorghum, 12, 16
 agridat, 14
 allcroft.lodging, 16, 22
 archbold.apple, 16, 23
 ars.earlywhitecorn96, 16, 25
 australia.soybean, 16, 26
- baker.barley.uniformity, 14, 28
 batchelor.apple.uniformity, 14
 batchelor.apple.uniformity
 (batchelor.uniformity), 30
 batchelor.lemon.uniformity, 14
 batchelor.lemon.uniformity
 (batchelor.uniformity), 30
 batchelor.navel1.uniformity, 14
 batchelor.navel1.uniformity
 (batchelor.uniformity), 30
 batchelor.navel2.uniformity, 14
 batchelor.navel2.uniformity
 (batchelor.uniformity), 30
 batchelor.uniformity, 30
 batchelor.valencia.uniformity, 14
 batchelor.valencia.uniformity
 (batchelor.uniformity), 30
 batchelor.walnut.uniformity, 14
 batchelor.walnut.uniformity
 (batchelor.uniformity), 30
 battese.survey, 16, 33
 beall.webworms, 16, 35
 beaven.barley, 16, 37
 becker.chicken, 16, 38
 besag.bayesian, 16, 39
 besag.beans, 17, 41
 besag.elbatan, 17, 43
 besag.endive, 17, 45
 besag.met, 17, 48
 besag.triticale, 17, 53
 blackman.wheat, 17, 55
 bliss.borers, 17, 57
 bond.diallel, 17, 58
 box.cork, 16, 59
 bradley.multi.uniformity, 14, 62
 brandle.rape, 17, 64
 brandt.switchback, 16, 66
 bridges.cucumber, 17, 67
 broadbalk.wheat, 20, 70
 burgueno.alpha, 17, 72
 burgueno.rowcol, 17, 76
 burgueno.unreplicated, 17, 78
 butron.maize, 17, 80
 byers.apple, 20, 84
- caribbean.maize, 17, 85
 carlson.germination, 17, 87
 carmer.density, 17, 88
 cate.potassium, 20, 90
 chinloy.fractionalfactorial, 17, 91
 christidis.competition, 17, 93

- christidis.wheat.uniformity, *14, 95*
 cleveland.soil, *20, 96*
 cochran.beets, *17, 97*
 cochran.bib, *17, 98*
 cochran.crd, *17, 99*
 cochran.eelworms, *17, 100*
 cochran.factorial, *17, 102*
 cochran.latin, *17, 103*
 cochran.lattice, *17, 104*
 cochran.wireworms, *17, 105*
 connolly.potato, *17, 106*
 cornelius.maize, *17, 109*
 corsten.interaction, *17, 111*
 cox.stripssplit, *17, 112*
 cramer.cucumber, *17, 114*
 crampton.pig, *16, 116*
 crossa.wheat, *17, 118*
 crowder.seeds, *17, 120*
 cullis.earlygen, *17, 124*

 darwin.maize, *17, 128*
 day.wheat.uniformity, *14, 130*
 denis.missing, *17, 131*
 denis.ryegrass, *17, 132*
 depalluel.sheep, *16, 133*
 digby.jointregression, *17, 135*
 diggle.cow, *16, 137*
 draper.safflower.uniformity, *14, 139*
 durban.competition, *17, 141*
 durban.rowcol, *17, 142*
 durban.splitplot, *17, 145*

 eden.nonnormal, *17, 147*
 eden.potato, *17, 150*
 engelstad.nitro, *17, 152*

 fan.stability, *17, 153*
 federer.diagcheck, *17, 154*
 federer.tobacco, *17, 159*
 fisher.barley, *18, 160*
 fisher.latin, *18, 163*
 foulley.calving, *16, 164*
 fox.wheat, *18, 166*

 garber.multi.uniformity, *14, 167*
 gartner.corn, *15, 169*
 gathmann.bt, *18, 171*
 gauch.soy, *18, 173*
 giles.wheat, *18, 174*

 gilmour.serpentine, *18, 176*
 gilmour.slatehall, *18, 179*
 gomez.fractionalfactorial, *18, 181*
 gomez.groupssplit, *18, 183*
 gomez.heteroskedastic, *18, 184*
 gomez.multilocsplitplot, *18, 185*
 gomez.nitrogen, *18, 186*
 gomez.nonnormal1, *18, 188*
 gomez.nonnormal2, *18, 189*
 gomez.nonnormal3, *18, 190*
 gomez.rice.uniformity, *14, 192*
 gomez.seedrate, *18, 193*
 gomez.splitplot.subsample, *18, 194*
 gomez.splitssplit, *18, 196*
 gomez.stripplot, *18, 197*
 gomez.stripssplitplot, *18, 199*
 gomez.wetdry, *18, 200*
 gotway.hessianfly, *18, 202*
 goulden.barley.uniformity, *14, 204*
 goulden.eggs, *16, 205*
 goulden.latin, *18, 206*
 goulden.splitssplit, *18, 207*
 graybill.heteroskedastic, *18, 208*
 gregory.cotton, *18, 210*
 gumpertz.pepper, *18, 212*

 hanks.sprinkler, *18, 214*
 hanover.whitepine, *16, 216*
 harris.multi.uniformity, *14, 218*
 harris.wateruse, *16, 221*
 harrison.priors, *20, 225*
 harvey.lsmmeans, *16, 226*
 harville.lamb, *16, 228*
 hayman.tobacco, *18, 229*
 hazell.vegetables, *18, 232*
 heady.fertilizer, *18, 234*
 henderson.milkfat, *16, 236*
 hernandez.nitrogen, *18, 238*
 hessling.argentina, *20, 240*
 hildebrand.systems, *18, 242*
 holland.arthropods, *16, 245*
 holshouser.splitstrip, *18, 246*
 holtsmark.timothy.uniformity, *14, 248*
 huehn.wheat, *18, 249*
 hughes.grapes, *18, 251*
 hunter.corn, *18, 253*

 ilri.sheep, *16, 254*
 immer.sugarbeet.uniformity, *15, 256*

- ivins.herbs, *18, 258*
 iyer.wheat.uniformity, *15, 259*
- jansen.apple, *18, 261*
 jansen.carrot, *18, 262*
 jansen.strawberry, *18, 263*
 jenkyn.mildew, *18, 264*
 john.alpha, *18, 266*
 johnson.blight, *18, 270*
- kadam.millet.uniformity, *15, 271*
 kalamkar.potato.uniformity, *15, 273*
 kalamkar.wheat.uniformity, *15, 274*
 kang.maize, *18, 275*
 kang.peanut, *18, 277*
 karcher.turfgrass, *18, 278*
 keen.potatodamage, *19, 279*
 kempton.barley.uniformity, *15, 281*
 kempton.competition, *19, 283*
 kempton.rowcol, *19, 284*
 kempton.slatehall, *19, 285*
 kenward.cattle, *16, 288*
 khin.rice.uniformity, *15, 291*
 kiesselbach.oats.uniformity, *15, 292*
 kreusler.maize, *20, 293*
 kristensen.barley.uniformity, *15, 296*
- lambert.soiltemp, *20, 298*
 lasrosas.corn, *15, 299*
 lavoranti.eucalyptus, *16, 301*
 lee.potatobligh, *19, 302*
 lehner.soybeanmold, *19, 305*
 lessman.sorghum.uniformity, *15, 307*
 li.millet.uniformity, *15, 309*
 lillemo.wheat, *19, 310*
 lin.superiority, *19, 312*
 lin.unbalanced, *19, 314*
 little.splitblock, *19, 315*
 lonnquist.maize, *19, 317*
 love.cotton.uniformity, *15, 319*
 lu.stability, *19, 320*
 lucas.switchback, *16, 322*
 lyon.potato.uniformity, *15, 324*
 lyons.wheat, *19, 325*
- masood.rice.uniformity, *15, 326*
 mcclelland.corn.uniformity, *15, 327*
 mconway.turnip, *19, 328*
 mcleod.barley, *19, 330*
 mead.cauliflower, *19, 332*
 mead.cowpeamaize, *19, 333*
 mead.germination, *19, 335*
 mead.lamb, *16, 336*
 mead.strawberry, *19, 338*
 mercer.mangold.uniformity, *15, 339*
 mercer.wheat.uniformity, *15, 341*
 minnesota.barley.weather, *19, 342*
 minnesota.barley.yield, *19, 345*
 montgomery.wheat.uniformity, *15, 347*
 moore.bushbean.uniformity, *15*
 moore.bushbean.uniformity
 (moore.uniformity), *349*
 moore.carrots.uniformity, *15*
 moore.carrots.uniformity
 (moore.uniformity), *349*
 moore.fallcauliflower.uniformity, *15*
 moore.fallcauliflower.uniformity
 (moore.uniformity), *349*
 moore.polebean.uniformity, *15*
 moore.polebean.uniformity
 (moore.uniformity), *349*
 moore.springcauliflower.uniformity, *15*
 moore.springcauliflower.uniformity
 (moore.uniformity), *349*
 moore.sweetcorn.uniformity, *15*
 moore.sweetcorn.uniformity
 (moore.uniformity), *349*
 moore.uniformity, *349*
- nass.barley, *20*
 nass.barley (nass.corn), *351*
 nass.corn, *20, 351*
 nass.cotton, *20*
 nass.cotton (nass.corn), *351*
 nass.hay, *20*
 nass.hay (nass.corn), *351*
 nass.rice, *20*
 nass.rice (nass.corn), *351*
 nass.sorghum, *20*
 nass.sorghum (nass.corn), *351*
 nass.soybean, *20*
 nass.soybean (nass.corn), *351*
 nass.wheat, *20*
 nass.wheat (nass.corn), *351*
 nebraska.farmincome, *20, 353*
 nonnecke.corn.uniformity, *15, 354*
 nonnecke.peas.uniformity, *15, 356*

- odland.soybean.uniformity, *15, 357*
odland.soyhay.uniformity, *15*
odland.soyhay.uniformity
(odland.soybean.uniformity),
357
omer.sorghum, *19, 358*
onofri.winterwheat, *19, 360*
ortiz.tomato, *19, 361*
- pacheco.soybean, *19, 363*
parker.orange.uniformity, *15, 364*
patterson.switchback, *16, 366*
pearce.apple, *16, 368*
pearl.kernels, *21, 369*
perry.springwheat, *19, 370*
piepho.cocksfoot, *19, 373*
polson.safflower.uniformity, *15, 374*
- ratkowsky.onions, *19, 375*
reid.grasses, *19, 376*
ridout.appleshoots, *19, 378*
rothamsted.brussels, *19, 380*
ryder.groundnut, *19, 382*
- salmon.bunt, *19, 383*
sawyer.multi.uniformity, *15, 385*
senshu.rice, *19, 387*
shafii.rapeseed, *19, 388*
silva.cotton, *19, 390*
sinclair.clover, *19, 393*
smith.beans.uniformity, *15, 395*
smith.corn.uniformity, *15, 397*
snedecor.asparagus, *19, 399*
snijders.fusarium, *19, 402*
stephens.sorghum.uniformity, *15, 404*
steptoe.morex.genotype, *19*
steptoe.morex.genotype
(steptoe.morex.phenotype), *406*
steptoe.morex.phenotype, *19, 406*
stickler.sorghum.uniformity, *15, 409*
stirret.borers, *21, 411*
streibig.competition, *19, 412*
stroup.nin, *19, 413*
stroup.splitplot, *19, 418*
student.barley, *19, 421*
- tai.potato, *19, 423*
talbot.potato, *19, 424*
theobald.barley, *19, 426*
theobald.covariate, *20, 427*
thompson.cornsoy, *20, 428*
turner.herbicide, *21, 431*
- urquhart.feedlot, *16, 432*
usgs.herbicides, *21, 434*
- vaneeuwijk.drymatter, *20, 436*
vaneeuwijk.fusarium, *20, 438*
vaneeuwijk.nematodes, *20, 439*
vargas.txe, *20, 441*
vargas.wheat1, *20, 444*
vargas.wheat2, *20, 446*
verbyla.lupin, *20, 448*
vold.longterm, *20, 452*
vsn.lupin3, *20, 454*
- wallace.iowaland, *21, 457*
walsh.cottonprice, *20, 459*
wassom.brome.uniformity, *15, 460*
waynick.soil, *21, 461*
wedderburn.barley, *20, 462*
weiss.incblock, *20, 464*
weiss.lattice, *20, 465*
welch.bermudagrass, *20, 467*
Wheat2, *415*
wheatley.carrot, *20, 470*
wiebe.wheat.uniformity, *15, 471*
wiedemann.safflower.uniformity, *15, 473*
williams.barley.uniformity, *15, 474*
williams.cotton.uniformity, *15, 475*
williams.trees, *16, 477*
- yan.winterwheat, *20, 478*
yang.barley, *20, 479*
yates.missing, *20, 480*
yates.oats, *20, 482*
- zuidhof.broiler, *16, 485*