

# Package ‘agridat’

August 29, 2016

**Version** 1.12

**Date** 2015-06-28

**Title** Agricultural Datasets

**Type** Package

**Author** Kevin Wright

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

**Description** Datasets from books, papers, and websites related to agriculture. Example analyses are included. Includes functions for plotting field designs and GGE biplots.

**Imports** graphics, grDevices, grid, lattice, reshape2, stats

**Suggests** AER, agricolae (>= 1.2), betareg, car, coin, corrgram, effects, equivalence, FrF2, gam, gstat, HH, knitr, latticeExtra, lme4 (>= 1.1-5), lucid, mapproj, maps, MASS, MCMCglmm, mgcv, nlme, ordinal, pls, pscl, qtl, sp, survival, vcd

**License** GPL-2

**LazyData** yes

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

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

**VignetteBuilder** knitr

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2015-06-29 19:25:08

## R topics documented:

aastveit.barley . . . . .	6
adugna.sorghum . . . . .	7
agridat . . . . .	9
allcroft.lodging . . . . .	15

archbold.apple . . . . .	16
ars.earlywhitecorn96 . . . . .	18
australia.soybean . . . . .	19
baker.barley.uniformity . . . . .	21
batchelor.uniformity . . . . .	22
beall.webworms . . . . .	24
beaven.barley . . . . .	26
besag.bayesian . . . . .	27
besag.beans . . . . .	28
besag.elbatan . . . . .	30
besag.endive . . . . .	32
besag.met . . . . .	33
besag.triticale . . . . .	36
blackman.wheat . . . . .	38
bliss.borers . . . . .	39
bond.diallel . . . . .	41
box.cork . . . . .	42
brandle.rape . . . . .	44
brandt.switchback . . . . .	45
bridges.cucumber . . . . .	46
broadbalk.wheat . . . . .	48
burgueno.alpha . . . . .	49
burgueno.rowcol . . . . .	51
burgueno.unreplicated . . . . .	52
butron.maize . . . . .	54
byers.apple . . . . .	56
caribbean.maize . . . . .	58
carlson.germination . . . . .	59
carmer.density . . . . .	60
cate.potassium . . . . .	62
cleveland.soil . . . . .	64
cochran.beets . . . . .	65
cochran.bib . . . . .	66
cochran.crd . . . . .	67
cochran.eelworms . . . . .	68
cochran.factorial . . . . .	70
cochran.latin . . . . .	71
cochran.lattice . . . . .	72
cochran.wireworms . . . . .	73
connolly.potato . . . . .	74
cornelius.maize . . . . .	76
corsten.interaction . . . . .	78
cox.stripsplit . . . . .	79
crossa.wheat . . . . .	81
crowder.seeds . . . . .	83
cullis.earlygen . . . . .	87
darwin.maize . . . . .	89
denis.missing . . . . .	91

denis.ryegrass . . . . .	92
desplot . . . . .	93
digby.jointregression . . . . .	95
diggle.cow . . . . .	97
durban.competition . . . . .	98
durban.rowcol . . . . .	100
durban.splitplot . . . . .	102
eden.potato . . . . .	103
engelstad.nitro . . . . .	105
fan.stability . . . . .	106
federer.diagcheck . . . . .	107
federer.tobacco . . . . .	111
fisher.latin . . . . .	112
foulley.calving . . . . .	113
fox.wheat . . . . .	115
garber.multi.uniformity . . . . .	116
gathmann.bt . . . . .	117
gauch.soy . . . . .	119
gge . . . . .	120
gilmour.serpentine . . . . .	123
gilmour.slatehall . . . . .	125
gomez.fractionalfactorial . . . . .	126
gomez.groupsplit . . . . .	128
gomez.multilocsplitplot . . . . .	129
gomez.nitrogen . . . . .	130
gomez.rice.uniformity . . . . .	132
gomez.seedrate . . . . .	133
gomez.splitplot.subsample . . . . .	134
gomez.splitsplit . . . . .	135
gomez.stripplot . . . . .	137
gomez.stripsplitplot . . . . .	138
gotway.hessianfly . . . . .	139
goulden.barley.uniformity . . . . .	140
goulden.latin . . . . .	141
graybill.heteroskedastic . . . . .	142
gumpertz.pepper . . . . .	143
hanks.sprinkler . . . . .	145
harris.multi.uniformity . . . . .	147
harris.wateruse . . . . .	149
harrison.priors . . . . .	153
hayman.tobacco . . . . .	155
hazell.vegetables . . . . .	157
heady.fertilizer . . . . .	159
henderson.milkfat . . . . .	161
hernandez.nitrogen . . . . .	163
hessling.argentina . . . . .	165
hildebrand.systems . . . . .	167
holland.arthropods . . . . .	169

holshouser.splitstrip . . . . .	170
hughes.grapes . . . . .	172
hunter.corn . . . . .	173
ilri.sheep . . . . .	174
immer.sugarbeet.uniformity . . . . .	177
ivins.herbs . . . . .	178
jansen.apple . . . . .	180
jansen.carrot . . . . .	181
jansen.strawberry . . . . .	182
jenkyn.mildew . . . . .	183
john.alpha . . . . .	184
johnson.blight . . . . .	186
kalamkar.potato.uniformity . . . . .	188
kang.maize . . . . .	189
kang.peanut . . . . .	190
karcher.turfgrass . . . . .	191
keen.potatodamage . . . . .	193
kempton.barley.uniformity . . . . .	194
kempton.competition . . . . .	196
kempton.rowcol . . . . .	197
kempton.slatehall . . . . .	198
lamert.soiltemp . . . . .	200
lasrosas.corn . . . . .	201
lavoranti.eucalyptus . . . . .	203
lee.potatobligh . . . . .	204
li.millet.uniformity . . . . .	207
lonnquist.maize . . . . .	207
lucas.switchback . . . . .	209
lyon.potato.uniformity . . . . .	210
lyons.wheat . . . . .	211
mcconway.turnip . . . . .	212
mcleod.barley . . . . .	214
mead.cauliflower . . . . .	215
mead.cowpeamaize . . . . .	216
mead.germination . . . . .	218
mead.strawberry . . . . .	219
mercet.mangold.uniformity . . . . .	220
mercet.wheat.uniformity . . . . .	221
minnesota.barley.weather . . . . .	223
minnesota.barley.yield . . . . .	225
nass.corn . . . . .	227
nebraska.farmincome . . . . .	229
odland.soybean.uniformity . . . . .	230
ortiz.tomato . . . . .	231
pacheco.soybean . . . . .	233
panel.outlinelevelplot . . . . .	234
patterson.switchback . . . . .	235
pearce.apple . . . . .	236

pearl.kernels . . . . .	237
perry.springwheat . . . . .	239
piepho.cocksfoot . . . . .	241
ratkowsky.onions . . . . .	242
RedGrayBlue . . . . .	243
ridout.appleshoots . . . . .	244
rothamsted.brussels . . . . .	246
ryder.groundnut . . . . .	248
salmon.bunt . . . . .	249
senshu.rice . . . . .	250
shafii.rapeseed . . . . .	252
sinclair.clover . . . . .	253
smith.corn.uniformity . . . . .	255
snedecor.asparagus . . . . .	257
snijders.fusarium . . . . .	259
steel.soybean . . . . .	261
stephens.sorghum.uniformity . . . . .	263
steptoe.morex.pheno . . . . .	264
stirret.borers . . . . .	266
streibig.competition . . . . .	267
stroup.nin . . . . .	269
stroup.splitplot . . . . .	271
student.barley . . . . .	274
talbot.potato . . . . .	276
theobald.barley . . . . .	277
theobald.covariate . . . . .	279
thompson.cornsoy . . . . .	280
turner.herbicide . . . . .	282
vargas.txex . . . . .	284
vargas.wheat1 . . . . .	286
vargas.wheat2 . . . . .	288
verbyla.lupin . . . . .	290
vold.longterm . . . . .	292
vsn.lupin3 . . . . .	294
wallace.iowaland . . . . .	297
walsh.cottonprice . . . . .	298
wassom.brome1.uniformity . . . . .	299
waynick.soil . . . . .	301
wedderburn.barley . . . . .	302
weiss.incblock . . . . .	303
weiss.lattice . . . . .	305
welch.bermudagrass . . . . .	306
wiebe.wheat.uniformity . . . . .	309
williams.barley.uniformity . . . . .	311
williams.cotton.uniformity . . . . .	312
williams.trees . . . . .	313
yan.winterwheat . . . . .	314
yang.barley . . . . .	315

yates.missing . . . . .	316
yates.oats . . . . .	318
zuidhof.broiler . . . . .	319

<b>Index</b>	<b>322</b>
--------------	------------

---

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

---

## Description

Average height for 15 genotypes of barley in each of 9 years. A second matrix contains 19 covariates in each of 9 years.

## Format

A list of two matrices, height and covs. See details below.

## Details

Experiments were conducted at As, Norway.

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

The covs matrix contains 19 environmental covariates for each year.

ST	Sowing date
T1-T6	Avg temp (deg Celsius) in period 1, ..., 6
R1-R6	Avg rainfall (mm/day) in period 1, ..., 6
S1-S6	Daily solar radiation (ca/cm <sup>2</sup> ) in period 1, ..., 6

## Source

Aastveit, A. H. and Martens, H. (1986). ANOVA interactions interpreted by partial least squares regression. *Biometrics*, 42, 829–844.

Used with permission of Harald Martens.

## References

Chadoeuf, J and Denis, J B (1991). Asymptotic variances for the multiplicative interaction model. *J. App. Stat.* 18, 331–353.

## Examples

```
data(aastveit.barley)

# First, PCA of each matrix separately
```

```

Z <- aastveit.barley$height
Z <- sweep(Z, 1, rowMeans(Z))
Z <- sweep(Z, 2, colMeans(Z)) # Double-centered
sum(Z^2)*4 # Total SS
sv <- svd(Z)$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")

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

# Now, PLS relating the two matrices
if(require(pls)){
m1 <- pls(Z~U)
loadings(m1)
# Aastveit Fig 2a (genotypes), not rotated as they did
biplot(m1, which="y", var.axes=TRUE)
# Fig 2b, 2c (not rotated)
biplot(m1, which="x", var.axes=TRUE)
}

```

---

adugna.sorghum

*Sorghum yields at 3 locations across 5 years*


---

## Description

Sorghum yields at 3 locations across 5 years

## Format

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

gen Genotype factor, 28 levels

trial Trial factor, 2 levels

env Environment factor, 13 levels

yield Yield kg/ha

year Year, 2001-2005

loc Loc factor, 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) and 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.

Used with permission of Asfaw Adugna.

## Examples

```
data(adugna.sorghum)

# Genotype means match Adugna
dat <- adugna.sorghum

require(lattice)
levelplot(yield ~ env*gen, data=dat, main="adugna.sorghum gxe heatmap",
          col.regions=RedGrayBlue)

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
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 published papers and books relating to agriculture including field crops, tree crops, animal studies, and a few others.

**Details**

Abbreviations in the 'other' column include: xy = coordinates, pls = partial least squares, 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.navel1.uniformity	20 x 50	xy	
batchelor.navel2.uniformity	15 x 33	xy	
batchelor.valencia.uniformity	12 x 20	xy	
batchelor.walnut.uniformity	10 x 28	xy	
garber.multi.uniformity	45 x 6	xy, 2 years/crops	
gomez.rice.uniformity	18 x 36	xy	aov
goulden.barley.uniformity	20 x 20	xy	
harris.multi.uniformity	2 x 23	xy, 23 crops	corrgram
immer.sugarbeet.uniformity	10 x 60	xy, 3 traits	
kalamkar.potato.uniformity	6 x 96	xy	
kempton.barley.uniformity	7 x 28	xy	
li.millet.uniformity	6 x 100	xy	
lyon.potato.uniformity	34 x 6	xy	
mercier.mangold.uniformity	10 x 25	xy, 2 traits	
mercier.wheat.uniformity	25 x 20	xy, 2 traits	spplot
odland.soybean.uniformity	25 x 42	xy	
odland.soyhay.uniformity	28 x 55	xy	
smith.corn.uniformity	6 x 20	xy, 3 years	rgl
stephens.sorghum.uniformity	100 x 20	xy	
wassom.brome1.uniformity	36 x 36	xy	
wassom.brome2.uniformity	36 x 36	xy	
wassom.brome3.uniformity	36 x 36	xy	
wiebe.wheat.uniformity	12 x 125	xy	medianpolish, loess
williams.barley.uniformity	48 x 15	xy	loess
williams.cotton.uniformity	24 x 12	xy	loess

**Animals**

name	gen	loc	years	trt	other	model
brandt.switchback	10			2		aov
diggle.cow				4	ts	
fouley.calving					ordinal	polr
henderson.milkfat						nls,lm,glm,gam
holland.arthropods	5					
ilri.sheep	4		6		diallel	lmer, asreml
lucas.switchback	12			3		aov
patterson.switchback	12			4		aov
zuidhof.broiler					ts	

### Trees

name	gen	loc	reps	years	trt	other	model
archbold.apple	2		5		24	split-split	lmer
box.cork						repeated	radial, asreml
harris.wateruse	2				2	repeated	asreml,lme
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
adugna.sorghum	28	13		5			
aastveit.barley	15			9		yr*gen~yr*trt	pls
allcroft.lodging	32	7				percent	tobit
ars.earlywhitecorn96	60	9				6 traits	dotplot
australia.soybean	58	4		2		4-way, 6 traits	biplot
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
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					
crossa.wheat	18	25					AMMI
crowder.seeds	2		21		2		glm,jags
cox.stripssplit			4		3,4,2		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	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.latin			5		5	xy,lating	lm
fox.wheat	22	14					lm
gathmann.bt	2		8				TOST
gauch.soy	7	7	4	12			AMMI
gilmour.serpentine	108		3			xy, serpentine	asreml
gilmour.slatehall	25		6			xy	asreml
gomez.fractionalfactorial			2		6	xy	lm
gomez.groupsplit	45		3		2	xy, 3 gen groups	aov
gomez.multilocsplitplot	2	3	3			rsm1,nitro	aov, lmer
gomez.nitrogen			4		8		aov, contrasts
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.stripssplitplot	6		3			xy, nitro	aov
gotway.hessianfly	16		4			xy	lmer
goulden.latin			5		5	xy, latin	lm
graybill.heteroskedastic	4	13				hetero	
gumpertz.pepper		2				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
hughes.grapes			3	6	binomial	lmer, aod, glmm
hunter.corn		12		3	1	rsm1
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
lasrosas.corn			3	2	6	xy
lee.potatobligh	337		4	11		xy, ordinal, repeated
lonnquist.maize	11					diallel
lyons.wheat		12		4		
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
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
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
sinclair.clover				5,5	rsm2,mitzerlich	nls,rgl
snedecor.asparagus			4	4	4	split-plot, antedependence
snijders.fusarium	17			3	4	percent

steel.soybean	12	3	3				
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
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
vargas.wheat1	7			6		gen*yr~gen*trt, yr*gen~yr*cov	pls
vargas.wheat2	8	7				env*gen~env*cov	pls
vargas.tx				10	24	yr*trt~yr*cov	pls
verbyla.lupin	9	8		2		rsm1, xy, density	
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
yan.winterwheat	18	9					GGE,biplot
yang.barley	6	18					biplot
yates.missing			10		3^2	factorial	lm, pca
yates.oats	3		6			xy, nitro	lmer

**Time series**

name	years	trt	other	model
byers.apple				lme
broadbalk.wheat	74	17		
hessling.argentina	30		temp,precip	
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
------	-------

<a href="#">cate.potassium</a>	cate-nelson
<a href="#">cleveland.soil</a>	loess 2D
<a href="#">harrison.priors</a>	nls, prior
<a href="#">nebraska.farmincome</a>	choropleth
<a href="#">pearl.kernels</a>	chisq
<a href="#">stirret.borers</a>	lm, 4 trt
<a href="#">turner.herbicide</a>	glm, 4 trt
<a href="#">wallace.iowaland</a>	lm, choropleth
<a href="#">waynick.soil</a>	spatial, nitro/carbon

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 these data is: 1. Validating published analyses (reproducible research). 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 almost the only option (even for small datasets) for modelling AR1xAR1 residual variance structures. Commercial use of `asreml` requires a license: <http://www.vsnr.co.uk/downloads/asreml/>.

### Comments on the package structure

The original sources for these data use several different words to refer to genotypes including *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 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).

Most data are presented as data frames. In a few cases, the data are lists of matrices.

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.

Note: In the U.S., raw data are generally not subject to copyright. See <http://www.lib.umich.edu/copyright-office-publishing/copyrightability-charts-tables-and-graphs> and <http://sciencecommons.org/about/towards> for some discussion.

Nonetheless, substantial effort has been made to contact people to secure permission to include data (published within the past few decades) in this package.

Data produced from work of the United States government is not subject to copyright. [http://en.wikipedia.org/wiki/Copyright\\_status\\_of\\_work\\_by\\_the\\_U.S.\\_government](http://en.wikipedia.org/wiki/Copyright_status_of_work_by_the_U.S._government)

### Other resources

The `SDaA` package <http://cran.r-project.org/package=SDaA> contains county-level data from the United States Census of Agriculture, along with a vignette to illustrate survey sampling analyses.

lmtest:ChickEgg has time series of annual chicken and egg production in the United States 1930-1983.

Payne 2013 - Design and Analysis of Long-Term Rotation Experiments. <https://www.agronomy.org/publications/aj/pdfs/107/> Data and R code.

### Author(s)

Kevin Wright, kw.stat@gmail.com

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

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

### References

Box G. E. P. (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.

---

allcroft.lodging

*Lodging data from a multi-environment trial cereal crop*

---

### 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 factor, 1-7

gen genotype factor, 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.

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 factor, 5 levels  
 row Row  
 pos Position within each row  
 spacing Spacing between trees, 6,10,14 feet  
 stock Rootstock factor, 4 levels  
 gen Genotype factor, 2 levels  
 yield Yield total in 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  $\times$  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  $\text{stock}=0,1,6,7$  for Seedling,MM111,MM106,M0007 and  $\text{gen}=1,2$  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.
desplot(spacing~row*pos, dat, col=stock, cex=1, num=gen,
  main="archbold.apple")

if(require("lme4") & require("lucid")){

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

vc(m1) # Variances/means on 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

}

## require(HH)
```

```
## interaction2wt(yield~spacing+stock+gen, dat)
```

---

ars.earlywhitecorn96 *Early White Food Corn Performance Tests*

---

### **Description**

Early White Food Corn Performance Tests of 60 white hybrids.

### **Format**

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

loc loc factor, 9 levels

gen gen factor, 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~staklodge, dat, group=loc, auto.key=list(columns=3),
        main="ars.earlywhitecorn96")
splom(~dat[,3:9], group=dat$loc, auto.key=list(columns=3))

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

```

---

australia.soybean      *Australia soybeans*

---

**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 Factor with 8 levels, first character of location and last two characters of year

loc Four locations: Brookstead, Lawes, Nambour, RedlandBay

year Year: 1970 or 1971

gen Genotype factor of soybeans, 1-58

yield Yield, metric tons / hectare

height Height, in meters

lodging Lodging

size Seed size in millimeters

protein Protein in percentage

oil Oil, in 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.

### Examples

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

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
# dmat2 <- dmat[44:58,]
# require("lattice")
# parallelplot(dmat2[,c(2:6,1)], main="australia.soybean", horiz=FALSE)
```

---

baker.barley.uniformity

*Ten years of barley uniformity trials on same ground*

---

### Description

Ten years of barley uniformity trials on same ground

### Format

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

```
row row
col column
year year, numeric
yield yield, (pound/acre)
```

### Details

Ten years of uniformity trials were sown on the same ground.

### 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.

**Examples**

```

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

asp <- (161*3+30)/827 # True aspect ratio
desplot(yield~col*row|year, data=dat,
        main="baker.barley.uniformity heatmaps by year", aspect=asp)
dat2 <- aggregate(yield ~ row*col, data=dat, FUN=mean, na.rm=TRUE)
desplot(yield~col*row, data=dat2, main="baker.barley.uniformity 10-year average", aspect=asp)
# Note low yield in upper right, slanting to left a bit due to sandy soil.

# 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
require(reshape2)
mat <- acast(dat, row+col~year)
round(cor(mat, use='pair'),2)

```

---

batchelor.uniformity *Uniformity trials of fruit tree yields.*

---

**Description**

Uniformity trials of apples, lemons, oranges, and walnuts.

**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 (*Malus sylvestris*) 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.

### **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.

### **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.

### **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.

### **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.

### **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.

### **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=Lil6AAAAMAJ&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.

**Examples**

```
require(lattice)

# Apple
data(batchelor.apple.uniformity)
desplot(yield~col*row, data=batchelor.apple.uniformity,
        aspect=30/16,
        main="batchelor.apple.uniformity",
        ylab="Jonathon apple tree yields")

# Lemon
data(batchelor.lemon.uniformity)
desplot(yield~col*row, data=batchelor.lemon.uniformity,
        aspect=26/14,
        main="batchelor.lemon.uniformity",
        ylab="Eureka lemon tree yields")

# Navel1 (Arlington)
data(batchelor.navel1.uniformity)
dat <- batchelor.navel1.uniformity
desplot(yield~col*row, dat, aspect=50/20,
        main="batchelor.navel1.uniformity",
        ylab="Navel orange tree yields (Arlington)")

# Navel2 (Antelope)
data(batchelor.navel2.uniformity)
desplot(yield~col*row, data=batchelor.navel2.uniformity,
        aspect=33/15,
        main="batchelor.navel2.uniformity",
        ylab="Navel orange tree yields (Antelope)")

# Valencia
data(batchelor.valencia.uniformity)
desplot(yield~col*row, data=batchelor.valencia.uniformity,
        aspect=20/12,
        main="batchelor.valencia.uniformity",
        ylab="Valencia orange tree yields")

# Walnut
data(batchelor.walnut.uniformity)
desplot(yield~col*row, data=batchelor.walnut.uniformity,
        aspect=32/10,
        main="batchelor.walnut.uniformity",
        ylab="Seedling walnut yields")
```



**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 factor

trt treatment factor

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.

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://www.jstor.org/stable/1930285>.

**Examples**

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

# Match Beall table 1
with(dat, table(y, trt))
# histogram(~y|trt, data=dat, layout=c(1,4), as.table=TRUE)

# Visualize Beall tble 6. Block effects may exist, but barely.
desplot(y ~ col*row, data=dat, out1=block, out2=trt, num=trt, flip=TRUE,
        main="beall.webworms (count of worms)",
        col.regions=colorRampPalette(c("white", "black"))(9))

# 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)
```

---

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 factor

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 copy was cleaner.

### Source

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

### 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
desplot(yield ~ col*row, data=dat, aspect="iso",
        main="beaven.barley", text=gen)
```

---

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 factor

## 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.

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',
       main="besag.bayesian")

## Not run:
# Use asreml to fit a model with AR1 gradient in rows
require(asreml)
dat <- transform(dat, cf=factor(col), rf=factor(rrow))
m1 <- asreml(yield ~ -1 + gen, data=dat, random=~ar1v(rf))

# 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)

```

---

besag.beans

*Competition experiment in beans with height measurements*


---

**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 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.

## Source

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

# Add a covariate = excess height of neighbors
require(reshape2)
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 Genetic variety, factor with 50 levels  
 block Block/column (numeric)  
 row Row (numeric)

## Details

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

## 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.

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

Used with permission of David Higdon.

**Examples**

```

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

desplot(yield~block*row, dat, main="besag.elbatan wheat yields",
num=gen)

# Besag figure 1
# xyplot(yield~row|block, dat, groups=block, type=c('l'),
#       layout=c(1,3), main="besag.elbatan wheat yields")

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

# Add smooth trend with GAM
if(require(mgcv)){
  m3 <- gam(yield ~ -1 + gen + factor(block) + s(row), data=dat)
  plot(m3, residuals=TRUE, cex=3, main="besag.elbatan")
  # pred <- cbind(dat, predict(m3, data=dat, type="terms"))

  # Compare estimates
  p2 <- coef(m3)[1:50]
  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")
}

# 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(FALSE & require(gam)){
  m2 <- gam(yield ~ -1 + gen + factor(block) + lo(row), data=dat)
  plot.gam(m2)

  plot(m2, residuals=TRUE, se=TRUE,
       terms="lo(row)", 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$"lo(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")
}

```

---

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 ordinate

row row ordinate

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

### Details

In a field of endives, does each plant have footrot, or not?

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.

### 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://dx.doi.org/10.1198/1061860043029>



**Examples**

```

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

# Incidence map. Figure 2 of Friel and Pettitt
desplot(disease~col*row, dat, col.regions=c('lightgray','black'),
        main="besag.endive")

## 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

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

```

**Description**

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

**Format**

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

county County factory, 1-6

row Row ordinate

col Column ordinate

rep Rep factor, 1-3

block Incomplete block factor, 1-8

yield Yield

gen Genotype factor, 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.

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 fields.

**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.

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

desplot(yield ~ col*row|county, out1=rep, out2=block, dat,
        main="besag.met")

## Not run:
# Heteroskedastic variance model (separate variance for each variety)
# asreml takes 1 second, lme 73 seconds, SAS PROC MIXED 30 minutes

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

```

# asreml Using 'rcov' ALWAYS requires sorting the data
require(asreml)
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

# 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 = F))^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 ...

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

## End(Not run)

## Not run:

# 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
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, classify="county:gen")
p3 <- p3$pred$pval
bi3 <- gge(predicted.value ~ gen*county, data=p3, scale=FALSE)
windows()
# Very similar to the coefficient biplot
biplot(bi3, stand=FALSE, title="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/m<sup>2</sup>  
 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.

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. <http://www.jstor.org/stable/2531047>

### References

None.

### Examples

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

require(lattice)
# desplot(yield ~ col*row, data=dat, main="besag.triticale")

# 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)
xyplot(res ~ col|as.character(row), data=dat,
       as.table=TRUE, type="s", layout=c(1,3),
       main="besag.tritical")
```

```
## Not run:
# Besag uses an adjustment based on neighboring plots.
# This analysis fits the standard AR1xAR1 residual model
# Needs asreml package (not on CRAN)
require(asreml)
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)
```

---

blackman.wheat

*Yield for conventional and semi-dwarf wheat varieties*


---

### 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 factor

loc Loc factor

nitro Nitrogen fertilizer factor, low/high

yield Yield (g/m<sup>2</sup>)

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.

## 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)
require(reshape2)

# 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)

# 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")
```

---

bliss.borers

*Corn borer infestation under four treatments*


---

## 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, 4 levels  
 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.

Geoffrey Beall. 1940. The Fit and Significance of Contagious Distributions when Applied to Observations on Larval Insects. *Ecology*, 21, 460-474. Page 463. <http://www.jstor.org/stable/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'))
  distplot(dat2$borers, type = "nbinomial",
           main="bliss.borers neg binomialness plot")
  # Better way is a rootogram
  g1 <- goodfit(dat2$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 factor  
male Male parent factor  
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.

**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.
```

---

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
N weight of cork deposit (centigrams), north direction
E east direction
S south direction
W west direction
```

**Source**

C.R. Rao (1948) *Tests of significance in multivariate analysis*. *Biometrika*, 35, 58-79.

**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

require(lattice)
splom(dat[,2:5], 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:
## Each tree is one line
require(plotrix)
radial.plot(dat[, 2:5], 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(dat)))

require(reshape2)
dat$tree <- factor(dat$tree)
d2 <- melt(dat)
names(d2) <- c('tree','dir','y')

require(asreml)
d2 <- d2[order(d2$tree, d2$dir), ]

# Unstructured covariance matrix
m1 <- asreml(y~dir, data=d2, rcov=~tree:us(dir, init=rep(200,10)))
## Note: 'rcor' is a personal function to extract the correlation
## round(rcor(m1)$dir, 2)
##      N      E      S      W
## N 290.41 223.75 288.44 226.27
## E 223.75 219.93 229.06 171.37
## S 288.44 229.06 350.00 259.54
## W 226.27 171.37 259.54 226.00

# 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)
##      N      E      S      W
## N 290.42 209.46 291.82 228.44
## E 209.46 219.95 232.85 182.28
## S 291.82 232.85 350.00 253.95
## W 228.44 182.28 253.95 225.99

## End(Not run)

```

brandle.rape

*Three-way table of rape seed yields***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 factor, 5 levels

year year, numeric

loc loc factor, 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=TRUE,
        main="brandle.rape, yields per location", ylab="Genotype")

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

# 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.

if(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  sdcor
## 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
}

```

---

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 factor, A,B  
cow cow factor, 10 levels  
trt treatment factor, 2 levels  
period period factor, 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/](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
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+cov:group+period:group, data=dat)
anova(m1)
```

---

bridges.cucumber

*Cucumber yields in 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 Loc factor

gen Cultivar factor

row Row position

col Column position

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.

## 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))

desplot(yield~col*row|loc, data=dat, text=gen, cex=1,
        main="bridges.cucumber")

## Not run:
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, 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
```

```
## 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 factor

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
levelplot(grain~year*plot, dat, main="broadbalk.wheat: Grain", col.regions=RedGrayBlue)
levelplot(straw~year*plot, dat, main="broadbalk.wheat: Straw", col.regions=RedGrayBlue)
```

---

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 factor, 3 levels  
 block block factor, 12 levels  
 row row ordinate  
 col column ordinate  
 gen genotype factor, 16 levels  
 yield yield, numeric

**Details**

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

**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.alpha)
dat <- burgueno.alpha

desplot(yield~col*row, dat, main='burgueno.alpha', out1=rep, out2=block,
        text=gen, cex=1,shorten="none")

if(require(lme4) && require(lucid)){
  # Inc block model
  m1 <- lmer(yield ~ gen + (1|rep/block), data=dat)
  vc(m1) # 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
}

## Not run:
require(asreml)
# Inc block model
m2 <- asreml(yield ~ gen, data=dat, random = ~ rep/block)
m2$loglik # Matches Burgueno p. 26
m2$coef$fixed # Matches solution on p. 27

# AR1 x AR1 model plus linear row effect, random spline row
dat <- transform(dat, xf=factor(col), yf=factor(row))
```

```
dat <- dat[order(dat$xf,dat$yf),]  
m3 <- asreml(yield ~ gen + lin(yf), data=dat, random = ~ spl(yf),  
            rcov= ~ar1(xf):ar1(yf))  
m3$loglik # Matches row 8 of Table 1  
plot(variogram(m3), main="burgueno.alpha") # Figure 1  
  
## 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 factor, 2 levels

row row ordinate

col column ordinate

gen genotype factor, 64 levels

yield yield, tons/ha

## Details

A field experiment with two contiguous replicates in 8 rows, 16 columns.

## 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**

```

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

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

if(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

}

## Not run:
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)

```

---

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 factor, 281 levels

col column ordinate

row row ordinate

yield yield, tons/ha

**Details**

A field experiment with 280 new genotypes. A check genotype is planted in every 4th column.

**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
desplot(yield ~ col*row, data=dat, main="burgueno.unreplicated",
        col=check, text=gen, cex=.75)

## Not run:
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$oddcol <- factor(dat$col
# The modulus operator throws a bug, so do it the hard way.
dat$oddcol <- factor(dat$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)

```

---

butron.maize

---

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


---

## 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), Pontecaldeas (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 Ord'as, 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.

Used with permission of Ana Butron.

## Examples

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

# require(lattice)
# densityplot(~yield|env, dat, layout=c(1,5))

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 Rcmdr 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

# 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) #

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, classify="gen")$pred$pvals
p4 <- predict(m4, 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

## End(Not run) # dontrun

```



**Description**

Measurements of the diameters of apples

**Format**

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

tree tree factor, 10 levels  
 apple apple factor, 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")

# Overall fixed linear trend, plus random intercept/slope deviations
# for each apple. Observations within each apple are correlated.
if(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
}
```

---

caribbean.maize      *Maize fertilization trial on Antigua and St. Vincent*

---

### Description

Maize fertilization trial on Antigua and St. Vincent.

### Format

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

isle Island factor, 2 levels

site Site factor

block Block, factor

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. <https://dl.sciencesocieties.org/publications/cs/abstracts/23/5/CS0230050882>.

**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

*Maize yield-density model*


---

**Description**

Maize yield-density model.

**Format**

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

gen Hybrid factor, 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

**Examples**

```
data(carmer.density)
dat <- carmer.density
dat$gen <- factor(dat$gen, levels=c('Hy2x0h7', 'WF9xC103', 'R61x187-2',
                                   'WF9x38-11', 'WF9xB14', 'C103xB14',
                                   '0h43xB37', 'WF9xH60'))
# xyplot(yield~pop|gen, dat, pch=16, as.table=TRUE,
#        main="carmer.density")

# 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://www.crops.org/publications/sssaj/abstracts/35/4/SS0350040658>

## 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='wheat')
```

---

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)
## levelplot(resistivity ~ easting + northing, data = dat,
##           col.regions=RedGrayBlue,
##           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 = .015),
                  northing = seq(.150, 3.645, by = .015))
fit1 <- with(dat,
             predict(loess(resistivity~easting*northing, span = 0.25,
                           degree = 2), sg1))
```



```

require(lattice)
levelplot(fit1 ~ sg1$easting * sg1$northing,
          col.regions=RedGrayBlue,
          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)))
surface3d(seq(.15, 1.410, by = .015), seq(.150, 3.645, by = .015),
          fit1/100, alpha=0.9, col=rep("wheat", length(fit1)),
          front="fill", back="fill")

## 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://www.jstor.org/stable/2527917>

## Examples

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

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

---

cochran.bib

*Balanced incomplete block design in corn*

---

## Description

Balanced incomplete block design in corn

## Format

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

loc Location/block factor, 13 levels

gen Genotype/line factor, 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)
levelplot(yield~loc*gen, dat,
          col.regions=RedGrayBlue,
          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)

if(require(agricolae)){
  BIB.test(dat$loc, dat$gen, y=dat$yield)
}

```

---

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 Numeric, 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.

### Examples

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

# Field plan
desplot(Inf~col*row, data=dat, text=trt, cex=1, 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 in field

col Column in field  
 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
desplot(initial ~ col*row, data=dat, flip=TRUE, main="cochran.eelworms")

# final counts are strongly related to initial counts
require(lattice)
xyplot(final~initial|dose, data=dat, group=fumigant,
        main="cochran.eelworms", 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

desplot(diff~col*row, dat, text=operator, cex=1, 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.

**Source**

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

**References**

Walter Federer. Combining Standard Block Analyses With Spatial Analyses Under a Random Effects Model. Cornell Univ Tech Report BU-1373-MA.

**Examples**

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

desplot(y~row*col|rep, dat, text=trt, cex=2, main="cochran.lattice")

# Random rep,row,column model often used by Federer
if(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)
}

```

---

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 22cm by 13cm. 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

# desplot(worms ~ col*row, data=dat, text=trt, cex=1)

# 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 ***
## trt0  1.7147    0.2275   7.537 4.80e-14 ***
}

```

---

connolly.potato

*Potato yields in single-drill plots*

---

## 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.

Used with permission of Iain Currie.

## Examples

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

# Field plan
desplot(yield~col*row, data=dat, out1=rep, main="connolly.potato yields")
# 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
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

*Maize yields 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*.

## Examples

```

data(cornelius.maize)
dat <- cornelius.maize
# dotplot(gen~yield|env,dat) # We cannot compare genotype yields easily
# Subtract environment mean from each observation
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    *Corn yield in multi-environment trial*

---

## Description

The data is the average 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 factor with 20 levels

loc Location factor with 7 levels

yield yield in 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, Structuring Interaction in Two-Way Tables By Clustering, *Biometrics*, 1990, 46, 207–215. Table 1.

Used with permission of Jean-Baptiste Denis.

**Examples**

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

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 Rep factor, 4 levels  
soil Soil factor, 3 levels  
fert Fertilizer factor, 4 levels  
calcium Calcium factor, 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.

**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/statug/63033/HTML/default/viewer.htm#statug\\_anova\\_sect030.htm](http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_anova_sect030.htm)

**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,
                main="cox.stripsplit")
}

# 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 Kenword-Rogers approach for denominator d.f.
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)
```

---

crossa.wheat

*Wheat yields 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 factor

gen genotype factor

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.

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.

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. <https://www.crops.org/publications/cs/abstracts/53/6/2332>.

### 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 ...
```

---

 crowder.seeds

*Germination of Orobanche seeds from Crowder (1978)*


---

### 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

*Orobanche aegyptiaca* (commonly known as Egyptian broomrape) 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.

### 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.  
 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
m1.glm <- glm(cbind(germ,n-germ) ~ gen*extract,
  data=dat, 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)
##
## Estimate Std. Error t value Pr(>|t|)
## (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)

# Unlike the BUGS docs, we use names that we can actually interpret
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(vc(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 ordinate
col  column ordinate
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.

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.)

## 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://www.jstor.org/stable/2348066>

## References

<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)
desplot(yield ~ col*row, dat, main="cullis.earlygen (yield)",
        col="check", cex=1, flip=TRUE)

desplot(weed ~ col*row, dat, main="cullis.earlygen (weed)", flip=TRUE)

require(lattice)
bwplot(yield ~ weed, dat, horizontal=FALSE, main="cullis.earlygen")

## Not run:
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)

```

---

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 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\\_F1249/1876\\_CrossandSelfFertilisation\\_F1249.html](http://darwin-online.org.uk/converted/published/1881_Worms_F1357/1876_CrossandSelfFertilisation_F1249/1876_CrossandSelfFertilisation_F1249.html).

## References

R. A. Fisher, (1935) *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")

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
```

---

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 factor, 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.

### Examples

```
data(denis.missing)
dat <- denis.missing

# View missingness structure
require(reshape2)
acast(dat, env~gen, value.var='yield')
require(lattice)
levelplot(yield ~ gen*env, data=dat,
          col.regions=RedGrayBlue,
          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

```

---

denis.ryegrass                      *Plant strength 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 factor, 21 levels

loc Location factor, 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.

### References

Gower, J.C. and Hand, D.J., 1996. *Biplots*. Chapman and Hall.

### Examples

```

data(denis.ryegrass)
dat <- denis.ryegrass
anova(aov(strength ~ gen + loc, dat))

```

desplot

*Plot the layout/data of a field experiment***Description**

Plot the layout/data of a field experiment

**Usage**

```
desplot(form=formula(NULL ~ x + y), data,
        num, col, text, out1, out2,
        col.regions=RedGrayBlue, col.text=NULL, text.levels=NULL,
        out1.gpar=list(col = "black", lwd = 3),
        out2.gpar=list(col = "yellow", lwd = 1, lty = 1),
        at, ticks=FALSE, flip=FALSE, main, xlab, ylab,
        shorten="abb", show.key=TRUE, key.cex, cex=0.4, strip.cex=0.75, ...)
```

**Arguments**

form	A formula like <code>yield~x*y location</code>
data	A data frame
num	The column of the data to use for plotting numbers
col	Column of the data for the color of the number
text	Column to use for text labels
out1	Column to use for outlining
out2	Column to use for outlining
col.regions	Colors for regions
col.text	Colors for text strings
text.levels	Character strings to use instead of default 'levels'
out1.gpar	A list of graphics parameters for outlining. Can either be an ordinary <code>list()</code> or A call to <code>gpar()</code> from the grid package.
out2.gpar	Second level of outlining.
at	Breakpoints for the color ribbon
ticks	If TRUE, show tick marks row/column
flip	If TRUE, vertically flip the image
main	Main title
xlab	Label for x axis
ylab	Label for y axis
shorten	Method for shortening text in the key
show.key	If TRUE, show the key
cex	Expansion factor for text/number in each cell
key.cex	Left legend cex
strip.cex	Strip cex
...	Other

## Details

Note, not all lattice parameters are passed down to `xyplot`, but it is possible to make almost any change to the plot by assigning the `desplot` object to a variable and then edit the object by hand or use `update` to modify the object. Then print it manually. See the first example below.

Ryder (1981) discusses the need to examine the layout of the experiment design, and not just the data. This function provides a tool for plotting the layout of a field experiment and also the observed data.

Use `col.regions` to specify fill colors. This can either be a vector of colors or a function that produces a vector of colors. If the response variable is a factor and `col.regions` is a *function*, it will be ignored and the cells are filled with default light-colored backgrounds and a key is placed on the left. If the response variable is *numeric*, the cells are colored according to `col.regions`, and a ribbon key is placed on the right.

The default `shorten='abb'` will shorten the cell text using the `abbreviate` function. Other choices include `shorten='sub'` to use a 3-character substring, and `shorten='no'` for no shortening.

Note that two sub-plots with identical levels of the split-plot factor can be adjacent to each other by virtue of appearing in different whole-plots. To correctly outline the split-plot factor, simply concatenate the whole-plot factor and sub-plot factor together.

To call this function inside another function, you can hack thusly: `vr <- "yield"; vx <- "x"; vy <- "y" eval(parse(text=paste("desplot(", vr, "~", vx, "*", vy, ", ", data=yates.oats))))`

## Author(s)

Kevin Wright

## References

K. Ryder (1981). Field plans: why the biometrician finds them useful, *Experimental Agriculture*, 17, 243–256.

## Examples

```
# Show how to customize any feature. In this case, make the strips bigger.
data(besag.met)
d1 <- desplot(yield ~ col*row|county, besag.met, main="besag.met",
             out1=rep, out2=block, out2.gpar=list(col="white"), strip.cex=3)
d1 <- update(d1, par.settings = list(layout.heights=list(strip=2)))
print(d1)
# Or: d1$par.settings = list(layout.heights=list(strip=2))

# Show experiment layout
data(yates.oats)
desplot(yield ~ x+y, yates.oats, out1=block, out2=gen)

desplot(block ~ x+y, yates.oats, col=nitro, text=gen, cex=1, out1=block,
        out2=gen, out2.gpar=list(col = "gray50", lwd = 1, lty = 1))
```

```
# Example from Ryder.
data(ryder.groundnut)
gnut <- ryder.groundnut
m1 <- lm(dry~block+gen, gnut)
gnut$res <- resid(m1)
# Note largest positive/negative residuals are adjacent
desplot(res ~ col + row, gnut, text=gen, cex=1,
        main="ryder.groundnut residuals from RCB model")
```

---

digby.jointregression *Multi-environment trial illustrating joint regression*

---

### 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 factor, 10 levels  
env Environment factor, 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.

### References

Hans-Pieter Piepho, 1997. Analyzing Genotype-Environment Data by Mixed-Models with Multiplicative Terms. *Biometrics*, 53, 761-766.

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
round(tapply(dat$yield, dat$gen, mean),3)

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")

# Dibly Table 2, modified joint regression

```



```

round(betas,3)
# 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

round(thetas,3)+1.164-.515 # re-parameterize to match Digby
# 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

```

---

 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 *Mycobacterium* 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/viewer.htm#statug\\_glimmix\\_sect018.htm](http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_glimmix_sect018.htm)

## Examples

```
data(diggle.cow)
dat <- diggle.cow

# Figure 1 of Verbyla 1999
require(lattice)
if(require(latticeExtra)){
  useOuterStrips(xyplot(weight ~ day|iron*infect, dat, type='b',group=animal,
                        main="diggle.cow"))
}

# Scaling
dat <- transform(dat, time = (day-122)/10)

## Not run:

# Smooth for each animal. No treatment effects. Similar to SAS Output 38.6.9
require(asreml)

m1 <- asreml(weight ~ 1 + lin(time) + animal + animal:lin(time), data=dat,
             random = ~ animal:spl(time))
p1 <- predict(m1, 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)
```

**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`.

**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/Trial11.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))

desplot(yield ~ col*block, dat, out1=block, text=gen, col=wheel,
        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 factor, 2 levels

gen Genotype factor, 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.

## 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.

Retrieved from <ftp://ftp.bioass.sari.ac.uk/pub/maria>.

Used with permission of Maria Durban.

**Examples**

```

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

desplot(yield~bed*row, dat, out1=rep, 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
require(gam)
m1lo <- gam(yield ~ gen + lo(row, span=10/16) + lo(bed, span=9/34), data=dat)
new1 <- expand.grid(row=unique(dat$row),bed=unique(dat$bed))
new1 <- cbind(new1, gen="G001")
p1lo <- predict(m1lo, new=new1)
wireframe(p1lo~row+bed, new1, aspect=c(1,.5), main="Field trend") # Figure 5

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))

# Figure 7
v7a <- asreml.variogram(x=dat$bedf, y=dat$rowf, z=resid(m1a3))
wireframe(gamma ~ x*y, v7a, aspect=c(1,.5)) # Fig 7a

v7b <- asreml.variogram(x=dat$bedf, y=dat$rowf, z=resid(m1a2))
wireframe(gamma ~ x*y, v7b, aspect=c(1,.5)) # Fig 7b

v7c <- asreml.variogram(x=dat$bedf, y=dat$rowf, z=resid(m1lo))
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 factor, 4 levels

gen Genotype factor, 70 levels

fung Fungicide factor, 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.

Retrieved from <ftp://ftp.bioss.sari.ac.uk/pub/maria>.

Used with permission of Maria Durban.

### Examples

```
data(durban.splitplot)
dat <- 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
require(gam)
m2lo <- gam(yield ~ gen*fung + lo(row, bed, span=.082), data=dat)
new2 <- expand.grid(row=unique(dat$row), bed=unique(dat$bed))
new2 <- cbind(new2, gen="G01", fung="F1")
p2lo <- predict(m2lo, new=new2)
wireframe(p2lo~row+bed, new2, aspect=c(1,.5), main="Field trend")

# Table 5, variance components. Table 6, F tests
require(asreml)
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)

```

---

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 Factor for year/type

yield Yield in pounds per plot

block Block factor

```

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')
desplot(trt~col*row, d5a, text=yield, cex=1, shorten='no',
        main="eden.potato: 1925 qualitative")
anova(m5a <- aov(yield~trt+factor(row)+factor(col), d5a)) # table 2

# 1926 qualitative
d6a <- subset(dat, year=='1926a')
desplot(trt~col*row, d6a, text=yield, cex=1, shorten='no',
        main="eden.potato: 1926 qualitative")
anova(m6a <- aov(yield~trt+factor(row)+factor(col), d6a)) # table 4

# 1925 quantitative
d5 <- subset(dat, year=='1925b')
desplot(yield ~ col*row, d5, out1=block, text=trt, cex=1,
        main="eden.potato: 1925 quantitative")
# Trt 't' not defined, seems to be the same as 'a'

```



```

require(lattice)
dotplot(trt~yield|block, d5,
        main="eden.potato: 1925 quantitative")
anova(m5 <- aov(yield~trt+block, d5)) # table 6

# 1926 quantitative
d6 <- subset(dat, year=='1926b')
desplot(yield ~ col*row, d6, out1=block, text=trt, cex=1,
        main="eden.potato: 1926 quantitative")
anova(m6 <- aov(yield~trt+block, d6)) # table 7

# 1927 qualitative + quantitative
d7 <- droplevels(subset(dat, year==1927))
desplot(yield ~ col*row, d7, out1=block, text=trt, cex=1, col=ptype,
        main="eden.potato: 1927 qualitative + quantitative")

# Table 8. Anova, mean yield tons / acre
anova(m7 <- aov(yield~trt+block+ptype + ptype:potash, d7))
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

*Corn yield response to 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 factor, 2 levels

year Year, 1962-1966

nitro Nitrogen fertilizer kg/ha

yield Corn 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']
```

---

fan.stability

*Maize hybrids grown in China*


---

**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 factor  
 maturity maturity, days  
 year year  
 loc location factor  
 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://www.soils.org/publications/aj/abstracts/99/1/220>

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 factor, 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.

**Source**

Federer, Walter T. 1998. Recovery of interblock, intergradient, and intervariety information in incomplete block and lattice rectangle design experiments. *Biometrics*, 54, 471–481.

**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")
desplot(yield ~ col*row, dat, text=gen, show.key=FALSE,
        main="federer.diagcheck",
        shorten='no', col=check, cex=.8, col.text=c("yellow", "gray"))

# 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
}

# 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  sdcor
## 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:

# asreml
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)
```

---

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.

## References

- R. D. Cook and S. Weisberg (1999). *Applied Statistics 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
require(lattice)
levelplot(resid(m1) ~ row * block, data=dat,
          col.regions=RedGrayBlue(21), main="federer.tobacco")

# Row-column analysis. Treatment now significant
m2 <- lm(height ~ rowf + blockf + dosef, data=dat)
anova(m2)
```

---

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.

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	<i>Calving difficulty by calf sex and age of dam</i>
-----------------	--

---

**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 male calves with young dams. Differences between male/female calving 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. *Gent Sel Evol*, 28, 249–273.

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')))
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

## Note 1.0196-.3244 = 0.6952 matches Foulley's '2-3' threshold estimate

## Coefficients:
##              Value Std. Error t value
## sexF          -0.500605    0.01518  -32.982
## age2.0-2.5    -0.234609    0.01350  -17.383
## age2.5-3.0    -0.758325    0.01803  -42.059
## age3.0-3.5    -1.037794    0.01649  -62.920
## age3.5-4.0    -1.218294    0.02218  -54.926
## age4.0-4.5    -1.271138    0.01960  -64.856
## age4.5-5.0    -1.374209    0.02589  -53.080
## age5.0-8.0    -1.416112    0.01529  -92.589
## age8.0+      -1.454782    0.02088  -69.680
```

```
## sexF:age2.0-2.5 -0.003035  0.01933 -0.157
## sexF:age2.5-3.0  0.076677  0.02611  2.937
## sexF:age3.0-3.5  0.080657  0.02464  3.274
## sexF:age3.5-4.0  0.135774  0.03293  4.124
## sexF:age4.0-4.5  0.124303  0.02982  4.169
## sexF:age4.5-5.0  0.198897  0.03831  5.192
## sexF:age5.0-8.0  0.135524  0.02280  5.943
## sexF:age8.0+    0.131033  0.03185  4.114

## Intercepts:
##      Value      Std. Error t value
## S1|S2  0.3244  0.0107  30.2148
## S2|S3  1.0196  0.0111  91.9371

## Residual Deviance: 216176.90
## AIC: 216214.90

if(require(ordinal)){
  m2 <- clm(score ~ sex*age, data=dat, weights=count, link='probit')
  summary(m2) # same as polr model
  predict(m2) # probability of each category
}
```

---

fox.wheat

*Wheat yields of 22 varieties at 14 sites in Australia*


---

## 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*

---

**Description**

Uniformity trials of oat straw and wheat

**Format**

A data frame with 270 observations on the following 4 variables.

```
row row ordinate
col column ordinate
oats yield of oat hay
wheat yield of wheat grain
```

**Details**

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

desplot(oats ~ col*row, data=dat, flip=TRUE, tick=TRUE,
        main="garber.multi.uniformity oats")
desplot(wheat ~ col*row, data=dat, flip=TRUE, tick=TRUE,
        main="garber.multi.uniformity wheat")
with(dat, cor(oats, wheat)) # = .37 matches Garber
```

---

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 Maize type factor, Bt ISO

thysan Numeric

aranei Numeric

**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/presentation7Hothorn.pdf>

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)
require(reshape2)
d2 <- melt(dat)
bwdot=function(x,y,...){
  panel.xyplot(jitter(as.numeric(x)),y,...)
  panel.bwplot(x,y,...)
}
bwplot(value ~ gen|variable, d2,
       main="gathmann.bt", ylab="Insect abundance", panel=bwdot,
       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], epsilon=1))
lapply(th0[c("mean.diff", "ci.diff")], round, 2)

ar0 <- with(dat, tost(aranei[1:8], aranei[9:16], alpha=.05, epsilon=.4))
lapply(ar0[c("mean.diff", "ci.diff")], round, 2)

# ----- Non-parametric exact CI. Same result.
```

```

require(coin)
th1 <- wilcox_test(thysan ~ gen, data=dat, conf.int=TRUE, conf.level=0.9)
lapply(confint(th1), round, 2)

ar1 <- wilcox_test(aranei ~ gen, data=dat, conf.int=TRUE, conf.level=0.9)
lapply(confint(ar1), round, 2)

# ----- 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))

# ----- 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

*New York soybean yields, 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 factor, 7 levels  
env Environment factor, 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](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, <http://www.soils.org/publications/aj/abstracts/80/3/AJ0800030388>.

### 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$bipplot
  with(bip, plot(yield, PC1, type='n', main="gauch.soy -- AMMI biplot"))
  with(bip, text(yield, PC1, rownames(bip),
    col=ifelse(bip$type=="GEN", "darkgreen", "blue"),
    cex=ifelse(bip$type=="GEN", 1.5, .75)))
}
```

### Description

Fit a GGE (genotype + genotype \* environment) model and display the results.



**Usage**

```

gge(x, ...)

## S3 method for class 'formula'
gge(formula, data = NULL, gen.group = NULL, env.group = NULL, ...)

## S3 method for class 'matrix'
gge(x, center = TRUE, scale = TRUE, gen.group=NULL,
    env.group = NULL, comps=c(1,2), method = "svd", ...)

## S3 method for class 'gge'
plot(x, title = substitute(x), ...)

## S3 method for class 'gge'
biplot(x, title = substitute(x), subtitle = "",
    cex.gen = 0.6, cex.env = 0.5,
    col.gen = "darkgreen", col.env="orange3",
    pch.gen=1, lab.env = TRUE, comps = 1:2, flip = "auto", res.vec=TRUE, ...)

```

**Arguments**

x	A matrix or data.frame.
formula	A formula
data	Data frame
...	Other arguments
title	title
center	If TRUE, center values for each environment
scale	If TRUE, scale values for each environment
subtitle	subtitle
gen.group	genotype group
env.group	env group
cex.gen	character expansion for genotypes
cex.env	character expansion for environments
col.gen	color for genotypes
col.env	color for envts
pch.gen	plot character for genotypes
lab.env	label envts
comps	comps
flip	If "auto" then each axis is flipped so that the genotype ordinate is positively correlated with genotype means. Can also be a vector like c(TRUE,FALSE) for manual control.
res.vec	If TRUE, for each group, draw residual vectors from the mean of the locs to the individual locs
method	method used to find principal component directions

### Details

If there is replication in G\*E, then the replications are averaged together before constructing the biplot.

The singular value decomposition of  $x$  is used to calculate the principal components for the biplot. Missing values are NOT allowed.

The NIPALS algorithm can be used when there are missing data.

The argument 'method' can be either 'svd' for complete-data, or 'nipals' for missing-data.

### Value

A list of class gge containing:

method

center            Data centered?

scale             Data scaled?

gen.group        This is only used for plotting. If not NULL, this specifies a classification of genotypes into groups.

env.group        If not NULL, this specifies a classification of environments into groups. When using the formula method, then envGroup is the name of a column in the data. When using the matrix method, then envGroup must be a vector the same length as the number of columns in data, the contents of the vector contain the grouping information.

### Author(s)

Jean-Louis Laffont, Kevin Wright

### References

Jean-Louis Laffont, Kevin Wright and Mohamed Hanafi (2013). Genotype + Genotype x Block of Environments (GGB) Biplots. *Crop Science*, 53, 2332-2341. <https://www.crops.org/publications/cs/abstracts/53/6/2332>.

Kroonenberg, Pieter M. (1997) *Introduction to Biplots for GxE Tables*, Research Report 51, Centre for Statistics, The University of Queensland, Brisbane, Australia. <http://three-mode.leidenuniv.nl/document/biplot.pdf>

Yan, W. and Kang, M.S. (2003) *GGE Biplot Analysis*. CRC Press.

### See Also

The 'agricolae' package has functions for AMMI biplots.

**Examples**

```
# Example 1. Data is a data.frame in 'matrix' format
B <- matrix(c(50, 67, 90, 98, 120,
             55, 71, 93, 102, 129,
             65, 76, 95, 105, 134,
             50, 80, 102, 130, 138,
             60, 82, 97, 135, 151,
             65, 89, 106, 137, 153,
             75, 95, 117, 133, 155), ncol=5, byrow=TRUE)
rownames(B) <- c("G1", "G2", "G3", "G4", "G5", "G6", "G7")
colnames(B) <- c("E1", "E2", "E3", "E4", "E5")

m1 = gge(B)
plot(m1)
biplot(m1, title="Example biplot")

# crossa.wheat biplot

# Specify env.group as column in data frame
data(crossa.wheat)
dat2 <- crossa.wheat
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")
m4 <- gge(yield~gen*loc, dat2, env.group=eg, scale=FALSE)
biplot(m4, lab.env=TRUE, title="crossa.wheat")
```

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, numeric

row Row, numeric

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).

## 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.

## Examples

```
data(gilmour.serpentine)
dat <- gilmour.serpentine

desplot(yield~ col*row, data=dat, num=gen, out1=rep,
        #aspect = 6/.75,
        main="gilmour.serpentine")

# Extreme field trend. Blocking insufficient--needs a spline/smoother
# xyplot(yield~col, data=dat, main="gilmour.serpentine")

## Not run:
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, 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, 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)

```

---

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 a balanced lattice with 25 varieties in 6 replicates. The plot size was 1.5 meters by 4 meters, so the total area was 15 x 1.5 meters by 10 x 4 meters. The 'rep' labels are arbitrary (no rep labels appeared in the source data). Each row within a rep is an (incomplete) block.

**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.

**Examples**

```

data(gilmour.slatehall)
dat <- gilmour.slatehall

desplot(yield ~ col * row, dat, num=gen,
        out1=rep, main="gilmour.slatehall")

## Not run:
# Model 4 of Gilmour et al 1997
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)

```

---

gomez.fractionalfactorial

*Rice fractional factorial experiment*

---

**Description**

A 1/2 fractional factorial experiment with 6 treatments. 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 factor, 2 levels  
 block Block factor within rep, 2 levels  
 trt Treatment factor, levels (1) to abcdef  
 x Column position in the field  
 y Row position in the field

**Details**

Note: In Gomez, page 171 has (for example) treatment (1) in Rep 1, Block 2 and Rep 2, Block 1. Page 172 has treatment (1) in Block 1 of both reps. 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
desplot(yield~x*y, dat, main="gomez.fractionalfactorial - treatment & yield",
        text=trt, shorten="none", show.key=FALSE, cex=1)

# 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)

if(require(FrF2)){
  aliases(m1)
  MEPlot(m1, select=3:8,
         main="gomez.fractionalfactorial - main effects plot")
  #IAPlot(m1, select=3:8)
}
```

---

gomez.groupspl

*Rice group balanced split-plot design*

---

### **Description**

Group balanced split-plot design

### **Format**

A data frame with 270 observations on the following 7 variables.

x Column ordinate

y Row ordinate

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.groupspllit)
dat <- gomez.groupspllit

# Gomez figure 3.10. Obvious fert and group effects
desplot(group~x*y, dat, out1=rep, col=fert, text=gen,
        main="gomez.groupspllit")

# 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.multilocsplitplot
          Rice multilocation Split-Plot

```

---

**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 factor, 3 levels  
nitro Nitrogen in kg/ha  
rep Rep factor, 3 levels  
gen Genotype factor, 2 levels  
yield Yield in 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)

# F values are somewhat similar to Gomez Table 8.21
if(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
}

```

---

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 Rep  
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)
require(HH)
interaction2wt(nitro ~ rep/trt + trt*stage, data=dat,
               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.rice.uniformity *Rice uniformity yield test*

### Description

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.

### Format

A data frame with 648 observations on the following 3 variables.

yield Grain yield in g/m<sup>2</sup>

x Column position

y Row position

### 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

# Raw data plot
require(lattice)
desplot(yield ~ x*y, data=dat, aspect=38/20, main="gomez.rice.uniformity")

# 3x3 moving average. Gomez figure 12.1
require(reshape2)
dmat <- melt(dat, id.var=c('x','y'))
dmat <- acast(dmat, y~x)
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=38/20,
```

```

      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, 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, 4 levels

manage management factor, 2 levels

rep rep/block factor, 3 levels

sample subsample, 2 levels

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
require(HH)
interaction2wt(height ~ rep + time + manage, data=dat,
               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

```

**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 Blocking factor, 3 levels  
 nitro Nitrogen fertilizer, in kilograms/hectare  
 management Plot management  
 gen Variety of rice  
 yield Yield of rice  
 x Column position in the field  
 y 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)

desplot(nf~x*y, dat, out1=rep, col=management, num=gen, cex=1,
        main="gomez.splitsplit")
desplot(yield~x*y, dat, 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",
                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	<i>Rice strip-plot experiment</i>
-----------------	-----------------------------------

---

**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  
x Ordinate in the field  
y Ordinate in the field

**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.*

**Examples**

```
data(gomez.stripplot)
dat <- gomez.stripplot

# Gomez figure 3.7
desplot(gen~x*y, data=dat, out1=rep, num=nitro, cex=1,
        main="gomez.stripplot")

# Gertheiss figure 1
# 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)
```

---

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 factor, 3 levels  
 nitro Nitrogen fertilizer in kg/ha  
 gen Genotype factor, G1 to G6  
 x Ordinate in the field  
 y Ordinate in the field

### Details

Note, this is a superset of the the 'gomez.striplot' 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
desplot(gen~x+y, 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 + y, dat, x.relation="free")

## End(Not run)
```

---

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.

## 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.

## References

The GLIMMIX procedure. <http://www.ats.ucla.edu/stat/SAS/glimmix.pdf>.

**Examples**

```
## Could the spaMM package be useful here...???
```

```
data(gotway.hessianfly)
dat <- gotway.hessianfly
dat$prop <- dat$y / dat$n
desplot(prop~long*lat, dat, out1=block, text=gen, cex=1, shorten='no',
        main="gotway.hessianfly")
```

```
# Block random. See Glimmix manual, output 1.18.
# Note: (Different parameterization)
if(require(lme4)){
## l2 <- glmer(cbind(y, n-y) ~ gen + (1|block), data=dat, family=binomial,
##   control=glmerControl(check.nlev.gtr.1="ignore"))
## coef(l2)
}

## Not run:
require('asreml')
a2 <- asreml(prop ~ gen, data=dat, random=~block,
            family=asreml.binomial(),
            weights=n)

coef(a2)

## End(Not run)
```

---

goulden.barley.uniformity

*Barley uniformity trial*

---

**Description**

Barley uniformity trial

**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.

**Source**

C. H. Goulden, 1939. *Methods of statistical analysis*. Wiley. Page 18. <http://archive.org/details/methodsofstatist031744mbp>

**References**

See LeClerc, Leonard, Clark. *Field Plot Technique*. Page 27.

**Examples**

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

desplot(yield ~ col*row, data=dat, main="goulden.barley.uniformity")

# See LeClerc, 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.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**

G. Goulden (1952). *Methods of Statistical Analysis*. 2nd ed. Page 216.

**Examples**

```
data(goulden.latin)
dat <- goulden.latin
desplot(yield ~ col*row, dat, text=trt, cex=1, main="goulden.latin")

# Matches Goulden.
m1 <- lm(yield~ trt + factor(row) + factor(col), data=dat)
anova(m1)
```

---

graybill.heteroskedastic

*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 factor, 13 levels

gen Genotype factor, 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.

**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.
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

```

---

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 (1) or absence (0) 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.
desplot(disease ~ row*quadrat|field, data=dat,
        main="gumpertz.pepper disease presence", aspect=1,
        col.region=c('lightgray','black'))

# Soil water. Gumpertz fig 1b, 2b
desplot(water ~ row*quadrat|field, data=dat,
        main="gumpertz.pepper soil moisture", aspect=1,
        col.regions=c("#D9D9D9", "#B3B3B3", "#848484", "#5A5A5A", "#252525"),
        at=c(5,7.5,10,12.5,15,18))

# Leaf assay. Gumpertz fig 1c, 2c
desplot(leaf ~ row*quadrat|field, data=dat,
        main="gumpertz.pepper leaf assay", aspect=1,
        col.regions=c("#D9D9D9", "#BDBDBD", "#969696",
                     "#737373", "#525252", "#252525"))
        at=c(0,1,2,3,4,5,6)-.5

# 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)
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 factor

row Row ordinate

subplot Column ordinate

gen Genotype factor, 3 levels

yield Yield (tons/ha)

irr Irrigation level, 1..6

dir Factor for 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://www.soils.org/publications/sssaj/abstracts/44/4/SS0440040886>

## 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\\_sect038.htm](http://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_mixed_sect038.htm)

## Examples

```
data(hanks.sprinkler)
dat <- hanks.sprinkler

# The line sprinkler is vertical between subplots 6 & 7
desplot(yield~subplot*row, dat, out1=block, out2=irr, cex=1,
        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')
       })

## Not run:

## 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;

dat <- transform(dat, subf=factor(subplot),
                irrf=factor(irr))
dat <- dat[order(dat$block, dat$gen, dat$subplot),]

# Reproduced with asreml
require(asreml)
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)

```

---

harris.multi.uniformity

*Multi-year uniformity trials with different crops at Huntley*

---

## Description

Multi-year uniformity trials with different crops at Huntley

## 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

All yields are given in pound per plot, except sugar beets, which are given in tons per acre.

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 plats 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
require(reshape2)
mat <- acast(dat, series+plot~harv, value.var='yield')
if(require(corrgram)) corrgram(mat, main="harris.multi.uniformity")
# Compare to Harris 1928, table 4. More positive than negative correlations.
# densityplot(as.vector(cor(mat)), xlab="correlations",
#             main="harris.multi.uniformity")
```

```

# 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))
# True field shape: aspect=.84 (536 ft North-South by 639 ft E-W)
# 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.
desplot(yield ~ xord*plot|harv, data=d2, aspect=.84, flip=TRUE,
        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, main="harris.multi.uniformity fertility")

```

---

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)))

```

```

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

if(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

## Not run:

## 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

## End(Not run)

# ----- 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

## Not run:

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. <http://dx.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)
}

```

---

 hayman.tobacco

*Two reps of full diallel cross of Aztec tobacco*


---

## Description

Two reps of full diallel cross of Aztec tobacco

## Format

year year  
 block block factor, 2 levels  
 male male parent, 8 levels  
 female female parent  
 day mean flowering time (days)

## Details

The hayman.tobacco 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://www.jstor.org/stable/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. Same data in 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 distr? Maybe.
require(lattice)
densityplot(~day, data=dat2, main="hayman.tobacco",
            xlab="flowering date")

d1 <- subset(dat2, block=='B1')
d2 <- subset(dat2, block=='B2')

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))

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 four vegetable crops in six years*


---

## Description

Gross profit for four vegetable crops in six 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.

**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.

**References**

Carlos Romero, Tahir Rehman. (2003). *Multiple Criteria Analysis for Agricultural Decisions*. Elsevier.

**Examples**

```
data(hazell.vegetables)
dat <- hazell.vegetables

## Not run:

require(linprog)
# colMeans(dat[, -1])
# 252.8333 442.6667 283.8333 515.8333

# Maximize c'x for Ax=b
A <- rbind(c(1,1,1,1), c(25,36,27,87), c(-1,1,-1,1))
cvec <- c(253, 443, 284, 516) # avg profit per acre
colnames(A) <- names(cvec) <- cc(carrot,celery,cucumber,pepper)
rownames(A) <- c('land','labor','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

# Brute-force search for optimum allocation that minimizes year-to-year
# income variability.

# 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
mat <- matrix(runif(4*100000), ncol=4)
mat <- 200*sweep(mat, 1, rowSums(mat), "/")
profit <- mat
ix <- apply(profit, 1, function(x) cov.wt(as.data.frame(x), wt=probs)$cov)
ix <- which.max(ix)
mat[ix,] # Optimal planting allocation that minimizes the weighted variance
## carrot celery cucumber pepper
## 71.67002 27.90306 84.69966 15.72726

## 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 factor  
 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. [http://lib.dr.iastate.edu/cgi/viewcontent.cgi?filename=12&article=1032&context=ag\\_researchbulletins&type=additional](http://lib.dr.iastate.edu/cgi/viewcontent.cgi?filename=12&article=1032&context=ag_researchbulletins&type=additional)

**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://dl.sciencesocieties.org/publications/sssaj/abstracts/20/2/SS0200020240>

**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

## 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("nls", "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      *Corn response to 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 factor, 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. <http://www.crops.org/publications/aj/abstracts/100/5/1221>.

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),
       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.
lm1 <- lm(inc ~ 1 + nitro + I(nitro^2), data=dat1) # Note, 'poly' gives weird coeffs
if(require("latticeExtra")){
# xyplot(inc~nitro, dat1, main="hernandez.nitrogen") +
#   xyplot(fitted(lm1) ~ nitro, dat1, type='l')
}
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), "
  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
plot(profile(nls1, "th2"), conf = c(50, 80, 90, 95)/100)

# 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://dx.doi.org/10.1175/1520-0493\(1922\)50<302:RBTWAT>2.0.CO;2](http://dx.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      *Maize yields 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 factor, 2 levels

farm Farm factor, 14 levels

system Cropping system factor, levels LM LMF CCA CCAF

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://www.agronomy.org/publications/aj/abstracts/76/2/AJ0760020271>.

**References**

H. P. Piepho, 1998. Methods for Comparing the Yield Stability of Cropping Systems. *Journal of Agronomy and Crop Science*, 180, 193–213.

**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")

## Not run:
```

```

# Environmental variance model, unstructured correlations
require("asreml")
dat <- dat[order(dat$system, dat$farm),]
m1 <- asreml(yield ~ system, data=dat, rcov = ~us(system):farm)

# Means, table 5
p1 <- predict(m1, 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, 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 0.000002 NA NA bound
## system_LMF!variance 0.5304 0.208 2.5 pos

## End(Not run)

# 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
```

---

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).

Used with permission of John Holland.

**Examples**

```

data(holland.arthropods)
dat <- holland.arthropods

require(lattice)
contourplot(linyphiidae ~ col*row, dat, at=c(0,40,80,120,160,200), region=TRUE,
            col.regions=c('gray90','gray70','gray50','gray30','gray10'),
            main="holland.arthropods - linyphiidae counts in winter wheat")

## Not run:
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, numeric

pop population, numeric (thousand per acre)

yield yield, a numeric vector

row row number

col column number

**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
desplot(spacing ~ col*row, data=dat, out1=block, out2=cultivar,
        col=cultivar, text=pop, cex=.8, shorten='none', col.regions=c('wheat','white'),
        main="holshouser.splitstrip experiment design")
desplot(yield ~ col*row, data=dat, out1=block,
        main="holshouser.splitstrip")

# Overall main effects and interactions
require(HH)
interaction2wt(yield~cultivar*spacing*pop, dat,
              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;

## Now lme4. This design has five error terms--four are explicitly given.
if(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  sdcor
##   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
}

```

---

`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

`diseased` 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.

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.

**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::glmmPQL
require(MASS)
m3 <- glmmPQL(y ~ trt + block, data=bdat,
             random=~1|plot/vine, family=binomial)
m3

## End(Not run)

```

---

hunter.corn

*Corn yield response to nitrogen*

---

## Description

Corn yield response to nitrogen

## Format

A data frame with 54 observations on the following 4 variables.

ni tro nitrogen fertilizer, pound/acre

year year  
 loc location factor  
 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.

### 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](https://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 of lamb, 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%20study%204.1.htm> Retrieved Dec 2011.

**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)

if(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)
}

## Not run:
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, classify="year")$predictions

## End(Not run)
```

---

immer.sugarbeet.uniformity

*Uniformity study of sugarbeets with measurements of yield, sugar, purity*

---

## Description

A uniformity study of sugarbeets.

## 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

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.

## Examples

```
data(immer.sugarbeet.uniformity)
dat <- immer.sugarbeet.uniformity

# Immer numbers rows from the top
desplot(yield~col*row, data=dat, tick=TRUE, flip=TRUE,
        main="immer.sugarbeet.uniformity - yield")
desplot(sugar~col*row, data=dat, tick=TRUE, flip=TRUE,
        main="immer.sugarbeet.uniformity - sugar")
desplot(purity~col*row, data=dat, 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 factor, 6 levels

gen Genotype factor, 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	Lolium perenne	Irish perennial ryegrass
G02	Lolium perenne	S. 23 perennial ryegrass
G03	Dactylis glomerata	Danish cocksfoot

G04	Dactylis glomerata	S. 143 cocksfoot
G05	Phleum pratense	American timothy
G06	Phleum pratense	S. 48 timothy
G07	Festuca pratensis	S. 215 meadow fescue
G08	Poa trivialis	Danish rough stalked meadow grass
G09	Cynosurus cristatus	New Zealand crested dogstail
G10	Trifolium pratense	Montgomery late red clover
G11	Medicago lupulina	Commercial black medick
G12	Trifolium repens	S. 100 white clover
G13	Plantago lanceolata	Commercial ribwort plantain

### Source

Ivins, JD. (1952). Concerning the Ecology of *Urtica Dioica* L., *Journal of Ecology*, 40, 380-382.

### References

Ivins, JD (1950). Weeds in relation to the establishment of the Ley. *Grass and Forage Science*, 5, 237-242.

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.

### 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
```

---

`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.

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")

if(require(lme4)){
  # Tentative model. Needs improvement.
  m1 <- glmer(cbind(y,n-y) ~ gen + factor(inoculum) + (1|block),
             data=dat, family=binomial)
  summary(m1)
}
```

---

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.

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")

if(FALSE & require(lme4)){
  # 'FALSE' used here because CRAN wants < 5 seconds per example
  # 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

```

---

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 factor, 4 levels

category a factor with levels C1 C2 C3

count number of plants in each category

**Details**

There were 3 male strawberry plants and 4 different female strawberry plants that were crossed to create 12 populations.

The experiment had four blocks with 12 plots each (one for each population). In each plot, 9 or 10 plants were assessed for damage from fungus and rated as belonging to category C1, C2, or C3 (increasing damage).

Red core disease is caused by a fungus, *Phytophthora fragariae*.

**Source**

J. Jansen, 1990. On the statistical analysis of ordinal data when extravariation is present. *Applied Statistics*, 39, 75-84, Table 1. <http://www.jstor.org/stable/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')))

ftable(xtabs(count~male + female + category, data=dat))
mosaicplot(xtabs(count~male + female + category, data=dat),
           color=c("lemonchiffon1", "moccasin", "lightsalmon1", "indianred"),
           main="jansen.strawberry disease ratings",
           xlab="Male parent", ylab="Female parent")

# Todo: Need an example model
```

---

 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.

**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 factor

block Incomplete block factor



gen Genotype (variety) factor  
 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

John JA, Williams ER. (1995) *Cyclic and computer generated designs*, Chapman and Hall, London. Page 146.

### References

Piepho, H.P. and M"ohring, J. (2007), Computing heritability and selection response from unbalanced plant breeding trials, *Genetics*, 177, 1881–1888.

### 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, data=means, main="john.alpha - means for RCB, IB, Intra-block")

}
```

```
## Not run:
# Heritability calculation of Piepho & Moring, Example 1
require(asreml)
m3 <- asreml(yield ~ 1 + rep, data=dat, random=~ gen + rep:block)
sg2 <- summary(m3)$varcomp[1,'component']
vblup <- predict(m3, classify="gen")$pred$saved ^ 2
m3 <- asreml(yield ~ 1 + gen + rep, data=dat, random = ~ rep:block)
vblue <- predict(m3, classify="gen")$pred$saved ^ 2
# H^2 = .803
sg2 / (sg2 + vblue/2)
# H^2c = .809
1-(vblup / 2 / sg2)

## 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 Vakoeh, D.L. (1996). Potato late blight forecasting models for the semiarid environment of south-central Washington. *Phytopathology*, 86, 480–484.

**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")
```

---

kalamkar.potato.uniformity

*Uniformity trial of potatoes.*

---

### Description

Uniformity trial of potatoes.

### Usage

```
data(kalamkar.potato.uniformity)
```

### Format

A data frame with 576 observations on the following 3 variables.

row row number, 1-96

col column number, 1-6

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.

### Source

Kalamkar, R.J. (1932). Experimental Error and the Field-Plot Technique with Potatoes. *The Journal of Agricultural Science*, 22, 373-385.

### Examples

```
data(kalamkar.potato.uniformity)
dat <- kalamkar.potato.uniformity
```

```
# True field dimension is 132 foot-long row horizontal, 96*3=288 feet vertical
# Similar to figure 1 of Kalamkar
```

```
desplot(yield~col*row, data=dat, main="kalamkar.potato.uniformity",
        flip=TRUE, tick=TRUE, aspect=288/132)
```

---

kang.maize

*Maize yields at 4 locs in 3 years in Louisiana.*


---

## Description

Maize yields at 4 locs in 3 years in Louisiana.

## Usage

```
data(kang.maize)
```

## Format

```
gen genotype factor, 17 levels
env environment factor, 12 levels
yield yield, tonnes/ha
  environment factor, 13 levels
year year, 85-87
loc loc factor, 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.
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

*Peanut yields for 10 genotypes in 15 environments*


---

## 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 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)
require(reshape2)
round(acast(dat, env~gen, value.var='yield', fun=mean),2)

# GGE biplot of Kang, p. 82.
m1 <- gge(yield~gen*env, data=dat, scale=FALSE)
biplot(m1, flip=c(1,1), title="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/m<sup>2</sup>, level 2 is 5 g/m<sup>2</sup>.

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 a numeric vector

### 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.

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 Spring Barley in 1978*

---

**Description**

Uniformity trial of Spring Barley in 1978

**Format**

A data frame with 196 observations on the following 3 variables.

row Row position  
col Column position  
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.

**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.

**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

require(lattice)
desplot(yield~col*row, dat, main="kempton.barley.uniformity")
asp <- (28*14)/(7*5) # aspect ratio (rows * ft)/(cols * ft)
desplot(yield~col*row, dat, aspect=asp,
        main="kempton.barley.uniformity (true aspect)") # true view

dat <- transform(dat, xf = factor(col), yf=factor(row))

## Not run:

# Kempton estimated auto-regression coefficients b1=0.10, b2=0.91
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)

# 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)

```

---

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 factor, 36 levels

rep Rep factor, 3 levels

row Row

col Column

yield Yield in 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.

**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	<i>Wheat row-column experiment</i>
----------------	------------------------------------

---

**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))

desplot(yield~col*row|rep, dat, num=gen, out1=rep,
        main="kempton.rowcol")

# Model with rep, row, col as random. Kempton, page 62.
# Use "-1" so that the vcov matrix doesn't include intercept
if(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
}

```

---

kempton.slatehall

*Slate Hall Farm 1976 spring wheat*


---

**Description**

Yields for a trial at Slate Hall Farm in 1976.

**Format**

A data frame with 150 observations on the following 5 variables.

rep Rep factor, 6 levels

row Row

col Column

gen Genotype factor, 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.

**Source**

R A Kempton and P N Fox. (1997). *Statistical Methods for Plant Variety Evaluation*, Chapman and Hall. Page 84.

**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
dat <- transform(dat, xf=factor(col), yf=factor(row))

desplot(yield ~ col * row, dat, num=gen, out1=rep,
        main="kempton.slatehall")

# Incomplete block model of Gilmour et al 1995
if(require(lme4) & require(lucid)){
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

}

## Not run:
# Incomplete block model of Gilmour et al 1995
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, classify="gen")$predictions$pvals

## End(Not run)

```

---

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 as a number

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](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](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.corn

*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.corn)
```

### 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
  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=RedGrayBlue)
}

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 factor

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.

**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)
levelplot(height~loc*gen, dat, main="lavoranti.eucalyptus - GxE heatmap")

dat <- droplevels(subset(dat, origin=="Ravenshoe"))
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 ordinate  
 row row ordinate  
 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 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. Licensed via Open Database License. Data retrieved from <https://researchspace.auckland.ac.nz/handle/2292/5240>.

**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)
desplot(y ~ col*row|date, dat,
        main="lee.potatobligh - maps of blight resistance",
        ylab="Year of testing",
        layout=c(10,11),skip=skp)

# 1983 only. I.Hardy succumbs 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)))
```

---

li.millet.uniformity    *Uniformity trial of 600 millet plots.*

---

**Description**

Uniformity trial of 600 millet plots.

**Format**

A data frame with 600 observations on the following 3 variables.

col column

row row

yield yield (grams)

**Details**

Field was 100 ft x 100 ft. Plots were 15 feet long by 1 foot wide.

**Source**

Li, HW and Meng, CJ and Liu, TN. 1936. Field Results in a Millet Breeding Experiment. *Agronomy Journal*, 28, 1–15.

**Examples**

```
data(li.millet.uniformity)
dat <- li.millet.uniformity

mean(dat$yield) # matches Li et al.
desplot(yield~col*row, dat, main="li.millet.uniformity", aspect=100/100)
```

---

lonnquist.maize    *Half diallel of maize*

---

**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.

C. O. Gardner and S. A. Eberhart. 1966. Analysis and Interpretation of the Variety Cross Diallel and Related Populations. *Biometrics*, 22, 439-452. <http://www.jstor.org/stable/2528181>

**Examples**

```
data(lonquist.maize)
dat <- lonquist.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)
levelplot(yield ~ p1*p2, dat, col.regions=RedGrayBlue,
  main="lonquist.maize - yield of diallel cross")

# Calculate the F1 means in Lonquist, table 1
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, p1
              p2)
d2 <- droplevels(d2)
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)

```

---

lucas.switchback	<i>Switchback trial in dairy with three treatments</i>
------------------	--

---

### 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.

**References**

Sanders, WL and Gaynor, PJ. 1987. Analysis of Switchback Data Using Statistical Analysis System. *Journal of Dairy Science*, 70, 2186–2191.

**Examples**

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

# Create a numeric period variable
dat$p <- as.numeric(substring(dat$period,2))

# Need to use 'terms' to preserve the order of the model terms
# Really, cow(block), p:cow(block), period(block)
m1 <- aov(terms(yield ~ block + cow:block + p: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:p 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

**Format**

A data frame with 204 observations on the following 3 variables.

row Row, numeric

section Section, numeric

yield Yield, in pounds

### Details

In 1909, potatoes were harvested from uniform land. There were 34 rows, 34 inches apart. Each row was harvested in 6 sections, each being 72 feet, 7 inches long.

### Source

Lyon, T.L. (1911). Some experiments to estimate errors in field plat tests. *Proc. Amer. Soc. Agron.*, 3, 89-114. <https://www.crops.org/publications/aj/abstracts/3/1/AJ0030010089>

### References

Harris, J.A. (1920). Practical universality of field heterogeneity as a factor influencing plot yields. *Journal of Agricultural Research*, 19, 279-314. <http://books.google.com/books?id=jyEXAAAYAAJ&pg=PA279>

### Examples

```
data(lyon.potato.uniformity)
dat <- lyon.potato.uniformity

desplot(yield ~ row*section, dat, tick=TRUE,
        main="lyon.potato.uniformity")
```

---

lyons.wheat

*Yield 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 loc factor, 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'))
```

---

mconway.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 factor
date Planting date factor, levels 21Aug1990 28Aug1990
density Planting density, 1, 2, 4, 8 kg/ha
block Block factor, 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,
    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

*Yield and yield components for barley*


---

## 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://dx.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(reshape2)
d2 <- melt(d1, id.var='rate')
require(lattice)
xyplot(value~rate|variable, data=d2,
  scales=list(y=list(relation="free")), type='b',
  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

**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.

**References**

Mick O'Neill. Regression & Generalized Linear (Mixed) Models. SStatistical 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=TRUE, main="mead.cauliflower",
       xlab="degree days", ylab="number of leaves (by year)", ) +
  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 factor, 3 levels  
 nitro Nitrogen factor, 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, 1992. Statistical Methods in Agriculture and Experimental Biology, 2nd ed. Chapman and Hall. Page 364.

**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

*Seed germination with different temperatures/concentrations*


---

## Description

Seed germination with different temperatures/concentrations

## Format

A data frame with 64 observations on the following 5 variables.

temp Temperature regime  
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.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 factor, 4 levels  
 gen Genotype factor, 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")
desplot(yield~col*row, dat, text=gen, cex=1, out1=block, out2=sub,
  main="mead.strawberry")
```

---

mercer.mangold.uniformity

*Mercer and Hall's uniformity trial of mangolds in 1910*

---

### Description

Mercer and Hall's uniformity trial of mangolds in 1910

### Usage

```
data(mercer.mangold.uniformity)
```

**Format**

A data frame with 200 observations on the following 4 variables.

roots root yields, pounds  
 leaves leaf yields, pounds  
 col column, 1-10  
 row row, 1-20

**Details**

Grown in 1910. Each plot was 3 "drills", each 2.4 feet wide. Plots were 1/200 acres, 7.2 feet by 30.25 feet long

**Source**

Mercer, WB and Hall, AD, 1911. The experimental error of field trials *The Journal of Agricultural Science*, 4, 107-132. Table 1.

**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(mercer.mangold.uniformity)
dat <- mercer.mangold.uniformity

# Show field in true aspect ratio
asp = (7.25 * 20) / (30.25 * 10)

desplot(roots~col*row, data=dat,
        main="mercer.mangold.uniformity roots", aspect=asp)
desplot(leaves~col*row, data=dat,
        main="mercer.mangold.uniformity leaves", aspect=asp)

# xyplot(roots~leaves, data=dat)
```

---

```
mercer.wheat.uniformity
```

*Mercer and Hall's uniformity trial of wheat in 1910*

---

**Description**

Uniformity trial of wheat in 1910

**Format**

A data frame with 500 observations on the following 4 variables.

row row, 1-20

col column, 1-25

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.

Row 1 is the south edge, row 20 is the north. Column 1 is the west edge. The dimensions of each plot were approximately 3.2 meters by 2.5 meters.

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.

**References**

D. G. Rossiter (2014). Tutorial: Using the R Environment for Statistical Computing An example with the Mercer & Hall wheat yield dataset.

The 'spdep' package includes the grain yields (only) and spatial positions of plot centres in its example dataset 'wheat'.

**Examples**

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

require('lattice')
desplot(grain ~ col*row, dat, main="mercet.wheat.uniformity - grain yield")

xyplot(straw ~ grain, data=dat, type=c('p','r'), main="mercet.wheat.uniformity - regression")

# require(hexbin)
# hexbinplot(straw ~ grain, data=dat)

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

## Not run:
# 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")

## End(Not run)

```

---

minnesota.barley.weather

*Monthly weather at 6 sites in Minnesota 1927-1936.*

---

## 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 factor, 6 levels  
year Year, 1927-1936, numeric  
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://dx.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
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
}

```

---

minnesota.barley.yield

*Yields of barley at 6 sites in Minnesota 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 Numeric

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=0UfxLocnpKkC&pg=PA19>.

## 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://dx.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)))
```

## 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, numeric
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')
## levelplot(datm, aspect=.7, col.regions=RedGrayBlue,
##           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 factor

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)

  # Raw, county-wide incomes. Note the outlier Cuming county
  mapplot(stco ~ crop + animal, data = dat,
          scales = list(draw = FALSE), colramp=RedGrayBlue,
          main="nebraska.farmincome: income per county",
          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,
          scales = list(draw = FALSE), colramp=RedGrayBlue,
          main="nebraska.farmincome: income per square mile (percentile breaks)",
          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))
}

```

---

odland.soybean.uniformity

*Uniformity trials of soy hay and soybeans.*

---

**Description**

Uniformity trials of soy hay and soybeans.

**Format**

Data frames with 3 variables.

row row ordinate  
 col column ordinate  
 yield yield

### Details

Grown at West Virginia Experiment Station in 1925/1926. Soybeans were planted in rows 30 inches apart.

In 1925 the crop was harvested for forage, 42 rows, each 200 feet long. Yields of 8-foot plats recorded.

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 seems to partially missing.

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.

### Examples

```
data( odland.soyhay.uniformity)
dat1 <- odland.soyhay.uniformity
desplot(yield ~ col*row, dat1, flip=TRUE,
        aspect=200/(42*30/12), # true aspect ratio
        main="odland.soyhay.uniformity")
```

```
data( odland.soybean.uniformity)
dat2 <- odland.soybean.uniformity
desplot(yield ~ col*row, dat2, flip=TRUE,
        aspect = 232/(55*30/12), # true aspect
        main="odland.soybean.uniformity")
```

---

 ortiz.tomato

---

*Tomato weight/yield and environmental covariates in Latin America*


---

### Description

Tomato weight/yield and environmental covariates in Latin America

### Format

A list of three matrices, yield, weight, and covs. See details below.

## Details

The yield matrix contains average marketable fruit yield (t / ha) for 8 open-pollinated and 7 hybrid tomatoes in 18 environments. The weight matrix contains average fruit weight (g).

The environment locations and codes are: Estanzuela, Guatemala (E04), Baja Verapaz, Guatemala, (E05), Cogutepeque, El Salvador (E06), San Andres, El Salvador (E07), Comayagua, Honduras (E11), Valle de Sabaco, Nicaragua (E14), San Antonio de Belen, Costa Rica (E15), San Cristobal, Dominican Republic (E20), Constanza, Dominican Republic (E21), Palmira, Colombia (E27), La Molina, Peru (E40), Santiago, Chile (E41), Chillan, Chile (E42), Curacavi, Chile (E43), Colina, Chile (44), Belem, Brazil (E50), Caacupe, Paraguay (E51), Centeno, Trinidad Tobago (E53).

The covs matrix contains 16 environmental covariates.

Lat	Latitude
Long	Longitude
MxT	Max temperature (C)
MnT	Min temperature (C)
MeT	Mean temperature (C)
Prec	Precipitation (mm)
Day	Degree days (base 10)
pH	Soil pH
OM	Organic matter (percent)
P	Phosphorous (ppm)
K	Potassium (me/100 g)
ExN	Extra nitrogen (kg / ha)
ExP	Extra phosphorous (kg / ha)
ExK	Extra potassium (kg / ha)
Trim	Trimming (0/1)
Driv	Drivings (0/1)
Irr	Irrigation (0/1)
Dha	Days to harvest

## 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.

Used with permission of Rodomiro Ortiz.

## Examples

```
data(ortiz.tomato)
# Double-centered yield matrix
Y <- ortiz.tomato$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
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: plsr deletes observations with missing values
if(require(plsr)){
  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

*Soybean yields in Brazil.*


---

### Description

Yields of 18 soybean genotypes at 11 environments in Brazil.

### Format

gen Genotype factor, 18 levels  
 env Environment factor, 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.

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"))))
}

```

---

panel.outlinelevelplot

*Panel Function for desplot*

---

### Description

This is a panel function for desplot which fills cells with a background color and adds outlines around blocks of cells.

### Usage

```

panel.outlinelevelplot(x, y, z, subscripts, at, ...,
alpha.regions = 1, out1f, out1g, out2f, out2g)

```

### Arguments

x,y	Coordinates
z	Value for filling each cell
subscripts	For compatability
at	Breakpoints for the colors
...	Other
alpha.regions	Transparency for fill colors. Not well tested.
out1f, out2f	Factors to use for outlining.
out1g, out2g	Graphics parameters to use for outlining.

### Details

This is a panel function for desplot which fills cells with a background color and adds outlines around blocks of cells.

It does not add the text labels, numbers, or colors.

The rule for determining where to draw outlines is to compare the levels of the factor used for outlining. If bordering cells have different levels of the factor, then a border is drawn. 'NA' values are ignored (otherwise, too many lines would be drawn).

The code works, but is probably overkill and has not been streamlined.

**Value**

None

**Author(s)**

Kevin Wright. Derived from panel.levelplot.

---

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**

None.

**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.

**Examples**

```

data(patterson.switchback)
dat <- patterson.switchback

# Create a numeric period variable
dat$p <- as.numeric(substring(dat$period,2))

# Need to use 'terms' to preserve the order of the model terms
m1 <- aov(terms(y ~ cow + p: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:p        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 Factor for ear 8, 9, 10, 11

obs Factor for 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.biolbull.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      *Yields 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:yor, 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
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?

# Results from lme4 look dangerous. Try asreml. No problem.
require(asreml)
m2a <- asreml(y ~ 1 + yor, data=dat, random = ~ env + env:yor)

# Compare the two packages

require(lucid)
vc(m2)
##      grp      var1 var2   vcov   sdcor
##      env (Intercept) <NA> 1.311   1.145
##      env.1         yor <NA> 0.000006 0.002465
## Residual          <NA> <NA> 4.179    2.044

```



```

vc(m2a)
##          effect component  std.error z.ratio    con
##   env!env.var  0.000106  0.00000653   16  Boundary
## env:yor!env.var  0.000006  0.00000203   3.1  Positive
##      R!variance  4.181    0.2581      16  Positive

# 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)
m3a <- asreml(y ~ 1 + yorc, data=dat, random = ~ env + env:yorc)

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:yor!env.var  0.00063  0.000236  2.7  Positive
##      R!variance  3.551    0.2231   16  Positive

# Now lme4 succeeds. Rate of gain is 100*0.0549 = 5.49

fixef(m3)
## (Intercept)          yorc
## 5.87492444  0.05494464

## 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.

### 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

if(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
}
```

---

 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** loc factor, 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')
}
```

---

RedGrayBlue

*Function to create a Red-Gray-Blue palette*

---

**Description**

Function to create a Red-Gray-Blue palette

**Usage**

```
RedGrayBlue(n)
```

**Arguments**

n                    Number of colors to create

**Details**

Function to create a Red-Gray-Blue palette. Using gray instead of white allows missing values to appear as white (transparent).

**Value**

A vector of n colors.

**Examples**

```
pie(rep(1,11), col=RedGrayBlue(11))  
title("RedGrayBlue(11)")
```

---

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/GENMODZIP/roots.pdf](http://support.sas.com/rnd/app/examples/stat/GENMODZIP/roots.pdf)

## 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=16,
# 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)

# 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"))
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")
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 factor, 9 levels

block Block factor, 4 levels

**Details**

The block numbers are arbitrary, and may not match the original source.

**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")
desplot(yield~col*row, data=dat, num=trt, out1=block, cex=1,
        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 factor

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.

### 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
desplot(res1 ~ col + row, dat, text=gen, main="ryder.groundnut")

# Swap the dry yields for the 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, text=gen, main="ryder.groundnut")

```

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**

Salmon, SC, 1938. Generalized standard errors for evaluating bunt experiments with wheat. *Agronomy Journal*, 30, 647–663. Table 1.

**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. [na1dc.nal.usda.gov/download/IND43969050/PDF](http://na1dc.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)
levelplot(pct~gen+bunt,d2, col.regions=RedGrayBlue,
          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...

```

**Description**

Response of rice to solar radiation and temperature

**Format**

A data frame with 40 observations on the following 7 variables.

country Country factor  
 loc Location name  
 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.

**References**

Walter W. Piegorsch, 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

*Rapeseed yield multi-environment trial, 3 years***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 factor, 14 levels

rep Rep factor, 3 levels

gen Genotype factor, 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**

Electronic version retrieved from <http://www.uiweb.uidaho.edu/ag/statprog/ammi/yld.data>.

Used with permission of Benjamin Price.

**References**

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, *JABES*, 3, 335–345.

**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
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

```

---

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.

**Examples**

```
data(sinclair.clover)
dat <- sinclair.clover

require(lattice)
xyplot(yield~P|factor(S), dat, layout=c(5,1), main="sinclair.clover")

# 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")

## End(Not run)

```

---

smith.corn.uniformity *Uniformity trial of corn in three consecutive years*

---

## Description

Uniformity trial of corn in three consecutive years, 1895-1897. The same field/plots were used in all three years.

## Format

A data frame with 360 observations on the following 5 variables.

plot Plot, consistent across years

row Row  
col Column  
year Year, numeric. Last two digits of 1895, 1896, 1897  
yield Yield in bushels / acre

### Details

The data values are from Smith (1910) and the field map is from Harris (1920).

The following (edited) text is 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 experiment with corn. *Agronomy Journal*, 1, 84–89. <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. <http://books.google.com/books?id=jyEXAAAYAAJ&pg=PA279>

### Examples

```
data(snedecor.asparagus)
dat <- snedecor.asparagus

dat = transform(dat, year=factor(year))
desplot(yield~col*row|year, dat, layout=c(1,3),
        main="snedecor.asparagus: yield across years 1895-1987")

# Outliers are obvious
# require(lattice)
# xyplot(yield~row|factor(col), dat, groups=year,
# auto.key=list(columns=3), main="snedecor.asparagus")

## Not run:
require(rgl)
# A few odd pairs of outliers in column 6
plot3d(dat$col, dat$row, dat$yield, col=dat$year)

## 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.

### 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
if(require(lme4)){
m1 <- lmer(yield ~ trt + year + trt:year + (1|block) + (1|block:trt), data=dat)
}

## Not run:
# Split-plot with asreml
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

# Antependence 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)
##          [,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.

Used with permission of Fred van Eeuwijk.

**References**

Fred A van Eeuwijk. 1995. Multiplicative interaction in generalized linear models. *Biometrics*, 51, 1017-1032. <http://www.jstor.org/stable/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)
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]
# AMMI 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)

```

---

steel.soybean

*Soybean yields in North Carolina*


---

## Description

Soybean yields in a multi-environment trial in North Carolina.

**Format**

A data frame with 108 observations on the following 4 variables.

gen Genotype factor, 12 levels

loc Location factor, 3 levels

block Block factor, 3 levels

yield Yield, grams per plot

**Details**

The experiment was planted as an RCB with 3 blocks at each location.

**Source**

Steel, R.G.D. and Torrie, J.H., 1980. *Principles and procedures of statistics: a biometrical approach*, McGraw-Hill. Page 400.

**References**

Mick O'Neill, 2010. *Anova and REML: A Guide to Linear Mixed Models In An Experimental Design Context*.

**Examples**

```
data(steel.soybean)
dat <- steel.soybean
require(lattice)
dotplot(gen~yield|loc, dat, main="steel.soybean - yield by location")

if(require(lme4)){
# REML analysis as in O'Neill, p. 130.
m1 <- lmer(yield ~ loc + (1|gen) + (1|gen:loc) + (1|loc:block), data=dat)
# Are gen:loc effects significant? No.
m2 <- lmer(yield ~ loc + (1|gen) + (1|loc:block), data=dat)
anova(m1,m2)
# h^2 deviations
ranef(m2)$gen
}
```

---

stephens.sorghum.uniformity  
*Uniformity trial of sorghum silage*

---

### Description

Uniformity trial of sorghum silage, 100x20 plots.

### Format

A data frame with 2000 observations on the following 3 variables.

row row ordinate  
 col column / rod ordinate  
 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.

### 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.
# Similar to Stephens Figure 7. North at top. East at right.
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
desplot(yield ~ row2*col, data=d4, flip=TRUE, main="stephens.sorghum.uniformity",
        at=c(500,680,780,1000), col.regions=c('white','gray','gray30'))
```

---

steptoe.morex.pheno     *Phenotypic and genotypic data for a barley population of Steptoe x Morex*

---

### 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 factor, 16 levels

amylase Alpha Amylase (20 Deg Units)

diapow Diastatic Power (Deg 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 qtl 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](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.

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.

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.

## Examples

```
data(steptoe.morex.pheno)
dat <- steptoe.morex.pheno
# 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) # or just use plot()
}

## Not run:
# This is a very rudimentary example.

# Fit a simple multi-environment mixed model
require(asreml)
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
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 CDO497 67.5 -0.089 0.001 5.6

## End(Not run)

```

---

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 factor, 15 levels

trt treatment factor, 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 factor  
 bfwf 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 Schaberger 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

*Nebraska Intrastate Nursery field experiment***Description**

The yield data from an advanced Nebraska Intrastate Nursery (NIN) breeding trial conducted at Alliance, Nebraska, in 1988/89.

**Format**

**gen** Genotype factor, 56 levels

**rep** Replicate factor, 4 levels

**yield** The yield in bu/ac

**x** Column ordinate

**y** Row ordinate

**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. The blocks are not rectangular but partially overlap columns.

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 expresses the yield in bushels per acre. The results 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 x/y. To obtain the correct value of 'long', multiply 'x' by 1.2. To obtain the correct value of 'lat', multiply 'y' 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.

## 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.

## 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*

## See Also

Identical data (except for the missing values) are available in the nlme package as [Wheat2](#).

## Examples

```
data(stroup.nin)
dat <- stroup.nin
dat <- transform(dat, xf=factor(x), yf=factor(y))

# Show experiment layout
# Note: all "Buckskin" plots are near left side
desplot(yield~x*y, dat, out1="rep", num=gen, cex=1, main="stroup.nin")

require(nlme)
# Random block 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, correlation = corLin(form = ~ x+y,nugget=TRUE), data=dat,
  na.action=na.omit)

# Compare the estimates from the two methods
eff = data.frame(ranblock=fixef(lme1), spat = coef(lme2))
rownames(eff) <- gsub("gen", "", rownames(eff))
plot(eff$ranblock, eff$spat, xlim=c(13,37), ylim=c(13,37),
  main="stroup.nin", xlab="RCB (random block)", ylab="corLin", type='n')
text(eff$ranblock, eff$spat, rownames(eff), cex=0.5)
abline(0,1)

## Not run:
require(asreml)

# RCB analysis
dat.rcb <- asreml(yield ~ gen, random = ~ rep, data=dat,
  na.method.X="omit")
pred.rcb <- predict(dat.rcb,classify="gen")$predictions
```

```
# Two-dimensional AR1xAR1 spatial model
dat <- dat[order(dat$xf, dat$yf),]
dat.sp <- asreml(yield~gen, rcov=~ar1(xf):ar1(yf),data=dat)
pred.sp <- predict(dat.sp,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(16,37), ylim=c(16,37),
      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)
```

---

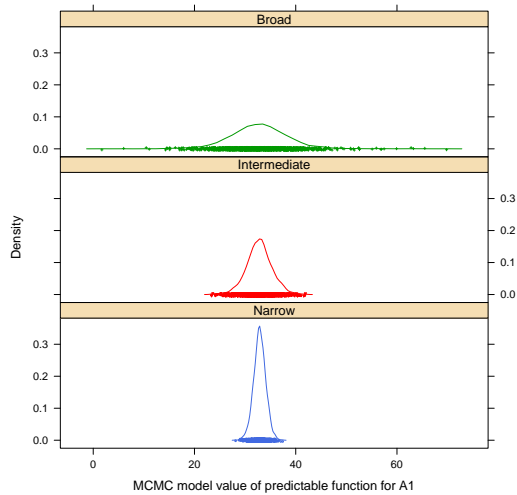
stroup.splitplot                      *Simulated split-plot experiment*

---

**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.



**Format**

y a numeric vector  
 rep Factor with 4 levels  
 b Factor with 2 levels  
 a Factor with 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.

**Examples**

```
data(stroup.splitplot)

## Not run:
dat <- stroup.splitplot

# ---- 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)
summary(m1)$var # Variance components match Stroup p. 41
#      Effect Estimate Std Err Z Ratio Con
# rep!rep.var   62.4    56.54    1.1 Pos
# a:rep!a.var   15.38   11.79    1.3 Pos
# R!variance    9.361   4.413    2.1 Pos

# Narrow space predictions
predict(m1, 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, 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, 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 %%% 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

*Barley yield of two varieties in a multi-environment trial*


---

## 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 (numeric), 1901-1906

farmer Factor of the farmer name

place Factor of the nearest town

district Factor of the geographical area  
 gen Genotype factor, Archer and Goldthorpe  
 yield Yield of the plot in '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.

### 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="")))

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)

# Kempton Table 9.6b
if(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"))
}

```

talbot.potato

*Potato yields and trait scores at 12 UK locations***Description**

Yield and 14 trait scores for each of 9 potato varieties at 12 locations in UK.

**Format**

A list of two matrices, `yield` and `char`. See details below.

**Details**

The `yield` matrix 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

The char matrix contains the scores of 14 traits on each of the 9 varieties.

### 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

```
data(talbot.potato)

if(require(pls)){

  char <- talbot.potato$char
  yield <- talbot.potato$yield

  # Transform columns to zero mean and unit variance
  char <- scale(char)
  yield <- scale(yield)

  m1 <- pls(yield ~ char, 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

*Barley yields at multiple locs, 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 factor

loc loc factor, 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.

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=TRUE, main="theobald.barley")
```

---

theobald.covariate      *Corn silage yields for Year \* Loc \* Variety with covariate*

---

### 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 factor, 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.

### 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")

# REML estimates (Means) in table 3 of Theobald 2002
if(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)
}
```

```

## 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("/files/theobald.bug", package="agridat"),
                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

*Corn & Soybean Yields, 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 (numeric) 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)

require(gam)
m1 <- gam(corn ~ lo(year) + lo(rain0,span=.8) + lo(rain7,span=.8) + lo(temp8,span=.8), data=dat.ia)
op <- par(mfrow=c(2,2))
plot(m1, residuals=TRUE, se=TRUE)
par(op)

## End(Not run)

```

**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

```

---

vargas.txe

*Treatment x environment interaction in agronomy trials*


---

## Description

Treatment x environment interaction in agronomy trials

## Format

A list of two matrices, yields and covs. See details below.

## Details

The yield matrix contains grain yields (kg/ha) of 24 treatments over 10 years. 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

The covs matrix contains environmental covariates for each year. The names of the covariates indicate:

MT	mean maximum temperature sheltered
mT	mean minimum temperature sheltered
mTU	mean minimum temperature unsheltered
PR	total monthly precipitation
SH	sun hours per day
EV	total monthly evaporation
D	December
J	January
F	February
M	March
A	April

## 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. Doi: 10.2134/agronj2001.934949x . <https://www.agronomy.org/publications/aj/articles/93/4/949>

Used with permission of Jose Crossa.

## Examples

```
data(vargas.txe)

require(lattice)
levelplot(vargas.txe$yields, col.regions=RedGrayBlue,
          main="vargas.txe", xlab="year", ylab="treatment")
```

```

Z <- vargas.txe$yields
Z <- sweep(Z, 1, rowMeans(Z))
Z <- sweep(Z, 2, colMeans(Z)) # Double-centered like AMMI
# Vargas figure 1
biplot(prcomp(Z, scale.=FALSE), main="vargas.txe")

## Not run:

# 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)

## End(Not run)

```

---

vargas.wheat1

*Durum 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.

## Format

A list of two data frames, `genvals` and `envvals`.

The `genvals` data frame contains 126 observations of genetic variables for each year\*rep\*gen:

year	Year, 1990-1995
rep	Replicate factor, 3 levels
gen	Genotype factor, 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 / m <sup>2</sup>
NGM	Grains / m <sup>2</sup>
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

The envals data frame has 17 variables for each of 6 years:

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	Sun hours per day December
SHJ	Sun hours per day January
SHF	Sun hours per day February
SHM	Sun hours per day March

## 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.

Data kindly provided by Jose Crossa.

## Examples

```
data(vargas.wheat1)

## Not run:
gvals <- vargas.wheat1$genvals
evals <- vargas.wheat1$envvals

require(pls)
```

```

# Yield as a function of gen variates
require(reshape2)
gv <- melt(gvals, id.var=c('gen','rep','year'))
Y0 <- subset(gv, variable=="yield")
Y0 <- acast(Y0, gen~year, mean)
Y0 <- sweep(Y0, 1, rowMeans(Y0))
Y0 <- sweep(Y0, 2, colMeans(Y0)) # GxE residuals
Y1 <- scale(Y0) # scaled columns
X1 <- subset(gv, variable!="yield")
X1 <- acast(X1, gen~variable, mean)
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 env variates
Y2 <- t(Y0)
X2 <- vargas.wheat1$envvals
rownames(X2) <- X2$year
X2 <- as.matrix(X2[,-1])
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

*Wheat yields for a multi-environment trial and environmental covariates*


---

### Description

The yield of 8 wheat genotypes was measured in 21 low-humidity environments. Each environment had 13 covariates recorded.

### Format

A list of two matrices, `yield` and `covs`. See below.

### Details

The yield matrix contains grain yields (kg/ha) for 8 wheat genotypes at 21 low-humidity environments grown during 1990-1994. The matrix is double-centered so that the rows and columns 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

The covs matrix has values of 13 covariates at the same 21 environments. The covariates are:

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

### 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.

Data kindly provided by Jose Crossa.

### Examples

```
data(vargas.wheat2)
dat <- vargas.wheat2
yield <- dat$yield
covs <- dat$covs

# The pls package centers, but does not (by default) use scaled covariates
# Vargas says you should
# yield <- scale(yield)
covs <- scale(covs)

require(pls)
m2 <- pls(yield ~ covs)
```

```

# Plot predicted vs observed for each genotype using all components
plot(m2)

# Loadings
# plot(m2, "loadings", xaxt='n')
# axis(1, at=1:ncol(covs), labels=colnames(covs), las=2)

# Biplots
biplot(m2, cex=.5, which="y", var.axes=TRUE,
       main="vargas.wheat2 - yield ~ covs") # 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

*Lupin yields for different varieties and locations, at many densities*


---

## Description

Yield of 9 varieties of lupin at different planting densities across 2 years and multiple locations.

## Format

gen Genotype factor, 9 varieties  
 site Site factor, 11 levels  
 rate seeding rate in plants/m<sup>2</sup>  
 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  
 loc Location factor

## 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 m<sup>2</sup>, and in 1992 was 20, 30, ..., 70 plants per m<sup>2</sup>. 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 an 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.

The year of release for the varieties is

Unicrop	1973
Illyarrie	1979
Yandee	1980
Danja	1986
Gungurru	1988
Yorrel	1989
Warrach	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.

Retrieved Oct 2010 from <http://www.blackwellpublishing.com/rss/Readmefiles/verbyla.htm>.

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
desplot(gen~col+row|site, data=dat, main="verbyla.lupin")

# 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", 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)))
} else {
  print(foo)
}
```

---

vold.longterm	<i>Long-term barley yields at different fertilizer levels</i>
---------------	---

---

**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, numeric

nitro nitrogen fertilizer, grams/m<sup>2</sup>

yield yield, grams/m<sup>2</sup>

**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 non-linear mixed effects model with an asymptotic curve.

**Source**

Vold, Arild (1998). A generalization of ordinary yield response functions. *Ecological modelling*, 108, 227-236.

**References**

Fernando E. Miguez (2008). Using Non-Linear Mixed Models for Agricultural Data.

**Examples**

```
data(vold.longterm)
dat <- vold.longterm

require(lattice)
require(nlme)
require(MASS) # for 'confint'

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 grouped by nitrogen", auto.key=list(columns=7))

## End(Not run)

## Not run:
# Global model
m1.nls <- nls(yield ~ SSasymp(nitro, max, int, lograte), data=dat)
summary(m1.nls)
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)
}

## End(Not run)

## Not run:
# Separate fit for each year. Overfitting with 3x19=57 params.
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)
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. Good 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","darkgreen","red"), points=FALSE, columns=3))
if(require(latticeExtra)){
  foo2 <- xyplot(pred~nitro|year, data=pdat2, type='l', col="red")
  foo3 <- xyplot(predf~nitro|year, data=pdat2, type='l', col="darkgreen",lwd=2)
  foo1 + foo2 + foo3
}

## # Income is maximized at about 15
## pdat2 <- transform(pdat2, income = predf*2 - 7*nitro)
## with(pdat2, xyplot(income~nitro))
```

---

vsn.lupin3

*Early generation lupin trial with 3 sites*


---

## 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 a factor with levels S1 S2 S3

col a numeric vector

row a numeric vector

gen a numeric vector

yield a numeric vector

## 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.vsnl.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)
})

desplot(yield~ col*row|site, dat, main="vsn.lupin3 - yield")
desplot(check~ col*row|site, dat,
  main="vsn.lupin3: check plot placement") # Site 1 & 2 used same randomization

## Not run:
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      1925 Iowa farmland values by county
```

---

**Description**

1925 Iowa farmland values by county

**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://www.jstor.org/stable/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))

if(require("latticeExtra")){ # for mapplot
  mapplot(polnm~fedval , data=dat, colramp=RedGrayBlue,
          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.

### 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.brome1.uniformity

*Uniformity trials of bromegrass.*

---

### Description

Uniformity trials of bromegrass.

### Usage

```
data(wassom.brome1.uniformity)
```

**Format**

A data frame with 1296 observations on the following 3 variables.

row row

col column

yield yield, pounds

**Details**

Experiments were conducted at Ames, Iowa. The response variable is forage yield in pounds of green weight. Each plot was 3.5 x 4 feet.

There are three datasets, hereafter referred to as 'brome1', 'brome2', 'brome3'.

Brome1 was broadcast seeded, harvested in 1950.

Brome2 was row planted, harvested in 1950.

Brome3 was broadcast seeded, harvested in 1951. This field contained a mixture of alfalfa and brome in about equal proportions.

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.

**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\\_researchbulletins/32/](http://lib.dr.iastate.edu/ag_researchbulletins/32/).

**Examples**

```
data(wassom.brome1.uniformity)
dat <- wassom.brome1.uniformity
desplot(yield~col*row, dat, main="wassom.brome1.uniformity", flip=TRUE)
```

```
dat <- wassom.brome2.uniformity
desplot(yield~col*row, dat, main="wassom.brome2.uniformity", flip=TRUE)
```

```
dat <- wassom.brome3.uniformity
desplot(yield~col*row, dat, main="wassom.brome3.uniformity", flip=TRUE)
```

---

`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
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)
splot(d1, zcol = "nitro", cuts=8, cex = 1.6,
```

```

main = "waynick.soil - Davis field - nitrogen",
col.regions = bpy.colors(8), key.space = "right")

# Variogram
v1 <- variogram(nitro~1, data=d1)
plot(v1, main="waynick.soil - Davis field - nitrogen") # Maybe hasn't reached sill

```

---

wedderburn.barley      *Percent of barley leaves affected by leaf blotch*

---

### 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://www.jstor.org/stable/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$pbins <- predict(m1, type="response")
dotplot(gen~pbins+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))

## 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)

```

**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.

**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/](http://lib.dr.iastate.edu/ag_researchbulletins/24/)

**Examples**

```
data(weiss.incblock)
dat <- weiss.incblock

# True aspect as shown in Weiss and Cox
desplot(yield~col*row, dat, text=gen, shorten='none', cex=.8, out1=block,
        main="weiss.incblock", xlab="Soybean yields", aspect=252/96)

## Not run:
# Standard inc block analysis used by Weiss and Cox
require(asreml)
m1 <- asreml(yield ~ gen + block , data=dat)
predict(m1, 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/](http://lib.dr.iastate.edu/ag_researchbulletins/24/)

**Examples**

```

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

desplot(yield~col*row|rep,dat, text=gen, aspect=3/16,
        main="weiss.lattice", xlab="Soybean yields", shorten="none", cex=.8)
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, 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://dl.sciencesocieties.org/publications/aj/abstracts/55/1/AJ0550010063>

## 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*n*100 + .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(vcov(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)

```

---

wiebe.wheat.uniformity

*Uniformity trial of wheat*

---

### Description

A uniformity trial of 1500 plots of wheat conducted in Idaho in 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.

Rows were 15 feet long, 1 foot apart.

Some authors recommend analyzing the square root of the yields.

## 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.

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

require("lattice")
# 125 rows = 125 feet tall. 12 cols * 15 feet = 180 feet wide.
desplot(yield~col+row, dat, aspect=125/180, flip=TRUE,
  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
xyplot(rowm~rep(1:8, length=125),
  main="wiebe.wheat.uniformity",
  xlab="Planter row unit", ylab="Row mean yield")
```

---

williams.barley.uniformity

*Yield from a barley uniformity trial*

---

### Description

Yield from a barley uniformity trial

### Format

A data frame with 720 observations on the following 3 variables.

row Row position

col Column position

yield 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.

### Source

Williams, ER and Lockett, 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.

### Examples

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

require("lattice")
asp = (15*5)/(48*.75) # true aspect
desplot(yield ~ col*row, dat, aspect=asp,
        main="williams.barley.uniformity")

# Smoothed contour/persp plot like Williams Fig 1b, 2b
dat$fit <- fitted(loess(yield~col*row, dat, span=.1))
contourplot(fit~col*row, data=dat, 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

*Lint yield from a cotton uniformity trial*


---

### Description

Lint yield from a cotton uniformity trial

### Format

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

row Row position

col Column position

yield Cotton 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.

### 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.

### Examples

```
data(williams.cotton.uniformity)
dat <- williams.cotton.uniformity

require("lattice")
asp = (12*12)/(24*1) # true aspect
desplot(yield ~ col*row, dat, aspect=asp, main="williams.cotton.uniformity")

# Smoothed contour/persp plot like Williams 1988 Fig 1a, 2a
dat$fit <- fitted(loess(yield~col*row, dat, span=.5))
contourplot(fit~col*row, data=dat, cuts=6, 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://aciar.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 factor

env Environment factor

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.

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.

### Examples

```
data(yan.winterwheat)
dat <- yan.winterwheat

require("reshape2")
mat <- melt(dat, id.var=c('gen','env'))
mat <- acast(mat, gen~env)

mat.sc <- scale(mat)

# Environment-standardized GGE biplot
# biplot(princomp(mat.sc), main="yan.winterwheat - GGE biplot")

m1 <- gge(yield ~ gen*env, data=dat)
biplot(m1, flip=c(0,1), title="yan.winterwheat - GGE biplot")
```

yang.barley

*Yield of 6 barley varieties at 18 locations in Alberta.***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.

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.

**Examples**

```
data(yang.barley)
dat <- yang.barley
require(reshape2)
dat <- acast(dat, gen~site, value.var='yield')

## Not run:

## For bootstrapping of a biplot, see the 'bbplot' and
## distfree.cr packages here:
## http://statgen.ualberta.ca/index.html?open=software.html

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
# this may not be working...

## 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 factor, 10 levels  
 y Numeric, 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("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

# 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-block experiment*


---

### Description

The yield of oats from a split-plot field trial using three varieties and four levels of manurial treatment. The experiment was laid out in 6 blocks of 3 main plots, each split into 4 sub-plots. The varieties were applied to the main plots and the manurial (nitrogen) treatments to the sub-plots.

### Format

```
block block factor with 6 levels
nitro nitrogen treatment in hundredweight per acre
gen genotype factor, 3 levels
yield yield in 1/4 lbs per sub-plot, each 1/80 acre.
x x ordinate
y y ordinate
```

### Source

Yates, Frank (1935) Complex experiments, *Journal of the Royal Statistical Society Suppl.* 2, 181–247.

### References

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth edition. Springer.

### Examples

```
data(yates.oats)
dat <- yates.oats

desplot(yield ~ x*y, dat, out1=block, text=gen, col=nitro, cex=1,
        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 ~ x|factor(nitro), dat, type = c('p', 'smooth'),
       xlab='x', span=1, as.table = TRUE, main="yates.oats")

# Typical split-plot analysis
```

```

if(require(lme4)){
m3 <- lmer(yield ~ nitro * gen + (1|block/gen), data=dat)
# Residuals still show structure
xyplot(resid(m3) ~ dat$x, xlab='x', type=c('p','smooth'),
        main="yates.oats")

# Add a linear trend for x
m4 <- lmer(yield ~ x + nitro * gen + (1|block/gen), data=dat)
# xyplot(resid(m4) ~ dat$x, type=c('p','smooth'), xlab='x')

## Compare fits
## AIC(m3,m4)
##   df      AIC
## m3  9 581.2372
## m4 10 557.9424
}

## Not run:

# Demonstrate the use of the lsmeans package and asreml::predict

# --- nlme ---
require(nlme)
require(lsmeans)
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) # warning.  FIXME
vc(m5a)

lsmeans(m5l, "gen")
predict(m5a, classify="gen")$predictions$pvals

## 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 (wk)", ylab="BW (g)",
      main="zuidhof.broiler, cage 1",
      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), col="red", type="s", lty=2)
axis(4, at=c(-50,-25,0,25,50,75,100,125,150,175), col="red")
mtext("Weight (g)", side=4, line=2)
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 /d", "Gain /d", "Egg wt"),  
      bty="n", cex=.65, 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**

adugna.sorghum, 7  
archbold.apple, 16  
ars.earlywhitecorn96, 18  
baker.barley.uniformity, 21  
beall.webworms, 24  
beaven.barley, 26  
besag.bayesian, 27  
besag.beans, 28  
besag.endive, 32  
besag.triticale, 36  
blackman.wheat, 38  
bliss.borers, 39  
bond.diallel, 41  
box.cork, 42  
brandle.rape, 44  
brandt.switchback, 45  
broadbalk.wheat, 48  
burgueno.alpha, 49  
burgueno.rowcol, 51  
burgueno.unreplicated, 52  
butron.maize, 54  
byers.apple, 56  
caribbean.maize, 58  
carlson.germination, 59  
carmer.density, 60  
cate.potassium, 62  
cleveland.soil, 64  
cochran.beets, 65  
cochran.crd, 67  
cochran.eelworms, 68  
cochran.factorial, 70  
cochran.latin, 71  
cochran.lattice, 72  
cochran.wireworms, 73  
connolly.potato, 74  
cornelius.maize, 76  
crossa.wheat, 81  
cullis.earlygen, 87  
darwin.maize, 89  
denis.ryegrass, 92  
digby.jointregression, 95  
diggle.cow, 97  
eden.potato, 103  
engelstad.nitro, 105  
fan.stability, 106  
federer.diagcheck, 107  
fisher.latin, 112  
foulley.calving, 113  
fox.wheat, 115  
garber.multi.uniformity, 116  
gilmour.slatehall, 125  
gomez.nitrogen, 130  
gomez.seedrate, 133  
gomez.splitplot.subsample, 134  
gotway.hessianfly, 139  
goulden.barley.uniformity, 140  
goulden.latin, 141  
gumpertz.pepper, 143  
harris.multi.uniformity, 147  
harris.wateruse, 149  
harrison.priors, 153  
hayman.tobacco, 155  
hazell.vegetables, 157  
heady.fertilizer, 159  
henderson.milkfat, 161  
hernandez.nitrogen, 163  
hessling.argentina, 165  
holland.arthropods, 169  
holshouser.splitstrip, 170  
hunter.corn, 173  
ilri.sheep, 174  
immer.sugarbeet.uniformity, 177  
ivins.herbs, 178  
jansen.apple, 180  
jansen.carrot, 181  
jansen.strawberry, 182  
jenkyn.mildew, 183

- john.alpha, 184
  - johnson.blight, 186
  - kalamkar.potato.uniformity, 188
  - kang.maize, 189
  - kang.peanut, 190
  - karcher.turfgrass, 191
  - keen.potatodamage, 193
  - kempton.slatehall, 198
  - lambert.soiltemp, 200
  - lasrosas.corn, 201
  - lavoranti.eucalyptus, 203
  - lee.potatobligh, 204
  - li.millet.uniformity, 207
  - lonnquist.maize, 207
  - lucas.switchback, 209
  - lyon.potato.uniformity, 210
  - lyons.wheat, 211
  - mcleod.barley, 214
  - mead.cauliflower, 215
  - mead.cowpeamaize, 216
  - mead.germination, 218
  - mercercorn.mangold.uniformity, 220
  - minnesota.barley.weather, 223
  - minnesota.barley.yield, 225
  - nass.corn, 227
  - nebraska.farmincome, 229
  - odland.soybean.uniformity, 230
  - ortiz.tomato, 231
  - pacheco.soybean, 233
  - patterson.switchback, 235
  - pearce.apple, 236
  - perry.springwheat, 239
  - piepho.cocksfoot, 241
  - ratkowsky.onions, 242
  - ridout.appleshoots, 244
  - ryder.groundnut, 248
  - salmon.bunt, 249
  - senshu.rice, 250
  - sinclair.clover, 253
  - snedecor.asparagus, 257
  - snijders.fusarium, 259
  - steel.soybean, 261
  - stephens.sorghum.uniformity, 263
  - steptoe.morex.pheno, 264
  - stirret.borers, 266
  - streibig.competition, 267
  - theobald.barley, 277
  - turner.herbicide, 282
  - vargas.txe, 284
  - vold.longterm, 292
  - vsn.lupin3, 294
  - wallace.iowaland, 297
  - walsh.cottonprice, 298
  - wassom.brome1.uniformity, 299
  - waynick.soil, 301
  - weiss.incblock, 303
  - weiss.lattice, 305
  - welch.bermudagrass, 306
  - wiebe.wheat.uniformity, 309
  - williams.barley.uniformity, 311
  - williams.cotton.uniformity, 312
  - yang.barley, 315
  - zuidhof.broiler, 319
- \*Topic **hplot**
- gge, 120
- aastveit.barley, 6, 10
  - adugna.sorghum, 7, 10
  - agridat, 9
  - allcroft.lodging, 10, 15
  - archbold.apple, 10, 16
  - ars.earlywhitecorn96, 10, 18
  - australia.soybean, 10, 19
  - baker.barley.uniformity, 9, 21
  - batchelor.apple.uniformity, 9
  - batchelor.apple.uniformity  
(batchelor.uniformity), 22
  - batchelor.lemon.uniformity, 9
  - batchelor.lemon.uniformity  
(batchelor.uniformity), 22
  - batchelor.navel1.uniformity, 9
  - batchelor.navel1.uniformity  
(batchelor.uniformity), 22
  - batchelor.navel2.uniformity, 9
  - batchelor.navel2.uniformity  
(batchelor.uniformity), 22
  - batchelor.uniformity, 22
  - batchelor.valencia.uniformity, 9
  - batchelor.valencia.uniformity  
(batchelor.uniformity), 22
  - batchelor.walnut.uniformity, 9
  - batchelor.walnut.uniformity  
(batchelor.uniformity), 22
  - beall.webworms, 10, 24
  - beaven.barley, 10, 26
  - besag.bayesian, 10, 27

- besag.beans, [10, 28](#)  
 besag.elbatan, [10, 30](#)  
 besag.endive, [10, 32](#)  
 besag.met, [10, 33](#)  
 besag.triticale, [10, 36](#)  
 biplot.gge (gge), [120](#)  
 blackman.wheat, [10, 38](#)  
 bliss.borers, [10, 39](#)  
 bond.diallel, [10, 41](#)  
 box.cork, [10, 42](#)  
 brandle.rape, [10, 44](#)  
 brandt.switchback, [10, 45](#)  
 bridges.cucumber, [10, 46](#)  
 broadbalk.wheat, [13, 48](#)  
 burgueno.alpha, [10, 49](#)  
 burgueno.rowcol, [10, 51](#)  
 burgueno.unreplicated, [10, 52](#)  
 butron.maize, [10, 54](#)  
 byers.apple, [13, 56](#)
- caribbean.maize, [11, 58](#)  
 carlson.germination, [11, 59](#)  
 carmer.density, [11, 60](#)  
 cate.potassium, [14, 62](#)  
 cleveland.soil, [14, 64](#)  
 cochran.beets, [65](#)  
 cochran.bib, [11, 66](#)  
 cochran.crd, [11, 67](#)  
 cochran.eelworms, [11, 68](#)  
 cochran.factorial, [11, 70](#)  
 cochran.latin, [11, 71](#)  
 cochran.lattice, [11, 72](#)  
 cochran.wireworms, [11, 73](#)  
 connolly.potato, [11, 74](#)  
 cornelius.maize, [11, 76](#)  
 corsten.interaction, [11, 78](#)  
 cox.stripsplit, [11, 79](#)  
 crossa.wheat, [11, 81](#)  
 crowder.seeds, [11, 83](#)  
 cullis.earlygen, [11, 87](#)
- darwin.maize, [11, 89](#)  
 denis.missing, [11, 91](#)  
 denis.ryegrass, [11, 92](#)  
 desplot, [93](#)  
 digby.jointregression, [11, 95](#)  
 diggle.cow, [10, 97](#)  
 durban.competition, [11, 98](#)  
 durban.rowcol, [11, 100](#)  
 durban.splitplot, [11, 102](#)  
 eden.potato, [11, 103](#)  
 engelstad.nitro, [11, 105](#)  
 fan.stability, [11, 106](#)  
 federer.diagcheck, [11, 107](#)  
 federer.tobacco, [11, 111](#)  
 fisher.latin, [11, 112](#)  
 foulley.calving, [10, 113](#)  
 fox.wheat, [11, 115](#)  
 garber.multi.uniformity, [9, 116](#)  
 gathmann.bt, [11, 117](#)  
 gauch.soy, [11, 119](#)  
 gge, [120](#)  
 gilmour.serpentine, [11, 123](#)  
 gilmour.slatehall, [11, 125](#)  
 gomez.fractionalfactorial, [11, 126](#)  
 gomez.groupsplit, [11, 128](#)  
 gomez.multilocsplitplot, [11, 129](#)  
 gomez.nitrogen, [11, 130](#)  
 gomez.rice.uniformity, [9, 132](#)  
 gomez.seedrate, [11, 133](#)  
 gomez.splitplot.subsample, [11, 134](#)  
 gomez.splitsplit, [11, 135](#)  
 gomez.striplot, [11, 137](#)  
 gomez.stripsplitplot, [11, 138](#)  
 gotway.hessianfly, [11, 139](#)  
 goulden.barley.uniformity, [9, 140](#)  
 goulden.latin, [11, 141](#)  
 graybill.heteroskedastic, [11, 142](#)  
 gumpertz.pepper, [11, 143](#)  
 hanks.sprinkler, [12, 145](#)  
 harris.multi.uniformity, [9, 147](#)  
 harris.wateruse, [10, 149](#)  
 harrison.priors, [14, 153](#)  
 hayman.tobacco, [12, 155](#)  
 hazell.vegetables, [12, 157](#)  
 heady.fertilizer, [12, 159](#)  
 henderson.milkfat, [10, 161](#)  
 hernandez.nitrogen, [12, 163](#)  
 hessling.argentina, [13, 165](#)  
 hildebrand.systems, [12, 167](#)  
 holland.arthropods, [10, 169](#)  
 holshouser.splitstrip, [12, 170](#)  
 hughes.grapes, [12, 172](#)  
 hunter.corn, [12, 173](#)

- ilri.sheep, [10](#), [174](#)  
 immer.sugarbeet.uniformity, [9](#), [177](#)  
 ivins.herbs, [12](#), [178](#)
- jansen.apple, [12](#), [180](#)  
 jansen.carrot, [12](#), [181](#)  
 jansen.strawberry, [12](#), [182](#)  
 jenkyn.mildew, [12](#), [183](#)  
 john.alpha, [12](#), [184](#)  
 johnson.blight, [12](#), [186](#)
- kalamkar.potato.uniformity, [9](#), [188](#)  
 kang.maize, [12](#), [189](#)  
 kang.peanut, [12](#), [190](#)  
 karcher.turfgrass, [12](#), [191](#)  
 keen.potatodamage, [12](#), [193](#)  
 kempton.barley.uniformity, [9](#), [194](#)  
 kempton.competition, [12](#), [196](#)  
 kempton.rowcol, [12](#), [197](#)  
 kempton.slatehall, [12](#), [198](#)
- lambert.soiltemp, [13](#), [200](#)  
 lasrosas.corn, [12](#), [201](#)  
 lavoranti.eucalyptus, [10](#), [203](#)  
 lee.potatobligh, [12](#), [204](#)  
 li.millet.uniformity, [9](#), [207](#)  
 lonnquist.maize, [12](#), [207](#)  
 lucas.switchback, [10](#), [209](#)  
 lyon.potato.uniformity, [9](#), [210](#)  
 lyons.wheat, [12](#), [211](#)
- mconway.turnip, [12](#), [212](#)  
 mcleod.barley, [12](#), [214](#)  
 mead.cauliflower, [12](#), [215](#)  
 mead.cowpeamaize, [12](#), [216](#)  
 mead.germination, [12](#), [218](#)  
 mead.strawberry, [12](#), [219](#)  
 mercer.mangold.uniformity, [9](#), [220](#)  
 mercer.wheat.uniformity, [9](#), [221](#)  
 minnesota.barley.weather, [12](#), [223](#)  
 minnesota.barley.yield, [12](#), [225](#)
- nass.barley, [13](#)  
 nass.barley (nass.corn), [227](#)  
 nass.corn, [13](#), [227](#)  
 nass.cotton, [13](#)  
 nass.cotton (nass.corn), [227](#)  
 nass.hay, [13](#)  
 nass.hay (nass.corn), [227](#)  
 nass.rice, [13](#)  
 nass.rice (nass.corn), [227](#)  
 nass.sorghum, [13](#)  
 nass.sorghum (nass.corn), [227](#)  
 nass.soybean, [13](#)  
 nass.soybean (nass.corn), [227](#)  
 nass.wheat, [13](#)  
 nass.wheat (nass.corn), [227](#)  
 nebraska.farmincome, [14](#), [229](#)
- odland.soybean.uniformity, [9](#), [230](#)  
 odland.soyhay.uniformity, [9](#)  
 odland.soyhay.uniformity  
     (odland.soybean.uniformity),  
     [230](#)
- ortiz.tomato, [12](#), [231](#)
- pacheco.soybean, [12](#), [233](#)  
 panel.outlinelevelplot, [234](#)  
 patterson.switchback, [10](#), [235](#)  
 pearce.apple, [10](#), [236](#)  
 pearl.kernels, [14](#), [237](#)  
 perry.springwheat, [12](#), [239](#)  
 piepho.cocksfoot, [12](#), [241](#)  
 plot.gge (gge), [120](#)
- ratkowsky.onions, [242](#)  
 RedGrayBlue, [243](#)  
 ridout.appleshoots, [12](#), [244](#)  
 rothamsted.brussels, [12](#), [246](#)  
 ryder.groundnut, [12](#), [248](#)
- salmon.bunt, [12](#), [249](#)  
 senshu.rice, [12](#), [250](#)  
 shafii.rapeseed, [12](#), [252](#)  
 sinclair.clover, [12](#), [253](#)  
 smith.corn.uniformity, [9](#), [255](#)  
 snedecor.asparagus, [12](#), [257](#)  
 snijders.fusarium, [12](#), [259](#)  
 steel.soybean, [13](#), [261](#)  
 stephens.sorghum.uniformity, [9](#), [263](#)  
 steptoe.morex.geno, [13](#)  
 steptoe.morex.geno  
     (step toe.morex.pheno), [264](#)  
 steptoe.morex.pheno, [13](#), [264](#)  
 stirret.borers, [14](#), [266](#)  
 streibig.competition, [13](#), [267](#)  
 stroup.nin, [13](#), [269](#)  
 stroup.splitplot, [13](#), [271](#)

student.barley, [13](#), [274](#)

talbot.potato, [13](#), [276](#)  
theobald.barley, [13](#), [277](#)  
theobald.covariate, [13](#), [279](#)  
thompson.cornsoy, [13](#), [280](#)  
turner.herbicide, [14](#), [282](#)

vargas.txe, [13](#), [284](#)  
vargas.wheat1, [13](#), [286](#)  
vargas.wheat2, [13](#), [288](#)  
verbyla.lupin, [13](#), [290](#)  
vold.longterm, [13](#), [292](#)  
vsn.lupin3, [13](#), [294](#)

wallace.iowaland, [14](#), [297](#)  
walsh.cottonprice, [13](#), [298](#)  
wassom.brome1.uniformity, [9](#), [299](#)  
wassom.brome2.uniformity, [9](#)  
wassom.brome2.uniformity  
    (wassom.brome1.uniformity), [299](#)  
wassom.brome3.uniformity, [9](#)  
wassom.brome3.uniformity  
    (wassom.brome1.uniformity), [299](#)  
waynick.soil, [14](#), [301](#)  
wedderburn.barley, [13](#), [302](#)  
weiss.incblock, [13](#), [303](#)  
weiss.lattice, [13](#), [305](#)  
welch.bermudagrass, [13](#), [306](#)  
Wheat2, [270](#)  
wiebe.wheat.uniformity, [9](#), [309](#)  
williams.barley.uniformity, [9](#), [311](#)  
williams.cotton.uniformity, [9](#), [312](#)  
williams.trees, [10](#), [313](#)

yan.winterwheat, [13](#), [314](#)  
yang.barley, [13](#), [315](#)  
yates.missing, [13](#), [316](#)  
yates.oats, [13](#), [318](#)

zuidhof.broiler, [10](#), [319](#)