

# Package ‘tigerstats’

May 23, 2015

**Type** Package

**Title** R Functions for Elementary Statistics

**Version** 0.2-6

**Date** 2015-05-23

**Author** Rebekah Robinson <rrobinson20@humana.com> and Homer White <hwhite0@georgetowncollege.edu>

**Maintainer** Homer White <hwhite0@georgetowncollege.edu>

**Description** A collection of data sets and functions that are useful in the teaching of statistics at an elementary level to students who may have little or no previous experience with the command line. The functions for elementary inferential procedures follow a uniform interface for user input. Some of the functions are instructional applets that can only be run on the R Studio integrated development environment with package 'manipulate' attached. R Studio is freely available ([www.rstudio.org](http://www.rstudio.org)), and includes package 'manipulate'. Other instructional applets are Shiny apps that may be run locally. In teaching the package is used alongside of package 'mosaic', 'mosaicData' and 'abd', which are therefore listed as dependencies.

**License** GPL (>= 3)

**Depends** R(>= 2.10), abd, mosaic, mosaicData

**Imports** ggplot2, lattice, MASS

**Suggests** knitr, shiny, shinyjs

**LazyLoad** yes

**LazyData** true

**VignetteBuilder** knitr

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2015-05-23 20:26:15

**R topics documented:**

tigerstats-package	4
alcohol	4
attitudes	5
barchartGC	6
beans	8
BinomNorm	8
BinomSkew	9
binomtestGC	9
cabrera	11
ChisqSimSlow	11
chisqtestGC	12
chugtime	14
CIMean	14
CIProp	15
colPerc	16
deathpen	16
DtrellHist	17
DtrellScat	17
EmpRule	18
EmpRuleGC	19
expCounts	20
FakeSchool	20
FindRegLine	21
fuel	21
galton	22
gfeeling	22
gcstudents	23
gss02	23
gss08	24
gss2012	25
hair_and_act	25
handheight	26
hanford1	26
hanford2	27
helpGC	27
henderson	28
hofbatting	29
hofpitching	30
imagpop	31
iqsiblings	32
knifeorgunblock	32
labels	33
ledgejump	33
lmGC	34
m111survey	35
m111surveyfa12	35

m11surveyfa13 . . . . .	36
MeanSampler . . . . .	37
music . . . . .	38
napkins . . . . .	39
nonresponse . . . . .	39
nosmokeday . . . . .	40
oldfaithful . . . . .	40
ostrichtemp . . . . .	41
Ovarian . . . . .	41
pbinomGC . . . . .	42
pchisqGC . . . . .	43
pennstate1 . . . . .	44
plot.GCIm . . . . .	44
plot.polyGC . . . . .	45
pnormGC . . . . .	46
Points2Watch . . . . .	47
polyfitGC . . . . .	47
popsamp . . . . .	48
predict.GCIm . . . . .	49
predict.polyGC . . . . .	50
PropSampler . . . . .	50
proptestGC . . . . .	51
ptGC . . . . .	53
pushups . . . . .	54
qnormGC . . . . .	54
RandomExp . . . . .	55
RegEstimate . . . . .	56
rowPerc . . . . .	57
saltmarsh . . . . .	57
SampDist2Means . . . . .	58
SampDist2Props . . . . .	59
SampDistMean . . . . .	60
sat . . . . .	61
sealsO2 . . . . .	61
ShallowReg . . . . .	62
SimpleRandom . . . . .	63
SimpleRandom2 . . . . .	63
Skewer . . . . .	64
SlowGoodness . . . . .	65
SmallExp . . . . .	65
stumps . . . . .	66
temperature . . . . .	66
tExplore . . . . .	67
theme.rpres . . . . .	67
themerpres . . . . .	68
tips . . . . .	68
tornado . . . . .	69
tSampler . . . . .	69

ttestGC . . . . .	70
Type12Errors . . . . .	72
ucdavis1 . . . . .	72
Variability . . . . .	73
VaryCorrelation . . . . .	74
verlander . . . . .	74
youthrisk03 . . . . .	75

## Index 77

---

tigerstats-package      *R Functions for Elementary Statistics*

---

### Description

tigerstats

### Details

Datasets and functions useful for teaching elementary statistics.

### Author(s)

Rebekah Robinson (<rrobinson20@humana.com>), Homer White (<hwhite0@georgetowncollege.edu>)

### References

<http://statistics.georgetowncollege.edu>

---

alcohol      *Alcohol at Georgetown College*

---

### Description

Alcohol policy violations on the Georgetown College campus over several years.

### Format

A data frame with 10 observations on the following 4 variables.

**year** Academic year ending with Spring of the given year.

**enrollment** Full-time equivalent enrollment.

**writeups** Number of write-ups for alcohol violations.

**writeups.per.100** Number of writeups per 100 students.

### Source

Collected by MAT 111 students as a project.

attitudes

*Attitudes Experiment 2001***Description**

Study conducted in November 2001 by students in MAT 111. Subjects were 267 Georgetown College students. Not all subjects got the same survey form.

**Format**

A data frame with 268 observations on the following 8 variables.

**def.race** Suggested race of the defendant in the survey form.

**vic.race** Suggested race of the victim in the survey form.

**conc.situation** Scenario described in the in the "rock concert" question on the survey form.

**sentence** Sentence, in years, recommended for the defendant.

**conc.decision** Whether or not the subject chose to buy a ticket (or buy another ticket).

**year** Class rank of the subject.

**sex** a factor with level Sex of the survey participant.

**major** possible values: humanities math.sci pre.prof social.sci Type of major the subject intends.

**Details**

Here is a sample survey form, with variants noted.

Attitudes Survey

Crime: You are on a jury for a manslaughter case in Lewistown, PA. The defendant has been found guilty, and in Pennsylvania it is part of the job of the jury to recommend a sentence to the judge. The facts of the case are as follows. The defendant, Tyrone Marcus Watson, a 35-year old native of Lewistown, was driving under the influence of alcohol on the evening of Tuesday July 17, 2001. At approximately 11:00 PM Watson drove through a red light, striking a pedestrian, Betsy Brockenheimer, a 20-year old resident of Lewistown. Brockenheimer was taken unconscious to the hospital and died of her injuries about one hour later. Watson did not flee the scene, nor did he resist arrest.

The prior police record for Mr. Watson is as follows: two minor traffic violations, and one previous arrest, five years ago, for DUI. No one was hurt in that incident.

Watson has now been convicted of DUI and manslaughter. The minimum jail term for this combination of offenses is two years; the maximum term is fifty years. In the blank below, write a number from 2 to 50 as your recommended length of sentence for Tyrone Marcus Watson.

\_\_\_\_\_

[In the question above, name of defedant could vary: either William Shane Winchester or Tyrone Marcus Watson. The name of the victim could also vary: either Betsy Brockenheimer or Latisha Dawes.]

**Spending Habits**

You have purchased a \$30 ticket to see a rock concert in Rupp Arena. When you arrive at the Arena on the night of the performance, you find that you have lost the ticket. You have no receipt, so it will not be possible to see the concert unless you purchase another ticket. Would you purchase another ticket? Circle below.

YES NO

[In other forms, the question above could have been: You plan to see a rock concert in Rupp Arena. Tickets for the performance are \$30. When you arrive at the Arena on the night of the performance, you find that you have lost two bills from your purse or wallet: a ten and a twenty. Would you buy the ticket anyway?]

**Respondent Data**

I am (circle one): freshman sophomore junior senior

I am (circle one) male female

(Optional) My intended major is: \_\_\_\_\_

**Source**

Georgetown College

---

barchartGC

*Easy Barcharts*

---

**Description**

Wrapper for `barchart` in package `lattice`. Creates a barchart from raw data using formula-data syntax similar to that of `xtabs`, or from a table. Defaults to a "standard" barchart in which the bars are vertical and unstacked. Supports percentage barcharts.

**Usage**

```
barchartGC(x, data=parent.frame(), type="frequency", flat=FALSE, auto.key=TRUE,
           horizontal=FALSE, stack=FALSE, ...)
```

**Arguments**

<code>x</code>	Either a formula or an object that can be coerced to a table. If formula, it must be of the form <code>~var</code> or <code>~var1+var2</code> .
<code>data</code>	Usually a data frame that supplies the variables in <code>x</code> . Variables not in the data argument are searched for in the parent environment.
<code>type</code>	Possible values are "frequency" and "percent".
<code>flat</code>	If set to TRUE, will produce barchart that resembles the layout of <code>xtabs</code>
<code>auto.key</code>	Provides a simple key
<code>horizontal</code>	Determines orientation of the bars (overridden by <code>flat</code> )

`stack` Determines whether bars for tallies are stacked on each other or placed next to one another (overridden by `flat`)

`...` other arguments passed to `barchart`: these include `main`, `sub`, and `xlab`, which are likely to be familiar to students from other lattice graphical functions. An error is possible if other arguments pertaining to legends are passed (hopefully anyone interested in such will have moved on to `barchart`).

### Value

A trellis object describing the barchart.

### Author(s)

Homer White <hwhite0@georgetowncollege.edu>

### Examples

```
#barchart of counts for one factor variable:
barchartGC(~sex,data=m111survey)

#barchart with percentages and title:
barchartGC(~sex,data=m111survey,
  type="percent",
  main="Distribution of Sex")

#barchart of counts, to study the relationship between
#two factor variables:
barchartGC(~sex+seat,data=m111survey)

#percentage barchart, two factor variables:
barchartGC(~sex+seat,data=m111survey,type="percent")

#From tabulated data:
sexseat <- xtabs(~sex+seat,data=m111survey)
barchartGC(sexseat,type="percent",main="Sex and Seating Preference")

#from tabulated data:
dieTosses <- c(one=8,two=18,three=11,four=7,five=9,six=7)
barchartGC(dieTosses,main="60 Rolls of a Die")

# a "flat" barchart, pictorial version of xtabs()
barchartGC(~sex+seat,data=m111survey,flat=TRUE,ylab="Sex")

# a "flat" barchart, pictorial version of xtabs()
barchartGC(~sex+seat,data=m111survey,type="percent",flat=TRUE,ylab="Sex")
```

---

 beans

*Beans*


---

**Description**

Experiment performed at UC-Davis; fifteen students participated. Each student was asked to place as many beans into a cup as he/she could, in 15 seconds. Each student performed this task once with the dominant hand, and once with the nondominant hand, but the order of performance was randomized. The purpose of the study was to see whether manual dexterity was better for the dominant hand. Terminology: your dominant hand is the hand you use the most.

**Format**

A data frame with 15 observations on the following 3 variables.

**Dom** Number of beans placed into cup with the dominant hand.

**NonDom** Number of beans placed with the nondominant hand.

**Diff** Difference in number of beans placed (dominant hand minus nondominant hand).

**Source**

Uts and Heckard, Mind on Statistics, 4th Edition.

---

 BinomNorm

*Binomial Distributions With Normal Approximation*


---

**Description**

An app to investigate the binomial family.

**Usage**

```
BinomNorm()
```

**Value**

no value. Graphical side-effects only.

**Author(s)**

Homer White (hwhite0@georgetowncollege.edu)

**Examples**

```
## Not run:
if (require(manipulate)) BinomNorm()

## End(Not run)
```



---

BinomSkew

*Skewness in the Binomial Family of Distributions*

---

### Description

An app to investigate how skewness in a binomial distribution vanishes when  $np$  is large enough. Sample size is set at  $n = 50$ , but the user can vary  $p$  with a slider.

### Usage

```
BinomSkew()
```

### Value

no value. Graphical side-effects only.

### Author(s)

Homer White (hwhite0@georgetowncollege.edu)

### Examples

```
## Not run:  
if (require(manipulate)) BinomSkew()  
  
## End(Not run)
```

---

binomtestGC

*Exact Procedures for a Single Proportion*

---

### Description

Wrapper for `binom.test` in package `stats`. Employs the binomial distribution in inferential procedures for a single proportion.

### Usage

```
binomtestGC(x, data=parent.frame(), n=numeric(), p=NULL,  
            alternative=c("two.sided", "less", "greater"),  
            success="yes", conf.level=0.95, graph=FALSE, verbose=TRUE)
```

**Arguments**

x	Either a formula or a numeric vector. If formula, it must be of the form $\sim x$ indicating the single variable under study. When summary data are provided, x is a numeric vector of success counts.
data	Data frame that supplies the variable x. If not found in data, the variable is searched for in the parent environment.
n	When not empty, this is a numeric vector giving the size of the sample.
p	Specifies Null Hypothesis value for population proportion. If not set, no test is performed.
alternative	"two.sided" requests computation of a two-sided P-value; other possible values are "less" and "greater".
success	When x is a formula, this argument indicates which value of variable x is being counted as a success. When working with formula-data input the value of this parameter <b>MUST</b> be set, even when the variable has only two values.
conf.level	Number between 0 and 1 indicating the confidence-level of the interval supplied.
graph	If TRUE, plot graph of P-value. Ignored if no test is performed.
verbose	Determines whether to return lots of information or only the basics

**Value**

an object of class GCbinomtest.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
#Confidence interval only:
binomtestGC(~sex,data=m111survey,success="female")

#Confidence interval and two-sided test, Null Hypothesis p = 0.5:
binomtestGC(~sex,data=m111survey,success="female",p=0.5)

#For confidence level other than 95%, use conf.level argument.
#For 90% interval:
binomtestGC(~sex,data=m111survey,success="female",conf.level=0.90)

#For one-sided test, set alternative argument as desired:
binomtestGC(~sex,data=m111survey,p=0.50,
  success="female",alternative="greater")

#Summary data:
#In one sample, 40 successes in 100 trials. Testing whether p = 0.45.
binomtestGC(x=40,n=100,p=0.45)
```

cabrera

*Miguel Cabrera***Description**

PITCHf/x data on nine-time All-Star Miguel Cabrera, who is thought to be one of the best pure hitters in baseball. Covers seasons 2009 through 2012.

**Format**

A data frame with 6265 observations on the following 12 variables. Each observation is a pitch to Cabrera.

**season** The year of play

**gamedate** Date of the game in which the pitch was thrown

**pitch\_type** Type of pitch thrown, as determined by a computer algorithm.

**balls** Current ball count

**strikes** Current strike count

**speed** speed of pitch (in mph). (When crossing plate?)

**px** x-coordinate of pitch (in feet, measured from center of plate)

**pz** vertical coordinate of pitch (in feet above plate)

**swung** Whether or not cabrera swung at the ball. Factor with levels "no", "yes".

**hitx** x-coordinate of landing point of ball (if it was hit). Relative to park.

**hity** y-coordinate of landing point of ball (if it was hit). Relative to park.

**hit\_outcome** Outcome when ball was hit. Factor with levels E (error), H (hit), O (batter out).

**Source**

Marchi and Albert: Analyzing Baseball Data with R, CRC Press 2014. For more on the PITCHf/x system, see <http://en.wikipedia.org/wiki/PITCHf/x>.

ChisqSimSlow

*Chi Square Resampler (One at a Time)***Description**

An app to illustrate use of the chi-square statistic to test for a relationship between two categorical variables. The P-value is computed by resampling, and the resamples are done one at a time. A histogram of resampled chi-square statistics is displayed after each resample, and summary information is output to the console.

**Usage**

```
ChisqSimSlow(form,data,effects=c("random", "fixed"))
```

**Arguments**

form	a formula of the form $\sim x+y$ . When using fixed effects (see below for explanation), x should be the variable that is considered the predictor variable.
data	A data frame from which x and y are drawn.
effects	When effects="fixed", the resampling is performed under the condition that the row sums in the resampled two-way table (with x for rows) are the same as the row sums in the twoway table based on the original data. When effects="random", then both row and column sums in the resampled table may vary: only the sum of the counts is constant. (Note: in the resampling procedure for <code>chisq.test</code> in the stats package of R, both row and column sums are required to equal the corresponding sums for the original data.)

**Value**

Graphical and numerical output

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:
ChisqSimSlow(~weather+crowd.behavior,data=ledgejump,effects="fixed")

## End(Not run)
```

---

chisqtestGC

*Chi-Square Test (GC version)*

---

**Description**

Perform chi-square test, either goodness of fit or test for association. Enter either formula-data input or a summary table. Simulation is optional.

**Usage**

```
chisqtestGC(x, data = parent.frame(), p = NULL, graph = FALSE,
            simulate.p.value = FALSE, B = 2000, verbose = TRUE)
```

**Arguments**

x	Could be a formula. If so, either ~var (for goodness of fit) or ~var1+var2 (for test for association). Otherwise either a table, matrix or vector of summary data.
data	dataframe supplying variables for formula x. If variables in x are not found in the data, then they will be searched for in the parent environment.
p	For goodness of fit, a vector of probabilities. This will be automatically scaled so as to sum to 1. Negative elements result in an error message.
graph	produce relevant graph for P-value (chi-square curve or histogram of simulation results).
simulate.p.value	If FALSE, use a chi-square distribution to estimate the P-value. Other possible values are "random" and "fixed" and TRUE. Random effects are suitable for resampling when the data are a random sample from a population. Fixed effects assume that the values of the explanatory variable (row variable for table, var1 in formula ~var1+var2) remain fixed in resampling, and values of response variable are random with null distribution estimated from the data. When set to TRUE, we implement an equivalent to R's routine. In our view procedure is most suitable when the data come from a randomized experiment in which the treatment groups are the values of the explanatory variable.
B	number of resamples to take.
verbose	If TRUE, include lots of information in the output.

**Value**

an object of class GCchisqtest

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
#Goodness of fit test for one factor variable:
chisqtestGC(~seat,data=m111survey,p=c(1/3,1/3,1/3))

#Test for relationship between two factor variables:
chisqtestGC(~sex+seat,data=m111survey)

#You can input a two-way table directly into chisqtestGC():
SexSeat <- xtabs(~sex+seat,data=m111survey)
chisqtestGC(SexSeat)

#Several types of simulation are possible, e.g.:
chisqtestGC(~weather+crowd.behavior,data=ledgejump,simulate.p.value="fixed",B=2500)

#For less output, set argument verbose to FALSE:
chisqtestGC(~sex+seat,data=m111survey,verbose=FALSE)
```

---

chugtime	<i>Time to Chug</i>
----------	---------------------

---

**Description**

College-aged males chugging a 12-ounce can of a certain beverage.

**Format**

A data frame with 13 observations on the following 2 variables.

**Weight** Weight of the subject (in pounds).

**ChugTime** How long (in seconds) the subject requires to drink the beverage.

**Source**

Utts and Heckard, Mind on Statistics, 4th Edition.

---

CIMean	<i>Confidence Intervals (for one population mean)</i>
--------	---

---

**Description**

An app to investigate how sample size and confidence level affect the width of a confidence interval. A sample is drawn from the input population and a confidence interval for the population mean is calculated. The kernel density plot for the population and the histogram for each new sample are plotted, along with the confidence interval. Summary information is output to the console to tally the number of times the computed confidence interval covers the true population mean and how many times it misses. There is an option to draw 100 or 1000 samples at a time.

**Usage**

```
CIMean(form, data)
```

**Arguments**

form	a formula of the form $\sim$ var.
data	A data frame from which var is drawn.

**Value**

Graphical and numerical output

**Author(s)**

Rebekah Robinson <rebekah\_robinson@georgetowncollege.edu>

**Examples**

```
## Not run:  
if (require(manipulate)) CIMean(~height,data=imagpop)  
  
## End(Not run)
```

---

CIProp

*Confidence Intervals (for one population proportion)*

---

**Description**

An app to investigate how many times a confidence interval for one population proportion captures the true population parameter. The true population proportion is plotted as a vertical red line and the user can visually see how changes to the sample, population proportion, sample size, and confidence level affect the width of the confidence interval. Summary information is output to the console to tally the number of times the computed confidence interval covers the true population mean and how many times it misses.

**Usage**

```
CIProp()
```

**Value**

Graphical and numerical output

**Note**

Uses manipulate from RStudio

**Author(s)**

Rebekah Robinson <rebekah\_robinson@georgetowncollege.edu>

**Examples**

```
## Not run:  
if (require(manipulate)) CIProp()  
  
## End(Not run)
```

---

colPerc	<i>Column Percents</i>
---------	------------------------

---

**Description**

Computes column percentages for a given twoway table.

**Usage**

```
colPerc(tab)
```

**Arguments**

tab                    A two way table, e.g., the result of `xtabs(~var1+var2, data=DataFrame)`.

**Value**

An object of class `table`, giving column percentages for the input table.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
MyTable <- xtabs(~weather+crowd.behavior, data=ledgejump)
colPerc(MyTable)
```

---

deathpen	<i>The Death Penalty and Race</i>
----------	-----------------------------------

---

**Description**

A dataset recreated from summary data that describes relationships between race of defendant, race of victim, and outcome of trial in a number of capital cases in Florida in 1976-1977.

**Format**

A data frame with 326 rows and 3 variables

**defrace** Race of the defendant in the capital case

**vicrace** Race of the victim

**death** Whether or not the defendant in the case received the death penalty

**Source**

Michael J. Radelet: "Racial Characteristics and the Imposition of the Death Penalty", *American Sociological Review*, 46 (1981). See also the JSTOR location: <http://www.jstor.org/stable/2095088>.



---

DtrellHist                      *Dynamic Trellising (Histogram)*

---

**Description**

A manipulative app that facilitates exploration of the distribution of a single numerical variable, conditioned upon the values of either a numerical variable or a factor.

**Usage**

```
DtrellHist(form,data)
```

**Arguments**

form	a formula of the form <code>~var cond</code> . <code>var</code> must be numeric; <code>cond</code> may be either numeric or factor.
data	A data frame fromm <code>var</code> and <code>cond</code> are drawn.

**Value**

Graphical output.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:  
if (require(manipulate)) DtrellHist(~dist|speed,data=cars)  
  
## End(Not run)
```

---

DtrellScat                      *Dynamic Trellising (Scatterplot)*

---

**Description**

An app to facilitate exploration of the relationship between two numerical variables, conditional upon the values of a third variable.

**Usage**

```
DtrellScat(form,data)
```

**Arguments**

form            A formula of the form  $y \sim x | c$ . All three variables in the formula should be from the data frame data. c May be a factor or numerical.

data            A data frame.

**Value**

Graphical and numerical output.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:
if (require(manipulate)) DtrellScat(sat~salary|frac,data=sat)

## End(Not run)
```

---

EmpRule

*Empirical Rule*

---

**Description**

An app to investigate how the Empirical Rule applies to symmetric data and skewed data. The user can select if they want to view a histogram of symmetric data or skewed data. Vertical bars are also plotted to signify one, two, and three standard deviations from the mean. Summary data is output to the console giving the proportion of the histogram that falls within one, two, and three standard deviations of the mean.

**Usage**

```
EmpRule()
```

**Value**

Graphical and numerical output

**Author(s)**

Rebekah Robinson <rebekah\_robinson@georgetowncollege.edu>

**Examples**

```
## Not run:
if (require(manipulate)) EmpRule()

## End(Not run)
```

---

EmpRuleGC

*Graphical Calculator for the Empirical Rule*

---

### Description

An app to facilitate visual understanding of Empirical Rule approximations of probabilities, percentages.

### Usage

```
EmpRuleGC(mean=0, sd=1, xlab="x")
```

### Arguments

mean	Mean of the distribution
sd	Standard deviation of the distribution
xlab	x-axis label

### Value

Returns no value. Used for the plotting side-effects.

### Note

Uses `manipulate` in RStudio

### Author(s)

Homer White <hwhite0@georgetowncollege.edu>

### Examples

```
## Not run:  
if(require(manipulate)) EmpRuleGC(mean=70, sd=3, xlab="Height (inches)")  
  
## End(Not run)
```

expCounts

*Expected Cell Counts From Two-Way Tables*

---

**Description**

Simple instructional function to compute expected cell counts from a table of observed counts.

**Usage**

```
expCounts(tab)
```

**Arguments**

**tab** A table with two dimensions, or an object that can be coerced to one.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

---

FakeSchool

*An Imaginary School*

---

**Description**

Hypothetical school, used for illustrative purposes

**Format**

A data frame with 28 observations on the following 5 variables.

**Students** Name of each student

**Sex** sex of the student

**class** class rank of the student

**GPA** grade point average

**Honors** whether or not the student is in the Honors Program

---

`FindRegLine`*Find the Regression Line*

---

**Description**

The regression minimizes the residual sum of squares (RSS). In this game, the player chooses slope and y-intercept of a line so as to approximate the regression line. The moveable line is set initially as a horizontal line with height equal to the mean of the y-coordinates of the scatterplot, so initially the residual sum of squares equals the total sum of squares (TSS). The player's score is the sum of the number of turns taken and the difference between the current RSS and the regression line's RSS (as a percentage of TSS-RSS for regression line). The aim is to lower one's score.

**Usage**`FindRegLine`**Value**

Graphical and numerical output.

**Note**

Requires package `manipulate`, available only in RStudio.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:  
if (require(manipulate)) FindRegLine()  
  
## End(Not run)
```

---

`fuel`*Speed and Fuel Efficiency (British Ford Escort)*

---

**Description**

A British Ford Escort was driven along a prescribed course. Each drive was done at a different speed, and the fuel efficiency was recorded for each drive.

**Format**

A data frame with 15 observations on the following 2 variables.

**speed** in kilometers per hour.

**efficiency** fuel efficiency, measured in liters of fuel required to travel 100 kilometers.

**Source**

The Basic Practice of Statistics, by Moore and McCabe.

---

galton

*Galton's Father-Son Data*

---

**Description**

Data on father-son pairs. Collected in 1885 by Francis Galton.

**Format**

A data frame with 1078 observations on the following 2 variables.

**fheight** Height of the father, in inches.

**sheight** Height of the son, in inches.

---

gcfeeling

*Feelings About Georgetown College*

---

**Description**

Results of a survey conducted by Georgetown College students on 47 Georgetown College upper-class students.

**Format**

A data frame with 47 observations on the following 6 variables.

**rating.fresh** how happy the subjects remembers being as a first-year student, on a scale of 1 to 10.

**rating.js** how happy the subjects feels now, on a scale of 1 to 10.

**greek** whether or not the subject belongs to a greek organization.

**athlete** whether or not the subject is a varsity athlete

**rating.diff** upper-level happiness rating minus remembered first-year rating

**happier** whether or not subject feels happier now than as a first-year student

**Source**

MAT 111 at Georgetown College

---

gcstudents

*Georgetown College Students*


---

**Description**

Data collected by GC students.

**Format**

A data frame with 62 observations on the following 4 variables.

**height** height of the survey participant, in inches

**GPA** grade-point average

**enough\_Sleep** Does the participant feel that he/she gets enough sleep?

**sex** sex of the survey participant

**Source**

MAT 111 at Georgetown College

---

gss02

*General Social Survey, 2002*


---

**Description**

The General Social Survey (GSS) is a nationwide poll that has been conducted since 1972 (semi-annually since 1994). Most interviews are done face-to-face. For further information, see below.

**Format**

A data frame with 2765 observations on the following 13 variables.

**sex** a factor with levels Female Male

**race** a factor with levels AfrAm Hispanic Other White

**degree** a factor with levels Bachelor Graduate HighSchool JunColl NotHS

**relig** a factor with levels Catholic Jewish Other Protestant

**polparty** a factor with levels Democrat Independent Other Republican

**cappun** a factor with levels Favor Oppose Whether or not the subject favors capital punishment.

**tvhours** the subject estimates number of hours per day he or she watches TV.

**marijuan** a factor with levels Legal NotLegal Whether or not subject believes that marijuana should be legalized.

**owngun** a factor with levels No Yes. Does the subject own a gun?

**gunlaw** a factor with levels Favor Oppose Whether or not the subject favors stricter gunlaws.

**age** age of the subject

**chldidel** the ideal number of children the subject would like to have.

**emailtime** estimated number of hours per week subject spends using email.

**Source**

National Opinion Research Center: <http://www3.norc.org/gss+website/>. Found in Uts and Heckard: Mind on Statistics, 4th Edition.

---

gss08

*General Social Survey, 2008*

---

**Description**

General Social Survey, 2008

**Format**

A data frame with 2023 observations on the following 12 variables.

**sex** a factor with levels Female Male

**race** a factor with levels AfrAm Other White

**degree** a factor with levels Bachelor Graduate HighSchool JunColl NotHS

**relig** a factor with levels Catholic Jewish None Other Protestant

**polparty** a factor with levels Democrat Independent Other Republican

**cappun** a factor with levels Favor Oppose

**tvhours** a numeric vector

**marijuan** a factor with levels Legal NotLegal

**owngun** a factor with levels No Yes

**gunlaw** a factor with levels Favor Oppose

**age** a numeric vector

**chldidel** a numeric vector

**Source**

National Opinion Research Center: <http://www3.norc.org/gss+website/>. Found in Uts and Heckard: Mind on Statistics, 4th Edition.

**References**

For more information see gss02



---

gss2012

*General Social Survey, 2012*

---

### Description

A selection of variables from the 2012 General Social Survey. The variables are as follows:

- age. Age of the subject.
- sex. Sex of the subject.
- race. Race of the subject.
- polviews. Subject's political views.
- relig. Religion of the subject.
- cappun. Opinion on capital punishment.
- owngun. Whether or not one owns a gun.
- emailhr. Number of hours per week spent on email.
- bigbang. Whether or not subject believes the Big Bang theory is true.
- premarsx. Opinion on premarital sex.
- pornlaw. Should pornography be legal?
- zodiac. Sign of the Zodiac under which the subject was born.

### Format

A data frame with 1976 rows and 12 variables

### Source

<http://www3.norc.org/gss+website/>.

---

hair\_and\_act

*Hair Color and ACT Score*

---

### Description

A study performed by MAT 111 students at Georgetown College.

### Format

A data frame with 100 observations on the following 3 variables.

**sex** a factor with levels female male

**hair.color** a factor with levels dark light

**act** composite ACT score of subject.

### Source

MAT 111 at Georgetown College

---

handheight

*Height and Handspan*

---

**Description**

Height and handspan of a few subjects.

**Format**

A data frame with 167 observations on the following 3 variables.

**Sex** a factor with levels Female Male

**Height** height of subject, in inches.

**HandSpan** handspan of subject, in centimeters.

**Source**

Uts and Heckard, Mind on Statistics, 4th Edition.

---

hanford1

*Handford Weather Station, 1984-2010*

---

**Description**

The station is located in Hanford, WA.

**Format**

A data frame with 27 observations on the following 2 variables.

**year** calendar year

**temp** average high temperature for that year.

**Source**

See <http://www.hanford.gov/hms/>

---

hanford2

*Hanford Weather Station, 1945-2010*

---

**Description**

The weather station is located in Hanford, WA. Note that this dataset is more complete than [hanford1](#)

**Format**

A data frame with 66 observations on the following 2 variables.

**year** calendar year

**temp** average high temperature for that year.

**Source**

See [www.hanford.gov/hms/](http://www.hanford.gov/hms/)

---

helpGC

*Quick Vignettes in the Viewer*

---

**Description**

A convenience function to show package vignettes. Vignette will show in a viewer selected by the front-end (e.g, the Viewer pane in R Studio), or in a browser if the "viewer" option is not set.

**Usage**

```
helpGC(topic, package="tigerstats")
```

**Arguments**

topic filename of the vignette, exclusive of the .html extension

package Name of the installed package containing the desired vignette

**Value**

side effects

**Author(s)**

Homer White (hwhite0@georgetowncollege.edu)

**Examples**

```
## Not run:  
helpGC(1mGC)  
  
## End(Not run)
```

---

henderson

*Ricky Henderson's Career Stats*

---

**Description**

Various year-by-year statistics for the Major League player Ricky Henderson

**Format**

A data frame with 23 observations on the following 18 variables.

**Season** season  
**TM** team played for  
**G** number of games played in  
**AB** number of at-bats  
**R** runs scored  
**H** number of base hits  
**X2B** doubles  
**X3B** triples  
**HR** home runs  
**RBI** runs batted in  
**BB** bases on balls  
**SO** number of times struck out  
**SB** Number of stolen bases  
**CS** number of times caught stealing  
**AVG** batting average  
**OBP** on-base percentage  
**SLG** slugging rate  
**OPS** on-base plus slugging

**Source**

unknown (possibly Albert: Teaching Statistics With Baseball)

---

hofbatting	<i>Hall of Fame Batters</i>
------------	-----------------------------

---

**Description**

Data on the 147 batters inducted into the Major League Baseball Hall of Fame as of the year 2013.

**Format**

A data frame with 147 observations on the following 29 variables.

**Name** Player name

**Rk** Unknown

**Inducted** Year inducted into Hall of Fame

**Yrs** Number of years played in the Majors

**From** First year in the Majors

**To** Last year in the Majors

**MidCareer** Middle year of player's career

**Era** Era of Baseball history in which the player was (for the most part) active. Values are: 19th century (before 1900), Dead Ball (1900-1919), Lively Ball (1920-1940), Integration, (1941-1959), Expansion (1960-1975), Free Agency (1976-1992), Long Ball (1993 +).

**ASG** Number of All-Star games played

**WAR.pos** Wins above replacement (WAR), as defined for a position player

**G** Games played

**PA** Number of plate appearances

**AB** Number of times at bat

**R** Runs scored

**H** Base hits

**X2B** Doubles

**X3B** Triples

**Triple.Rate** Number of triples divided by number of times at bat

**HR** Home runs

**HR.Rate** Number of home runs divided by number of times at bat

**RBI** Runs batted in

**SB** Number of successful stolen base attempts

**CS** Number of times thrown out while attempting to steal a base

**BB** Base on Balls (number of times "walked")

**SO** Number of times struck out

**BA** Batting average

**OBP** On base percentage

**SLG** Slugging average

**OPS** OBP plus SLG

**Source**

Modified from Marchi and Albert: Analyzing Baseball Data with R, CRC Press 2014.

---

hofpitching

*Hall of Fame Pitchers*

---

**Description**

Data on the 70 pitchers inducted into the Major League Baseball Hall of Fame as of the year 2013.

**Format**

A data frame with 147 observations on the following 32 variables.

**Name** Player name

**Rk** Unknown

**Inducted** Year inducted into Hall of Fame

**Yrs** Number of years played in the Majors

**From** First year in the Majors

**To** Last year in the Majors

**MidCareer** Middle year of player's career

**Era** Era of Baseball history in which the player was (for the most part) active. Values are: 19th century (before 1900), Dead Ball (1900-1919), Lively Ball (1920-1940), Integration, (1941-1959), Expansion (1960-1975), Free Agency (1976-1992), Long Ball (1993 +).

**ASG** Number of All-Star games played

**WAR** Wins above replacement

**W** Games won

**L** Games lost

**W.L.** proportion of games won

**ERA** Earned run average

**G** Games played

**GS** Games started

**GF** Games finished

**CG** Complete games

**SHO** Shut-outs

**SV** Saves

**IP** Innings pitched

**H** Hits allowed

**R** Runs allowed

**ER** Earned Runs allowed

**HR** Home Runs allowed  
**BB** Bases on Balls (number of "walks")  
**IBB** Intentional walks  
**SO** Strikeouts  
**HBP** Hit batter with pitch (?)  
**BK** Balks  
**WP** Wild Pitches  
**BF** Total batters faced

### Source

Modified from Marchi and Albert: Analyzing Baseball Data with R, CRC Press 2014.

---

imagpop

*An Imaginary Population*

---

### Description

An imaginary population, used for instructional purposes. The variables are as follows:

- sex. (male, female).
- math. Whether or not you were a mathematics major.
- income. Your annual income, rounded to the nearest \$100.
- cappun. Opinion about the death penalty (favor, oppose).
- height. Height in inches.
- idealheight. The height you would like to be, in inches.
- diff. Idealheight - actual height.
- kkardashtemp. Your feelings about Kim Kardashian on a 0-100 scale (0=very cold, 100=very warm).

### Format

A data frame with 10000 rows and 8 variables

---

 iqsiblings

*IQs of Siblings*


---

**Description**

IQs of pairs of siblings.

**Format**

A data frame with 80 observations on the following 2 variables.

**First** IQ of the older sibling.

**Second** IQ of the younger sibling.

**Source**

William Harris, Georgetown College

---

knifeorgunblock

*Knife or Gun?*


---

**Description**

What will make you yell louder: being killed with a knife or being killed with a gun? Results of an entirely imaginary experiment performed on very strange volunteers. Members of the Knife group are killed by a knife, and members of the Gun group are killed by a gun. The volume of the screams of each subject during slaying is recorded. In order to ensure that the two groups are similar with respect to how loud they can yell in the first place, subjects are blocked by whether or not they have participated in hog-hollering contests. After blocking, subjects are randomly assigned to groups.

**Format**

A data frame with 20 observations on the following 3 variables.

**hoghollerer** a factor with levels no yes whether or not the subject competes in hog-hollerin' contests

**means** a factor with levels gun knife means by which subject is slain

**volume** volume of expiring subject's cries.

**Source**

A morbid imagination.



---

labels	<i>Labels and Perception of Quality</i>
--------	---

---

**Description**

Students in MAT 111 performed an experiment to see whether the perception of the quality of peanut butter was affected by the labeling on the peanut butter jar. Each subject tasted from two jars, one of which was labeled Jiff, and the other of which was labeled Great Value (a cheaper brand). Unknown to the subjects, both jars contained Great Value peanut butter. Each subject rated the quality of the peanut butter on a scale of 1 to 10.

**Format**

A data frame with 30 observations on the following 3 variables.

**jiffrating** rating subject gave to the PB in the jar with the Jiff label

**greatvaluerating** rating subject gave to the PB in the jar with the Great Value label

**sex** a factor with levels female male

**Source**

MAT 111 at Georgetown College

---

ledgejump	<i>Crowd Behavior at Ledge-Jumping Incidents</i>
-----------	--

---

**Description**

A dataset recreated from summary data describing the relationship between weather and crowd behavior during 21 recorded incidents in England, in which a (suicidal) person was contemplating jumping from a ledge or other high structure and a crowd gathered to watch. The variables are as follows:

- weather. Warm or cool, based on the time of year when the incident occurred.
- crowd.behavior. The crowd either baited the would-be jumper, or was polite.

**Format**

A data frame with 21 rows and 2 variables

**Source**

"The baiting crowd in episodes of threatened suicide", *Journal of Personality and Social Psychology*, 41, 703-709. See also dataset 59 in *A Handbook of Small Datasets* by Hand et. al. See also <http://www.ncbi.nlm.nih.gov/pubmed/7288565>.

---

`lmGC`*Linear Regression*

---

**Description**

Regression analysis (one numerical predictor variable) with simplified output. Wrapper function for `lm` in package `stats`.

**Usage**

```
lmGC(form,data=parent.frame(),graph=FALSE,check=FALSE)
```

**Arguments**

<code>form</code>	formula of form $y \sim x$ , both variables numeric
<code>data</code>	dataframe supplying $y$ and $x$ above. If one or more of the variables is not in <code>data</code> , then they will be searched for in the parent environment.
<code>graph</code>	Produce scatterplot with fitted polynomial, together with prediction standard error bands
<code>check</code>	Asks to produce a lowess or gam curve with approximate 95 fitted line wanders outside the band, then perhaps a linear fit is not appropriate.

**Value**

A list of class "GCIm". Elements that may be queried include "slope", "intercept", "s" (residual standard error), "R^2" (unadjusted).

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
#To study the relationship between two numerical variables:  
lmGC(fastest~GPA,data=m111survey,graph=TRUE)
```

---

m111survey

*MAT 111 Survey*

---

**Description**

Results of a survey of MAT 111 students at Georgetown College.

- height. How tall are you, in inches?
- ideal\_ht. A numeric vector How tall would you LIKE to be, in inches?
- sleep. How much sleep did you get last night?
- fastest. What is the highest speed at which you have ever driven a car?
- weight\_feel. How do you feel about your weight?
- love\_first. Do you believe in love at first sight?
- extra\_life. Do you believe in extraterrestrial life?
- seat. When you have a choice, where do you prefer to sit in a classroom?
- GPA. What is your college GPA?
- enough\_Sleep. Do you think you get enough sleep?
- sex. What sex are you?
- diff. Your ideal height minus your actual height.

**Format**

A data frame with 71 rows and 12 variables

**Source**

Georgetown College, MAT 111.

---

m111surveyfa12

*MAT 111 Survey, Fall 2012*

---

**Description**

Results of a survey given at beginning of semester, to all students in MAT 111.

**Format**

A data frame with 89 observations on the following 14 variables.

**height** Your height in inches.

**ideal\_ht** How tall you would LIKE to be, in inches.

**sleep** How much sleep you got last night, in hours.

**fastest** What is the highest speed at which you have ever driven a car (in mph)?

**wt\_feel** a factor with levels 1\_underweight 2\_about\_right 3\_overweight How do you feel about your weight?

**love\_first** a factor with levels no yes Do you believe in love at first sight?

**et\_life** a factor with levels no yes Do you believe in life on other planets?

**seat** a factor with levels 1\_front 2\_middle 3\_back When you have a choice, where do you prefer to sit in a classroom?

**GPA** What is your current GPA?

**engh\_slp** a factor with levels no yes Do you think you get enough sleep?

**sex** a factor with levels female male What sex are you?

**anchor** a factor with levels australia united\_states (Anchor for the next question.) For the next question, either Australia or the US, along with its population, was given in the leadup information to the question. The "anchor" variable records which version of the question you were given.

**canada** "The population of country XXX is YYY million. About what is the population of Canada, in millions?" XXX was either the U.S. or Australia.

**diff.ih.ah.** Your ideal height minus your actual height.

**Source**

MAT 111 at Georgetown College

---

m111surveyfa13

*MAT 111 Survey, Fall 2013*

---

**Description**

Results of a survey given at beginning of semester, to all students in MAT 111.

**Format**

A data frame with 85 observations on the following 14 variables.

**height** Your height in inches.

**ideal\_ht** How tall you would LIKE to be, in inches.

**sleep** How much sleep you got last night, in hours.

- fastest** What is the highest speed at which you have ever driven a car (in mph)?
- weight\_feel** a factor with levels 1\_underweight 2\_about\_right 3\_overweight How do you feel about your weight?
- love\_first** a factor with levels no yes Do you believe in love at first sight?
- extra\_life** a factor with levels no yes Do you believe in life on other planets?
- seat** a factor with levels 1\_front 2\_middle 3\_back When you have a choice, where do you prefer to sit in a classroom?
- GPA** What is your current GPA?
- enough\_Sleep** a factor with levels no yes Do you think you get enough sleep?
- sex** a factor with levels female male What sex are you?
- diff** ideal height minus actual height
- symbol** a factor with levels a b (Anchor for the next question.) For the next question, either Australia or the US, along with its population, was given in the leadup information to the question. The "anchor" variable records which version of the question you were given. If "a", the population of Australia was given. If "b", the U.S. population was given.
- pop\_Canada** "The population of country XXX is YYY million. About what is the population of Canada, in millions?" XXX was either the U.S. or Australia.

### Source

MAT 111 at Georgetown College

---

MeanSampler

*Repeated Sampling for a Mean (Slow)*

---

### Description

An app to explore the sampling distribution of the sample mean. The user takes one sample at a time from a given population. Output to the console describes relevant features of the sample, and graphical output updates the empirical distribution of the sample mean.

### Usage

```
MeanSampler(form,data,max.sample.size=30,show.sample=FALSE)
```

### Arguments

- |                              |   |
|------------------------------|---|
| <code>form</code>            | an object of class formula, of the form $\sim x$ , where $x$ is a numeric variable from the data frame supplied by: |
| <code>data</code>            | A dataframe, representing the imaginary population.   |
| <code>max.sample.size</code> | Maximum sample size on the slider.  |
| <code>show.sample</code>     | If TRUE, the complete sample will be output to the console, in addition to the summary information.                 |

**Value**

Graphical and numerical output.

**Note**

Uses manipulate.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:  
data(imagpop)  
if (require(manipulate)) MeanSampler(~income,data=imagpop)  
  
## End(Not run)
```

---

music

*Music and Reading Comprehension*

---

**Description**

An experiment performed by a student at Georgetown College. Forty-four subjects were randomized into four groups. All subjects read an article; one group read in a silent environment, while the other three groups heard each three different genres of music. Each subject took a reading comprehension test afterward.

- sex a factor with levels Female Male
- year class rank of subject
- type type of music subject listened to while reading
- score number of questions correct on reading comprehension test

**Format**

A data frame with 44 observations on 4 variables.

**Source**

Matt Doolin, MAT 111 at Georgetown College

---

 napkins

*Napkin Use*


---

**Description**

Students at GC observed their fellow students in the Cafe at lunch.

**Format**

A data frame with 86 observations on the following 2 variables.

**napkins** number of napkins used by the subject during the meal.

**sex** a factor with levels female male Sex of the person being observed

**Source**

MAT 111 at Georgetown College

---

 nonresponse

*Non-Response to Surveys*


---

**Description**

Results of a study on non-response to a mail survey. Subjects were residents of Denmark.

**Format**

A data frame with 4229 observations on the following 3 variables.

**residence** where the subject resides: either in Copenhagen, a city outside of Copenhagen, or in the countryside

**gender** sex of the subject

**response** Whether or not eh subject responded to the mail survey

**Source**

Rebuilt from a contingency table in E. B. Andersen (1991), *The Statistical Analysis of Categorical Data*. 2nd edition. Springer-Verlag, Berlin. Table found in package vcd.

---

nosmokeday

*Nicotine Withdrawal and Accidents*


---

**Description**

Results of study conducted in Great Britain to see if nicotine withdrawal increases the risk of an accident.

**Format**

A data frame with 10 observations on the following 3 variables.

**year** calendar year

**Injuries.before.NSD** number of injury accidents on the day one week prior to National No Smoke Day in the United Kingdom

**Injuries.on.NSD** number of injury accidents on National No Smoke Day in the United Kingdom

**Source**

J. Knwles, "Nicotine withdrawal and road accidents", *Science*, 400, 128, (8 July 1999). Found in Whitlock and Schluter, *The Analysis of Biological Data*.

---

oldfaithful

*Old Faithful*


---

**Description**

Old faithful geyser at Yellowstone Park.

**Format**

A data frame with 299 observations on the following 2 variables.

**Duration** duration of eruption, in minutes

**TimeNext** time until the next eruption, in minutes

**Source**

Unknown



---

ostrichtemp

*Body and Brasin Temperatures of Ostriches*

---

### Description

Body and brain temperatures of six free-ranging ostriches.

### Format

A data frame with 6 observations on the following 2 variables.

**body.temp** carotid arterial blood temperature of the ostrich (degrees Centigrade).

**brain.temp** brain temperature (near hypothalamus) of the ostrich (degrees Centigrade).

### Source

Fuller et. al., "Variability in brain and arterial blood temperatures in free-ranging ostriches in their natural habitat", The Journal of Experimental Biology 206: pp. 1171 to 1181. April 2003. <http://jeb.biologists.org/content/206/7/1171.long>. Found in Whitlock and Schluter, The Analysis of Biological Data.

---

Ovarian

*Ovarian Cancer Study*

---

### Description

Results of a retrospective study, conducted in 1973, on 299 women who been surgically treated for ovarian cancer 10 years before.

### Format

A data frame with 299 observations on the following 4 variables.

**stage** factor indicating the stage of the cancer at the time of operation (early, advanced)

**operation** factor indicating the amount of tissue removed during surgery (radical,limited)

**survival** whether or not the subject was still alive after ten years (yes,no)

**xray** factor indicating whether or not the subject also received x-ray treatments (yes,no)

### Source

Rebuilt from a contingency table in E. B. Andersen (1991), The Statistical Analysis of Categorical Data. 2nd edition. Springer-Verlag, Berlin. Table found in package vcd.

---

pbinomGC

*Graphical Calculator for Binomial Curve Probabilities*

---

### Description

Shades desired areas under rectangles of probability histogram for binomial, returns numerical value of the area.

### Usage

```
pbinomGC(bound,region="below",size=100,prob=0.5,graph=FALSE)
```

### Arguments

bound	A numerical vector of length 1 or 2, range of shaded rectangles
region	A character string. Default is "below". Possible values are "between" (when boundary consists of two numbers), "below", "above", and "outside" (again when boundary consists of two numbers)
size	Number of trials
prob	Probability of success
graph	produce graph?

### Value

Numerical value of probability.

### Author(s)

Homer White <hwhite0@georgetowncollege.edu>

### Examples

```
#This gives P(X <= 6) for binom X with 10 trials, chance of success 0.70 on each trial:
pbinomGC(6,region="below",size=10,prob=0.70)
```

```
#This gives P(45 <= X <= 55), where X is binom with 100 trials,
#chance of success on each trial p = 0.50:
pbinomGC(c(45,55),region="between",size=100,prob=0.50)
```

```
#This gives P(X >= 7) = P(X > 6), for binom X with 10 trials,
#70% chance of success on each trial
pbinomGC(6,region="above",size=10,prob=0.7)
```

---

pchisqGC

*Graphical Calculator for Chi-Square Probabilities*

---

### Description

Shades desired areas under a specified chi-square curve, returns numerical value of the area.

### Usage

```
pchisqGC(bound,region="above",df=NA,xlab="chi_square_statistic",graph=FALSE)
```

### Arguments

bound	A numerical vector of length 1, indicating boundary of shaded region on horizontal axis
region	A character string. Possible values are "below" and "above"
df	Degrees of freedom of the chi-square distribution
xlab	Label for the horizontal axis
graph	produce graph?

### Value

Numerical value of area under curve over region. Also plots the chi-square curve with the shaded area.

### Author(s)

Homer White <hwhite0@georgetowncollege.edu>

### Examples

```
#This gives P(X < 6.8) where X is chisq with 3 degrees of freedom:  
pchisqGC(6.8,df=3,region="below")  
  
#This gives P(X >= 6.8), where X is chisq with 3 degrees of freedom  
pchisqGC(6.8,df=3,region="above")
```

pennstate1

*Penn State #1*

---

**Description**

A study of students at Penn State University.

**Format**

A data frame with 190 observations on the following 9 variables.

**Sex** a factor with levels F M

**HrsSleep** how many hours of sleep the subject gets per night

**SQpick** a factor with levels Q S. Each subject was presented with two letters (S and Q), and asked to pick one. This variable indicates which letter the subject picked.

**Height** height in inches

**RandNumb** a numeric vector: Each subject was asked to choose randomly an integer from 1 to 10.

**Fastest** highest speed, in mph, at which subject has ever driven a car

**RtSpan** span of the right hand, in centimeters.

**LftSpan** span of the left hand, in centimeters.

**Form** a factor with levels QorS SorQ. The order of presentation of the S and Q options to the subject varied from one survey form to another. This variable indicates which letter was presented first on the form.

**Source**

Uts and Heckard, Mind on Statistics, 4th Edition.

---

plot.GC1m

*Diagnostic Plots for GC Linear Regression*

---

**Description**

Used by generic plot function

**Usage**

```
## S3 method for class 'GC1m'  
plot(x, ...)
```

**Arguments**

x                    An object of class GClm  
...                    ignored

**Value**

two diagnostic plots

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
SpeedModel <- lmGC(fastest~GPA,data=m111survey)  
plot(SpeedModel)
```

---

plot.polyGC

*Diagnostic Plots for GC Polynomial Regression*

---

**Description**

Used by generic plot function

**Usage**

```
## S3 method for class 'polyGC'  
plot(x,...)
```

**Arguments**

x                    An object of class polyGC  
...                    ignored

**Value**

two diagnostic plots

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
mpgModel <- polyfitGC(mpg~wt,data=mtcars)  
plot(mpgModel)
```

---

 pnormGC

*Graphical Calculator for Normal Curve Probabilities*


---

**Description**

Shades desired areas under a specified normal curve, returns numerical value of the area.

**Usage**

```
pnormGC(bound,region="below",mean=0,sd=1,graph=FALSE)
```

**Arguments**

bound	A numerical vector of length 1 or 2, indicating boundary(ies) of shaded region on horizontal axis
region	A character string. Default is "below". Possible values are "between" (when boundary consists of two numbers), "below", "above", and "outside" (again when boundary consists of two numbers)
mean	Mean of the distribution
sd	Standard deviation of the distribution
graph	Will produce graph of the probability

**Value**

Numerical value of area under curve over region.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
#This gives P(X < 75) for X normal with mean=70 and sd=4:
pnormGC(75,region="below",mean=70,sd=4)
```

```
#This gives P(X > 71) for X normal with mean=70 and sd=4:
pnormGC(71,region="above",mean=70,sd=4)
```

```
#This gives P(-1 < X < 1), for standard normal X:
pnormGC(c(-1,1),region="between")
```

```
#This gives P(X < 68 or X > 71), for X normal with mean =70 and sd=4:
pnormGC(c(68,71),region="outside",mean=70,sd=4)
```

---

Points2Watch

*A study of Influence*

---

**Description**

An app to explore the idea of influence. Note how the influence of the blue point wanes as the number of points in the central cloud increases, and also wanes as the correlation of the central cloud increases.

**Usage**

```
Points2Watch()
```

**Value**

Graphical output.

**Note**

Requires package `manipulate`, available only in RStudio. Uses `mvrnorm` from package `MASS`.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:  
if (require(manipulate)) Points2Watch()  
  
## End(Not run)
```

---

polyfitGC

*Polynomial Regression*

---

**Description**

Regression analysis (one numerical predictor variable) with simplified output. Wrapper function for `lm` in package `stats`.

**Usage**

```
polyfitGC(form,data=parent.frame(),degree=2,graph=TRUE,check=FALSE)
```

**Arguments**

form	formula of form $y \sim x$ , both variables numeric
data	dataframe supplying y and x above. If one or more of the variables is not in data, then they will be searched for in the parent environment.
degree	desired degree of polynomial (for degree 1 use <code>lmG</code> )
graph	Produce scatterplot with fitted ploynomial.
check	Asks to produce a lowess or gam curve with approximate 95 fitted line wanders outside the band, then perhaps a linear fit is not appropriate.

**Value**

A list of class "polyGC". Elements that may be queried include "s" (residual standard error) and "R^2" (unadjusted).

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
#To study the relationship between two numerical variables:
polyfitGC(mpg~wt,data=mtcars,degree=2,graph=TRUE)
#check the second-fdegree fit:
polyfitGC(mpg~wt,data=mtcars,degree=2,check=TRUE)
```

---

popsamp

*Sampling From a Population*

---

**Description**

Instructional function, and possibly a utility function for some apps.

**Usage**

```
popsamp(n,pop,...)
```

**Arguments**

n	number of items to sample
pop	data frame, from which to sample n rows
...	other arguments passed to function

**Value**

The sample, as a data frame.



**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
data(imagpop)
popsamp(10,imagpop) #Simple random sampling (no replacement)
popsamp(10,imagpop,replace=TRUE) #Random sampling with replacement
```

---

*predict.GCIm*                      *Prediction Function for GC Linear Regression*

---

**Description**

Used by generic predict function

**Usage**

```
## S3 method for class 'GCIm'
predict(object,x,level=NULL,...)
```

**Arguments**

object	An object of class GCIm
x	value of the predictor variable
level	desired level of prediction interval
...	ignored

**Value**

numeric prediction

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
#predict fastest speed driven, for person with GPA=3.0:
SpeedModel <- lmGC(fastest~GPA,data=m111survey)
predict(SpeedModel,x=3.0)
#include prediction interval:
predict(SpeedModel,x=3.0,level=0.95)
```

predict.polyGC

*Prediction Function for GC Polynomial Regression*

---

**Description**

Used by generic predict function

**Usage**

```
## S3 method for class 'polyGC'  
predict(object,x,level=NULL,...)
```

**Arguments**

object	An object of class polyGC
x	value of the predictor variable
level	desired level of prediction interval
...	ignored

**Value**

numeric prediction

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
#predict mpg for a car weighing 3 tons:  
mpgModel <- polyfitGC(mpg~wt,data=mtcars,degree=2)  
predict(mpgModel,x=3.0)  
#include prediction interval:  
predict(mpgModel,x=3.0,level=0.95)
```

---

PropSampler*Repeated Sampling for a Proportion (Slow)*

---

**Description**

An app to explore the sampling distribution of the sample proportion. The user takes one sample at a time from a given population. Output to the console describes relevant features of the sample, and graphical output updates the empirical distribution of the sample proportion.

**Usage**

```
PropSampler(form,data,max.sample.size=110,show.sample=FALSE)
```

**Arguments**

form	An object of class formula, of the form $\sim x$ , where $x$ is a factor from the data frame supplied by:
data	A dataframe, representing the imaginary population.
max.sample.size	Maximum sample size on the slider.
show.sample	If TRUE, the complete sample will be output to the console, in addition to the summary information.

**Value**

Graphical and numerical output.

**Note**

Uses manipulate.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:
data(imagpop)
if (require(manipulate)) PropSampler(~cappun,data=imagpop)

## End(Not run)
```

---

proptestGC

*Proportions Procedures*

---

**Description**

Employs the normal approximation to perform test for one or two proportions.

**Usage**

```
proptestGC(x,data=parent.frame(),n=numeric(),p=NULL,
           alternative=c("two.sided","less","greater"),
           success="yes",first=NULL,conf.level=0.95,
           correct=TRUE,graph=FALSE,verbose=TRUE)
```

**Arguments**

x	Either a formula or a numeric vector. If formula, it must be of the form $\sim x$ indicating the single variable under study, or of the form $\sim x+y$ , in which case $x$ is the explanatory grouping variable (categorical with two values) and $y$ is the response categorical variable with two values. When summary data are provided, $x$ is a numeric vector of success counts.
data	Data frame that supplies the variables $x$ and $y$ . If any are not in data, then they will be searched for in the parent environment.
n	When not empty, this is a numeric vector giving the size of each sample.
p	Specifies Null Hypothesis value for population proportion. If not set, no test is performed.
alternative	"two.sided" requests computation of a two-sided P-value; other possible values are "less" and "greater".
success	When $x$ is a formula, this argument indicates which value of variable $x$ (in case of $\sim x$ ) or $y$ (in case of $\sim x+y$ ) is being counted as a success. When working with formula-data input the value of this parameter MUST be set, even when the variable has only two values.
first	When performing 2-sample procedures, this argument specifies which value of the explanatory variable constitutes the first group.
conf.level	Number between 0 and 1 indicating the confidence-level of the interval supplied.
correct	Applies continuity correction for one-proportion procedures. It is ignored when 2-proportions are performed.
graph	If TRUE, plot graph of P-value.
verbose	Indicates how much output goes to the console

**Value**

A list, either of class "gcp1test" (one-proportion) or "gcp2test" (two proportions). Components of this list that may be usefully queried include: "statistic", "p.value", and "interval".

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
data(m111survey)
#2-proportions, formula-data input, 95%-confidence interval only:
proptestGC(~sex+seat,data=m111survey,success="2_middle")

#For other confidence levels, use argument conf.level. For 90%-interval for one proportion p:
proptestGC(~sex,data=m111survey,success="male",conf.level=0.90)

#one proportion, formula-data input, confidence interval and two-sided test with H_0: p = 0.33:
proptestGC(~seat,data=m111survey,success="1_front",p=0.33)
```

```

#Summary data:
#In first sample, 23 successes out of 100 trials. In second sample, 33 out of 110.
proptestGC(x=c(23,33),n=c(100,110))

#Summary data:
#In one sample, 40 successes in 100 trials. Testing whether p = 0.45.
proptestGC(x=40,n=100,p=0.45,correct=TRUE)

#Want less output? Set argument verbose to FALSE:
proptestGC(~sex+seat,data=m111survey,success="2_middle",p=0.33,verbose=FALSE)

```

---

ptGC

*Graphical Calculator for t-Curve Probabilities*


---

## Description

Shades desired areas under a specified t-curve, returns numerical value of the area.

## Usage

```
ptGC(bound,region="between",df=1,graph=FALSE)
```

## Arguments

bound	A numerical vector of length 1 or 2, indicating boundary(ies) of shaded region on horizontal axis
region	A character string. Possible values are "between" (when boundary consists of two numbers), "below", "above", and "outside" (again when boundary consists of two numbers)
df	degrees of freedom of the distribution
graph	produce graph?

## Value

Numerical value of area under curve over region. Also plots the t-curve with the shaded area.

## Author(s)

Homer White <hwhite0@georgetowncollege.edu>

## Examples

```
#This gives P(-2 < t < 2) for a t-random variable with 1 degree of freedom:
ptGC(c(-2,2),region="between",df=1)
```

```
#This gives P(t < -1) for a t-random variable with 5 degrees of freedom:
ptGC(-1,region="below",df=5)
```

```
#This gives P( t < -2 OR t >2), for a t-random variable with 5 degrees of freedom:
ptGC(c(-2,2),region="outside",df=5)
```

---

 pushups

*Pushups by Football Players at Georgetown College*


---

**Description**

Two football players at GC asked their team-mates to do as many pushups as they could in two minutes.

**Format**

A data frame with 30 observations on the following 3 variables.

**weight** weight of subject in pounds.

**pushups** number of pushups completed.

**position** a factor with levels LINE SKILL: type of position played by the subject. Line positions require high body mass, skill positions require a lot of running around.

**Source**

MAT 111, Georgetown College

---

 qnormGC

*Graphical Calculator for Normal Curve Percentiles*


---

**Description**

When you know a certain area under a normal density curve, this function returns the x-axis values of the boundary of that area.

**Usage**

```
qnormGC(area, region="below", mean=0, sd=1, graph=FALSE)
```

**Arguments**

area	The known percentile
region	A character string. Default is "below". Other possible values are "between" (when known area is symmetric around the mean two numbers), "below", "above", and "outside" (when know area is outside a region symmetric around the mean)
mean	Mean of the distribution
sd	Standard deviation of the distribution
graph	Will produce graph of the area

**Value**

Numerical value of the percentile, and a vector when there are two bounds.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
#80th percentile of a normal distribution with mean=70 and sd=4:
```

```
qnormGC(0.80,region="below",mean=70,sd=4)
```

```
#Return value x so that P(X > x) = 0.10 (same as the 90th percentile)
```

```
qnormGC(0.10,region="above",mean=70,sd=4)
```

```
#This gives the multiplier for 95%-confidence intervals based on the z-statistic
```

```
qnormGC(0.95,region="between")
```

```
#This gives critical values for a two-sided z-test with alpha = 0.01:
```

```
qnormGC(0.01,region="outside")
```

---

 RandomExp

*Randomized Experimental Designs*


---

**Description**

Randomizes subjects into treatment groups according to specified criteria.

**Usage**

```
RandomExp(data,sizes=NULL,groups=NULL,block=NULL,seed=NULL)
```

**Arguments**

data	A data frame containing the subjects to be randomized
sizes	a numeric vector indicating the sizes of the treatment groups. Vector must sum to the number of subjects. If not provided, subjects will be randomized into two groups of size as nearly equal as possible.
groups	a character vector giving the names of the groups. Names correspond to sizes specified in previous sizes. Length of groups must equal length of sizes.
block	Variable(s) in the data frame with respect to which blocking is performed. In order to block with respect to more than one variable at once, enter as character vector, e.g.: c("Var1","Var2").
seed	randomization seed, for reproducibility of results.

**Value**

A data frame: the input frame data augmented with a variable `treat.grp` indicating the assignment of subjects to groups.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
data(SmallExp) #small hypothetical list of subjects

#completely randomized design
RandomExp(SmallExp)

#Block with respect to sex:
RandomExp(SmallExp,sizes=c(8,8),groups=letters[1:2],block="sex")

#Block for both sex and athletic status:
RandomExp(SmallExp,sizes=c(8,8),groups=letters[1:2],block=c("sex","athlete"))
```

---

RegEstimate

*Estimation of Regression Coefficients*

---

**Description**

An app to explore estimation of coefficients in simple regression.

**Usage**

```
RegEstimate(x=1:10)
```

**Arguments**

`x` A numerical vector, specifying the fixed set of x-values.

**Value**

Graphical and numerical output.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:
if (require(manipulate)) RegEstimate()

## End(Not run)
```



---

`rowPerc`*Row Percents*

---

**Description**

Computes row percentages for a given twoway table.

**Usage**

```
rowPerc(tab)
```

**Arguments**

`tab` A table, e.g., the result of `xtabs(~var1+var2, data=DataFrame)`.

**Value**

An object of class `table`, giving row percentages for the input table.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
data(ledgejump)
MyTable <- xtabs(~weather+crowd.behavior, data=ledgejump)
rowPerc(MyTable)
```

---

`saltmarsh`*Effect of Soil Salinity on Plant Growth*

---

**Description**

Result of an experiment conducted to investigate the effect of salinity level in soil on the growth of plants.

**Format**

A data frame with 24 observations on the following 3 variables.

**salt** amount of salt applied to the plot (in parts per million)

**biomass** total biomass of plot at the end of the study period (units unknown)

**block** field in which the plot was located

**Details**

From the source (see below): "Experimental fields of land were located at an agricultural field station, and each field was divided into six smaller plots. Each of the smaller plots was treated with a different amount of salt (measured in ppm) and the biomass at the end of the experiment was recorded."

**Source**

The Course Notes of Carl Schwarz, Simon Fraser University: <http://people.stat.sfu.ca/~cschwarz/CourseNotes/>

---

SampDist2Means

*Distribution of the Difference of Sample Means*

---

**Description**

An app to explore the Central Limit Theorem in the context of the difference of sample means.

**Usage**

```
SampDist2Means(pop,max.samp.sizes=50,sim.reps=1000)
```

**Arguments**

pop	A data frame representing the population from which samples are taken.
max.samp.sizes	Largest sample sizes shown on the sliders.
sim.reps	Number of simulation repetitions to construct empirical distribution of difference of sample means.

**Value**

Graphical and numerical output.

**Note**

Uses `manipulate` in RStudio. Also requires package `lattice`.

**Author(s)**

Homer White <[hwhite0@georgetowncollege.edu](mailto:hwhite0@georgetowncollege.edu)>

**Examples**

```
## Not run:  
data(imagpop)  
if (require(manipulate)) SampDist2Means(imagpop)  
  
## End(Not run)
```

---

SampDist2Props	<i>Difference of Two Sample Proportions</i>
----------------	---

---

**Description**

An app to explore the sampling distribution of the difference of two sample proportions.

**Usage**

```
SampDist2Props(form,data,max.sample.sizes=100,sim.reps=1000)
```

**Arguments**

form	An object of class formula, of the form $\sim x+y$ where x and y are factors supplied by:
data	A dataframe, representing the imaginary population. In the formula, both factors should have exactly two levels. The variable x represents the explanatory variable.
max.sample.sizes	Maximum sample sizes allowed on the sliders.
sim.reps	Number of samples to construct the empirical distribution.

**Value**

Graphical and numerical output.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:  
data(imagpop)  
SampDist2Props(~sex+cappun,data=imagpop)  
  
## End(Not run)
```

---

SampDistMean	<i>Distribution of the Sample Mean</i>
--------------	--

---

**Description**

An app to explore the Central Limit Theorem.

**Usage**

```
SampDistMean(pop,max.samp.size=50,sim.reps=1000)
```

**Arguments**

pop	A data frame representing the population from which samples are taken.
max.samp.size	Largest sample size shown on the slider.
sim.reps	Number of simulation repetitions to construct empirical distribution of the sample mean.

**Value**

Graphical and numerical output.

**Note**

Uses manipulate in RStudio.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:  
data(imagpop)  
if (require(manipulate)) SampDistMean(imagpop)  
  
## End(Not run)
```

---

 sat

*SAT Scores*


---

### Description

SAT scores by state. The variables are as follows:

- state. A state in the U.S.
- expend. Mean annual expenditure per student (in 1000\$).
- ratio. Mean student-teacher ratio.
- salary. Mean annual teacher salary.
- frac. Percentage of students in the state who take the SAT.
- verbal. Mean SAT Verbal score for the state.
- math. Mean SAT Math score for the state.
- sat. Sum of mean Verbal and mean Math.

### Format

A data frame with 50 rows and 8 variables

### Source

Deborah Lynn Guber, "Getting what you pay for: the debate over equity in public school expenditures" (1999), *Journal of Statistics Education* 7(2). <http://www.amstat.org/publications/jse/secure/v7n2/datasets.guber.cfm>.

---

 sealsO2

*Weddell Seal Oxygen Consumptions*


---

### Description

Results of an experiment conducted on ten Weddell seals.

### Format

A data frame with 10 observations on the following 2 variables.

**o2.nonfeeding** Oxygen consumption during recovery time after a dive during which no plankton was consumed by the seal, in ml of O2 per kilogram of weight

**o2.feeding** Oxygen consumption during recovery time after a dive during which plankton was consumed, in ml of O2 per kilogram of weight

### Source

Williams, T. M., L. A. Fuiman, M. Horning, and R. W. Davis. 2004. *The Journal of Experimental Biology* 207: 973 to 982. <http://jeb.biologists.org/content/207/6/973.full>

---

`ShallowReg`*Regression Line Too Shallow?*

---

**Description**

The regression line is not as steep as the SD Line (line through point of averages, with slope =  $sd(y)/sd(x)$ ). The difference is especially noticeable when the scatterplot is the result of a sample from a bivariate normal distribution. This app explains why we use the regression line to predict  $y$  from  $x$ , even though the SD line appears to be a better linear summary of the scatterplot. Can be used as a starting-point for a discussion of "regression to the mean."

**Usage**

```
ShallowReg(n=900, rho=0.5)
```

**Arguments**

<code>n</code>	Number of points in the scatterplot.
<code>rho</code>	Target correlation for the scatterplot. Points are selected from a standardized bivariate normal distribution, with correlation $\rho$ .

**Value**

Graphical output.

**Note**

Uses `manipulate`, available only in RStudio, and `mvrnorm` from package `MASS`.

**Author(s)**

Homer White <[hwhite0@georgetowncollege.edu](mailto:hwhite0@georgetowncollege.edu)>

**Examples**

```
## Not run:  
if (require(manipulate)) ShallowReg()  
  
## End(Not run)
```

---

`SimpleRandom`*Simple Random Sample*

---

**Description**

An app to investigate the visual and numerical differences between a sample and a population. A sample is drawn from the input population and then a variable of choice is selected by the user. If a categorical variable is chosen, the user sees a bar chart with red bars designating the population and blue bars designating the sample. Simultaneously, a summary table (of percents) is output to the console for both the population and the sample. If a numerical variable is chose, the kernel density plot for the population is plotted in red and the histogram for each new sample is plotted in blue. Simultaneously, the summary information for minimum, maximum, quartiles, median, mean, and standard deviation are output to the console for both the population and the sample. The size of the sample can be changed to explore how this affects statistics and the plots.

**Usage**`SimpleRandom()`**Value**

Graphical and numerical output

**Author(s)**

Rebekah Robinson &lt;rebekah\_robinson@georgetowncollege.edu&gt;

**Examples**

```
## Not run:  
if (require(manipulate)) SimpleRandom()  
  
## End(Not run)
```

---

`SimpleRandom2`*Simple Random Sample (v2)*

---

**Description**

Similar to [SimpleRandom](#), but with a fixed sample size set by the user.

**Usage**`SimpleRandom2(n=100)`

**Arguments**

n                    the desired sample size

**Value**

Graphical and numerical output

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:  
if (require(manipulate)) SimpleRandom()  
  
## End(Not run)
```

---

Skewer

*SkewR*

---

**Description**

An app to illustrate the effect of skewness on the shape of a boxplot.

**Usage**

Skewer()

**Value**

Graphical output.

**Note**

Requires `manipulate`; uses functions from package `lattice`

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:  
if (require(manipulate)) Skewer()  
  
## End(Not run)
```



---

SlowGoodness	<i>Chi Square Resampler (One at a Time) for Goodness-of-Fit</i>
--------------	---

---

**Description**

An app to illustrate use of the chi-square statistic to test for goodness of fit. The P-value is computed by resampling, and the resamples are done one at a time. A histogram of resampled chi-square statistics is displayed after each resample, and summary information is output to the console.

**Usage**

```
SlowGoodness(x, p)
```

**Arguments**

x	a one-dimensional table, or a vector of observed counts
p	vector of null probabilities

**Value**

Graphical and numerical output

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:  
throws <- c(one=8, two=18, three=11, four=7, five=9, six=7)  
SlowGoodness(throws, p=rep(1/6, 6))  
  
## End(Not run)
```

---

SmallExp	<i>A Small Experiment</i>
----------	---------------------------

---

**Description**

Subjects in a hypothetical experiment

**Format**

A data frame with 16 observations on the following 3 variables.

**name** name of the subject

**sex** sex of the subject

**athlete** whether or not the subject is an athlete

---

 stumps

*Larvae on Stumps*


---

### Description

Biologists were interested in whether beetles prefer areas where beavers have cut down cottonwood trees. (The tree-stumps produce tender green shoots that beetles are thought to like.) 23 circular plots, all of equal area, were studied. For each plot the researchers counted the number of cottonwood stumps, and also the number of clusters of beetle larvae found in the plot.

### Format

A data frame with 23 observations on the following 2 variables.

**stumps** number of stumps in the plot

**larvae** number of larvae clusters in the plot

### Source

Basic Practice of Statistics, by Moore and McCabe.

---

 temperature

*Temperature in U.S. Cities*


---

### Description

Average temperatures for cities in the United States.

- **city** Name of the city
- **latitude** latitude of the city, in degrees north of the Equator
- **JanTemp** mean temperature of the city in January.
- **AprTemp** mean temperature of the city in April.
- **AugTemp** mean temperature of the city in August.

### Format

A data frame with 20 observations on 5 variables.

### Source

Mind on Statistics, 4th edition, Uts and Heckard.

---

`tExplore`*Exploration of the t-Distributions*

---

**Description**

Plot the density curve of a t random variable at various degrees of freedom. Compare with the standard normal curve.

**Usage**

```
tExplore()
```

**Value**

Used only for graphical side effects.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:  
if (require(manipulate)) tExplore()  
  
## End(Not run)
```

---

`theme.rpres`*Lattice Theme or R Presentations*

---

**Description**

Modifies the current theme for use with lattice graphics in R Presentation documents. Increases size of title, axis labels and axis numbers, thickens some lines, etc.

**Usage**

```
theme.rpres()
```

**Value**

Returns a list to be supplied as the theme to the lattice function `trellis.par.set()`.

**Note**

Deprecated in favor of `themerpres()`. May not appear in future versions.

**See Also**

[trellis.par.set](#), [show.settings](#)

**Examples**

```
trellis.par.set(theme=theme.rpres())
```

---

themerpres

*Lattice Theme or R Presentations*

---

**Description**

Modifies the current theme for use with lattice graphics in R Presentation documents. Increases size of title, axis labels and axis numbers, thickens some lines, etc.

**Usage**

```
themerpres()
```

**Value**

Returns a list to be supplied as the theme to the lattice function [trellis.par.set\(\)](#).

**See Also**

[trellis.par.set](#), [show.settings](#)

**Examples**

```
trellis.par.set(theme=themerpres())
```

---

tips

*Restaurant Tips*

---

**Description**

A waiter recorded all tips earned during a 2.5 month period in early 1990, along with other information about the customers who gave the tips. The variables are as follows:

- obs Observation number of event
- totbill The total bill
- tip Amount of the tip
- sex Sex of tipper (F or M)
- smoker Whether the tipper smokes: Yes or No
- day Day of the week
- time Whether the meal was during the Day or the Night
- size Number of people in the dining party

**Format**

A data frame with 244 rows and 8 variables

**Source**

Bryant, P. G. and Smith, M. A.(1995), Practical Data Analysis: Case Found in Interactive and Dynamic Graphics for Data Analysis: With Examples Using R and GGobi, Dianne Cook and Deborah F. Swayne. See also <http://www.ggobi.org/book/>.

---

tornado	<i>Tornado Damage</i>
---------	-----------------------

---

**Description**

Tornado damage in the U.S., by state. Also includes Puerto Rico.

- state the state
- damage mean annual damage from tornados, over a five-year period, in millions of dollars

**Format**

A data frame with 51 observations on 2 variables.

**Source**

Moore and McCabe, The Basic Practice of Statistics.

---

tSampler	<i>Illustrating the t-statistic</i>
----------	-------------------------------------

---

**Description**

An app to explore the distribution of the t-statistic. The user takes one sample at a time from a given population. Graphical output updates the empirical distribution of the sample mean.

**Usage**

```
tSampler(form,data,max.sample.size=30,show.sample=FALSE)
```

**Arguments**

form	An object of class formula, of the form $\sim x$ , where $x$ is a numeric variable from the data frame supplied by:
data	A dataframe, representing the imaginary population.
max.sample.size	Maximum sample size on the slider.
show.sample	If TRUE, the complete sample will be output to the console, in addition to the summary information.

**Value**

Graphical and numerical output.

**Note**

Uses `manipulate`.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:
data(imagpop)
if (require(manipulate)) tSampler(~income,data=imagpop)

## End(Not run)
```

---

ttestGC

*t-Procedures*


---

**Description**

t-tests and confidence intervals for one and two samples.

**Usage**

```
ttestGC(x=NULL,data=parent.frame(),mean=numeric(),sd=numeric(),n=numeric(),
mu=NULL,alternative=c("two.sided","less","greater"),var.equal=FALSE,
conf.level=0.95,graph=FALSE,first=NULL,verbose=TRUE)
```

**Arguments**

x	If not <code>NULL</code> , then must be a formula. If a formula, then data must be a dataframe. For one sample t-procedures, x is of the form <code>~var</code> . For two-sample procedures, x is of the form <code>resp~exp</code> , where <code>exp</code> is factor with two values. If x is of form <code>~var1-var2</code> , then matched pairs procedures are performed.
data	A data frame containing variables in formula x. If some variables are not in data, then they are searched for in the parent environment.
mean	When not <code>NULL</code> , contains sample mean(s). Length 1 for one sample t-procedures, Length 2 for two-sample procedures.
sd	When not <code>NULL</code> , contains sample standard deviation(s).
n	When not <code>NULL</code> , contains sample size(s).
mu	Contains the null value for the parameter of interest. If not set, no test is performed.

alternative	"two.sided" requests computation of a two-sided P-value; other possible values are "less" and "greater".
var.equal	When FALSE, use Welch's approximation to the degrees of freedom.
conf.level	Number between 0 and 1 indicating the confidence-level of the interval supplied.
graph	If TRUE, plot graph of P-value.
first	If assigned, gives the value of the explanatory variable that is to count as the first sample.
verbose	Indicate how much output goes to console

### Value

A list of class "GCttest" Components of the list that may be usefully queried include: "statistic", "p.value", and "interval".

### Author(s)

Homer White <hwhite0@georgetowncollege.edu> for matched pairs.

### Examples

```
#One-sample t, 95%-confidence interval only:
ttestGC(~fastest,data=m111survey)

#For other confidence levels, set argument conf.level as desired. For 90%-interval:
ttestGC(~fastest,data=m111survey,conf.level=0.90)

# One-sample t, 95%-confidence interval and two-sided test with H_0: mu = 100:
ttestGC(~fastest,data=m111survey,mu=100)

#Two-sample t, 95%-confidence interval only:
ttestGC(fastest~sex,data=m111survey)

#control order of groups with argument first:
ttestGC(fastest~sex,data=m111survey,first="male")

# Matched pairs, confidence interval with one-sided test, H_0: mu-d = 0:
ttestGC(~ideal_ht-height,data=m111survey,mu=0,alternative="greater")

#Summary data, one sample, one-sided test with H_0: mu = 52.5:
ttestGC(mean=55,sd=4,n=16,mu=52.5,alternative="greater")

#Summary data, two samples:
ttestGC(mean=c(50,55),sd=c(3,4),n=c(25,40),mu=0)
```

---

 Type12Errors

*Type I and Type II Errors*


---

**Description**

An app to explore the concepts of Type I and Type II errors, and the concept of power. We take samples from a population that is imagined to be normal, and perform the t-procedures for one mean. The Null Hypothesis is  $H_0: \mu=170$ . A slider allows us to vary the true mean  $\mu$ .

**Usage**

Type12Errors()

**Value**

Graphical and numerical output.

**Note**

Uses manipulate.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:
if (require(manipulate)) Type12Errors()

## End(Not run)
```

---

 ucdavis1

*UC Davis #1*


---

**Description**

Results of a survey of students at UC-Davis.

- Sex a factor with levels Female Male
- TV Number of hours spent watching TV per week
- computer number of hours spent on computer per week
- Sleep hours of sleep per night
- Seat a factor with levels Back Front Middle Where do you prefer to sit in class, when you have a choice?



- alcohol number of alcoholic drinks consumed per week
- Height height in inches
- momheight height of mother, in inches
- dadheight height of father, in inches
- exercise number of hours of exercise per week
- GPA grade point average
- class a factor with levels LibArts NonLib Student Category: liberal arts or not

**Format**

A data frame with 173 observations on 12 variables..

**Source**

Mind on Statistics, 4th edition, Uts and Heckard.

---

Variability

*Variability*

---

**Description**

An app to investigate how the variance and sample size affects the shape of a histogram and violin plot generated from normal data. Summary data (minimum, median, mean, maximum, and quartiles) are displayed in the output for each random sample drawn.

**Usage**

```
Variability()
```

**Value**

Graphical and numerical output

**Note**

Uses histogram and bwplot from the lattice package.

**Author(s)**

Rebekah Robinson <rebekah\_robinson@georgetowncollege.edu>

**Examples**

```
## Not run:  
if (require(manipulate)) Variability()  
  
## End(Not run)
```

---

VaryCorrelation      *Vary Correlation*

---

**Description**

An app to illustrate the effectiveness of the correlation coefficient as a measure of the strength of a linear relationship.

**Usage**

```
VaryCorrelation(n=300)
```

**Arguments**

n                      number of randomly generated-points in the scatterplot.

**Value**

Graphical output.

**Note**

Uses manipulate in RStudio, and mvrnorm from package MASS.

**Author(s)**

Homer White <hwhite0@georgetowncollege.edu>

**Examples**

```
## Not run:  
if(require(manipulate)) VaryCorrelation(n=500)  
  
## End(Not run)
```

---

verlander                      *Justin Verlander*

---

**Description**

PITCHf/x data on Justin Verlander, winner of the 2011 Cy Young Award. Covers his 2009 through 2012 seasons.

**Format**

A data frame with 15307 observations on the following 12 variables. Each observation is a single pitch.

**season** The year of play

**gamedate** Date of the game in which the pitch was thrown

**pitch\_type** Type of pitch thrown: CH (Change-up), CU (Curveball), FF (Four-Seam Fastball), FT (Two-Seam Fastball), SL (Slider). Pitch type is determined by computer algorithm.

**balls** Current ball count

**strikes** Current strike count

**pitches** number of pitches previously thrown in the game

**speed** speed of pitch (in mph). (When crossing plate?)

**px** x-coordinate of pitch (in feet, measured from center of plate)

**pz** vertical coordinate of pitch (in feet above plate)

**px\_x** the horizontal movement, in inches, of the pitch between the release point and home plate, as compared to a theoretical pitch thrown at the same speed with no spin-induced movement. Measured at 40 feet from home plate.

**px\_z** the vertical movement, in inches, of the pitch between the release point and home plate, as compared to a theoretical pitch thrown at the same speed with no spin-induced movement. Measured at 40 feet from home plate.

**batter\_hand** A factor with two values: L (left) and R (right).

**Source**

Marchi and Albert: Analyzing Baseball Data with R, CRC Press 2014. For more on the PITCHf/x system, see <http://en.wikipedia.org/wiki/PITCHf/x>.

---

 youthrisk03

 Youth Risk 2003
 

---

**Description**

A Study of Risky Behaviors in High School Seniors, from year 2003.

- Sex
- Grades Typical grades you earn in school
- WtAction What do you plan to do about your weight?
- Seatbelt How often do you wear a seat-belt?
- Sunscreen How much do you wear sunscreen?
- Grades\_1 Same as grades, but with some groups combined
- Sun\_1 Same as sunscreen, but with some groups combined

**Format**

A data frame with 3042 observations on 7 variables.

**Source**

Mind on Statistics, 4th Edition, by Uts and Heckard.

# Index

## \*Topic **datasets**

- alcohol, 4
- attitudes, 5
- beans, 8
- cabrera, 11
- chugtime, 14
- deathpen, 16
- FakeSchool, 20
- fuel, 21
- galton, 22
- gcfeeling, 22
- gcstudents, 23
- gss02, 23
- gss08, 24
- gss2012, 25
- hair\_and\_act, 25
- handheight, 26
- hanford1, 26
- hanford2, 27
- henderson, 28
- hofbatting, 29
- hofpitching, 30
- imagpop, 31
- iqsiblings, 32
- knifeorgunblock, 32
- labels, 33
- ledgejump, 33
- m111survey, 35
- m111surveyfa12, 35
- m111surveyfa13, 36
- music, 38
- napkins, 39
- nonresponse, 39
- nosmokeday, 40
- oldfaithful, 40
- ostrichtemp, 41
- Ovarian, 41
- pennstate1, 44
- pushups, 54

- saltmarsh, 57
- sat, 61
- seals02, 61
- SmallExp, 65
- stumps, 66
- temperature, 66
- tips, 68
- tornado, 69
- ucdavis1, 72
- verlander, 74
- youthrisk03, 75

## \*Topic **graphics**

- theme.rpres, 67
- themerpres, 68

## \*Topic **package**

- tigerstats-package, 4

- alcohol, 4
- attitudes, 5
  
- barchart, 6
- barchartGC, 6
- beans, 8
- BinomNorm, 8
- BinomSkew, 9
- binomtestGC, 9

- cabrera, 11
- ChisqSimSlow, 11
- chisqtestGC, 12
- chugtime, 14
- CIMean, 14
- CIProp, 15
- colPerc, 16

- deathpen, 16
- DtrellHist, 17
- DtrellScat, 17

- EmpRule, 18
- EmpRuleGC, 19

- expCounts, 20
- FakeSchool, 20
- FindRegLine, 21
- fuel, 21
- galton, 22
- gcfeeling, 22
- gcstudents, 23
- gss02, 23
- gss08, 24
- gss2012, 25
- hair\_and\_act, 25
- handheight, 26
- hanford1, 26, 27
- hanford2, 27
- helpGC, 27
- henderson, 28
- hofbatting, 29
- hofpitching, 30
- imagpop, 31
- iqsiblings, 32
- knifeorgunblock, 32
- labels, 33
- ledgejump, 33
- lmGC, 34
- m111survey, 35
- m111surveyfa12, 35
- m111surveyfa13, 36
- MeanSampler, 37
- music, 38
- napkins, 39
- nonresponse, 39
- nosmokeday, 40
- oldfaithful, 40
- ostrichtemp, 41
- Ovarian, 41
- pbinomGC, 42
- pchisqGC, 43
- pennstate1, 44
- plot.GClm, 44
- plot.polyGC, 45
- pnormGC, 46
- Points2Watch, 47
- polyfitGC, 47
- popsamp, 48
- predict.GClm, 49
- predict.polyGC, 50
- PropSampler, 50
- proptestGC, 51
- ptGC, 53
- pushups, 54
- qnormGC, 54
- RandomExp, 55
- RegEstimate, 56
- rowPerc, 57
- saltmarsh, 57
- SampDist2Means, 58
- SampDist2Props, 59
- SampDistMean, 60
- sat, 61
- seals02, 61
- ShallowReg, 62
- show.settings, 68
- SimpleRandom, 63, 63
- SimpleRandom2, 63
- Skewer, 64
- SlowGoodness, 65
- SmallExp, 65
- stumps, 66
- temperature, 66
- tExplore, 67
- theme.rpres, 67
- themerpres, 68
- tigerstats (tigerstats-package), 4
- tigerstats-package, 4
- tips, 68
- tornado, 69
- trellis.par.set, 67, 68
- tSampler, 69
- ttestGC, 70
- Type12Errors, 72
- ucdavis1, 72
- Variability, 73
- VaryCorrelation, 74
- verlander, 74

xtabs, [6](#)

youthrisk03, [75](#)