

# Package ‘dafs’

July 22, 2025

**Version** 1.0-38

**Date** 2022-04-11

**Title** Data Analysis for Forensic Scientists

**Author** James Curran, Danny Chang

**Maintainer** James Curran <j.curran@auckland.ac.nz>

**Depends** s20x

**Description** Data and miscellanea to support the book ``Introduction to Data analysis with R for Forensic Scientists." This book was written by James Curran and published by CRC Press in 2010 (ISBN: 978-1-4200-8826-7).

**License** GPL (>= 2)

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2022-04-11 09:12:33 UTC

## Contents

abaz.df . . . . .	3
abduct.age.df . . . . .	4
anneal.df . . . . .	4
anscombe.df . . . . .	5
bennett.df . . . . .	5
bottle.df . . . . .	6
casework.df . . . . .	6
cck196.df . . . . .	7
cck45.df . . . . .	7
docexam.df . . . . .	8
dpd.df . . . . .	8
fiber.color.df . . . . .	9
fitPvalue . . . . .	9
formatScientificTeX . . . . .	10
gc.df . . . . .	11
ghb.df . . . . .	11

gustafson.df . . . . .	12
halfnorm . . . . .	12
intPlot . . . . .	13
jauhari1.df . . . . .	15
jauhari2.df . . . . .	15
kent.df . . . . .	16
kerckring.df . . . . .	17
last . . . . .	17
liver.df . . . . .	18
morphine.df . . . . .	18
nasal.spline.df . . . . .	19
newton.df . . . . .	19
nzglass.df . . . . .	20
pairsDAFS . . . . .	20
palatal.df . . . . .	21
plotBoth . . . . .	21
plotBoth.control . . . . .	22
plotRegDiagPlots . . . . .	23
printANODEVTable . . . . .	23
printANOVATable . . . . .	24
printCI . . . . .	25
printRegTable . . . . .	26
pvalue . . . . .	26
ri.calibration.df . . . . .	27
ri.calibration2.df . . . . .	28
Rsq . . . . .	28
salting1.df . . . . .	29
salting2.df . . . . .	30
sex.age.df . . . . .	31
shadeDens . . . . .	31
shotgun.df . . . . .	32
squamous.df . . . . .	33
StangleBook . . . . .	33
SweaveBook . . . . .	34
tryptase.df . . . . .	35
velocity.df . . . . .	35
vhcdt.df . . . . .	36
wong.df . . . . .	37

---

abaz . df                      *DNA from drinking containers*

---

## Description

The amount of DNA left on different types of drinking containers.

## Usage

data(abaz . df)

## Format

A data frame with 21 columns.

[,1]	person	factor	A label (A..F) for the 6 experimental subjects
[,2]	sample	factor	A treatment factor indicating the different beverage/container combinations
[,3]	ab.sample	factor	An abbreviated treatment label
[,4]	time	factor	time when DNA concentration was measured. Levels: 24hrs, 48hrs
[,5]	amylase	numeric	the relative amount of alpha-amylase activity
[,6]	quant	numeric	
[,7]	amp.volume	numeric	
[,8]	dna.conc	numeric	
[,9]	gel.profile	factor	
[,10]	failed.profile	factor	the failure or success of obtaining a usable DNA profile
[,11:21]	d3..fga	numeric	Total peak heights at each locus

## Author(s)

Abaz et al.

## References

Abaz, J., Walsh, S.J., Curran, J.M., Moss, D.S., Cullen, J., Bright, J.A., Crowe, G.A., Cockerton, S.L. and Power, T.E. 'Comparison of the variables affecting the recovery of DNA from common drinking containers' *Forensic Sci Int.* 2002 May 23;126(3):233-40.

---

abduct.age.df                      *Victim abduct and the age data*

---

### Description

This data was studied to investigate whether there was a relationship between whether the victim had been abducted and the age of the victims in certain crimes. The age of the victims had been classified as 0-10 and 11+.

### Usage

```
data(abduct.age.df)
```

### References

C. G. G. Aitken, T. Connolly, A. Gammernan, G. Zhang, D. Bailey, R. Gordon, and R. Oldfield. *Statistical modelling in specific case analysis*. Science & Justice, 36(4):245-255, October 1996.

---

anneal.df                              *Annealing float glass*

---

### Description

Rushton was interested in the effect of annealing on the refractive index of glass. It is well known that annealing float glass changes the refractive index (RI). The change in RI - called delta-RI - can tell a forensic scientist something about the glass that they are examining.

In this data set, 3 replicate measurements were made from 150 squares of glass from a single pane. Each fragment was measured pre and post annealing.

### Usage

```
data(anneal.df)
```

### Format

A data frame with 900 observations on 3 variables.

[,1]	ri	numeric	The fragment's refractive index
[,2]	temp	numeric	The fragment's match temperature - this will be almost perfectly correlated with ri
[,3]	anneal	factor	either pre or post for pre or post annealing

### Author(s)

K.P. Rushton

**References**

Rushton, K.P., Analysis of the variation of glass refractive index with respect to annealing, (2009) MSc Thesis, Forensic Science, University of Auckland.

---

anscombe.df	<i>Anscombe quartet</i>
-------------	-------------------------

---

**Description**

This data is known as the Anscombe quartet. Each of sets has two variables, x and y. In each data set, x and y have the same mean (9 and 7.5), the same standard deviation (3.32 and 2.03) and the same correlation (0.82).

**Usage**

```
data(anscombe.df)
```

**References**

F. J. Anscombe. *Graphs in statistical analysis*. The American Statistician, 27(1):17-21, 1973.

---

bennett.df	<i>Bennett data</i>
------------	---------------------

---

**Description**

This data has 10 rows and 49 columns corresponding 10 refractive index (RI) measurements from 49 different locations in a windowpane.

**Usage**

```
data(bennett.df)
```

**References**

R. L. Bennett, N. D. Kim, J. M. Curran, S. A. Coulson, and A. W. N. Newton. *Spatial variation of refractive index in a pane of float glass*. Science & Justice, 43(2):71-76, April 2003.

---

`bottle.df`*Bottle data*

---

**Description**

This data contains the elemental concentration of five different elements (Manganese, Barium, Strontium, Zirconium, and Titanium) in samples of glass taken from six different Heineken beer bottles at four different locations (Base, Body, Shoulder, and Neck). Five repeat measurements are made on each sample at each location.

**Usage**

```
data(bottle.df)
```

**References**

R. L. Bennett. *Aspects of the analysis and interpretation of glass trace evidence*. Master's thesis, Department of Chemistry, University of Waikato, 2002.

---

`casework.df`*Casework RI data*

---

**Description**

This data has 10 refractive index (RI) measurement from recovered glass fragments and .....

**Usage**

```
data(casework.df)
```

**Author(s)**

James Curran

---

cck196.df	<i>CCK-196G/A data</i>
-----------	------------------------

---

**Description**

This data has genotype of Cholecystokinin (CCK) gene promoter regions of -196G/A from selected suicide victims (S) and from control subjects (C).

**Usage**

```
data(cck196.df)
```

**References**

Shindo S, Yoshioka N. *Polymorphisms of the cholecystokinin gene promoter region in suicide victims in Japan*. Forensic Science International, 150(1):85-90, May 2005.

---

cck45.df	<i>CCK-45C/T data</i>
----------	-----------------------

---

**Description**

This data has genotype of Cholecystokinin (CCK) gene promoter regions of -45C/T from selected suicide victims (S) and from control subjects (C).

**Usage**

```
data(cck45.df)
```

**References**

S. Shindo and N. Yoshioka. *Polymorphisms of the cholecystokinin gene promoter region in suicide victims in Japan*. Forensic Science International, 150(1):85-90, May 2005.

---

docexam.df

*Misleading signatures data*

---

### Description

This data has number of misleading signatures in comparisons of 16 genuine signatures and 64 simulated signatures from 15 document examiners.

### Usage

```
data(docexam.df)
```

### References

B. Found and D. K. Rogers. *Investigating forensic document examiners' skill relating to opinions on photocopied signatures*. *Science & Justice*, 45(4):199-206, 2005.

---

dpd.df

*Deoxyypyridinoline data*

---

### Description

This data measured deoxyypyridinoline (DPD) to estimate human age. DPD is a nonreducible collagen crosslink that can be measured in human dentin samples extracted from permanent individual molars. Measurements were made in dentin samples from 22 patients with ages ranging from 15 to 73.

### Usage

```
data(dpd.df)
```

### References

S. Martin-de las Heras, A. Valenzuela, and E. Villanueva. *Deoxyypyridinoline crosslinks in human dentin and estimation of age*. *International Journal of Legal Medicine*, 112(4):222-226, June 1999.



---

fiber.color.df	<i>Textile fibres in human hair data</i>
----------------	--

---

**Description**

This data comes from a population study of textile fibres in human hair in Cambridgeshire, UK, 2002. It was carried out using 26 volunteers, and 12149 fibres were recovered from a variety of hair lengths using low adhesive tape and classified according to colour, generic type and fibre length. This data has percentage distribution of fibres in head hair according to colour.

**Usage**

```
data(fiber.color.df)
```

**References**

R. Palmer and S. Oliver. *The population of coloured fibres in human head hair*. Science & Justice, 44(2):83-88, April 2004.

---

fitPvalue	<i>Return the P-value from an F-test for a linear model</i>
-----------	---

---

**Description**

This functions the P-value from the (null) hypothesis that all of the linear predictors are zero or not-significant.

**Usage**

```
fitPvalue(fit)
```

**Arguments**

`fit` a `lm` object.

**Details**

`summary` returns an invisible vector called `fstatistic` that contains the F-statistic and the degrees of freedom used to test the hypothesis that all of the linear predictors are zero or not-significant. This function takes those values and returns the appropriate upper tail probability from the F-distribution.

**Value**

A P-value

**Author(s)**

James Curran

**See Also**

[summary.lm](#)

**Examples**

```
x = runif(100,1,10)
y = 2 + 3*x + rnorm(100)
fit = lm(y~x)
fitPvalue(fit)
```

---

formatScientificTeX     *Format a number in scientific notation into LaTeX.*

---

**Description**

Format a number in standard scientific format XXX.XXEXXX into a string that can be typeset by LaTeX

**Usage**

```
formatScientificTeX(x, width, digits)
```

**Arguments**

x	The number to be formatted. This number does not need to be in scientific format.
width	width and digits are interpreted in the same way that they are for the C function sprintf. The number will be formatted so that it takes x times 10 to the power of y, where x is formatted so that the number has width characters in total (including a decimal point) and digits decimal places.
digits	See width.

**Details**

If  $x = 300.123$ ,  $width = 4$  and  $digits = 1$ , then this function will return "3.0\texttimes 10^2"

**Value**

A string that will format in a LaTeX inline math environment.

**Author(s)**

James Curran

**See Also**[sprintf](#)**Examples**

```
fmtST(300.123, 1, 4)
```

---

`gc.df`*FBI-Gc data*

---

**Description**

This data has the human group-specific component (Gc) genotypes of African Americans, Caucasians, and South Western Hispanics. The Gc locus has alleles A, B, and C and hence, the possible genotype are AA, AB, AC, BB, BC, and CC.

**Usage**

```
data(gc.df)
```

**References**

FBI reference database

---

`ghb.df`*Gamma-hydroxybutyric acid data*

---

**Description**

This data comes from a gamma-hydroxybutyric acid (GHB) experiment, which has three groups of individuals: (1) volunteers with no alcoholic history, (2) alcoholics who were waiting to start a GHB treatment program but had not yet received the drug, (3) alcoholics currently being treated with GHB, and measured the endogenous GHB concentration (ug/mL) in each.

**Usage**

```
data(ghb.df)
```

**References**

F. Mari, L. Politi, C. Trignano, M. Grazia Di Milia, M. Di Padua, and E. Bertoli. *What constitutes a normal ante-mortem urine GHB concentration?* Journal of Forensic and Legal Medicine, 16(3):148-151, April 2009.

gustafson.df

*Teeth data*

---

**Description**

This data is collected from 41 teeth, which were scored on a number of variables.

**Usage**

```
data(gustafson.df)
```

**References**

G. Gustafson. *Age determinations on teeth*. Journal of the American Dental Association, 41(1):45-54, 1950.

---

halfnorm

*Produce a half normal plot*

---

**Description**

Produce a half normal plot for a fitted `lm` or `glm` object. This function should work for any class that implements residuals.

**Usage**

```
halfnorm(fit)
```

**Arguments**

`fit`            A `lm` or `glm` object.

**Details**

The absolute value of the residuals are plotted against the positive quantiles of the normal distribution. The largest 5 percent of the empirical quantiles are labelled to help identify potential outliers.

**Value**

No values are returned.

**Note**

The labelling of the largest 5 percent is utterly arbitrary.

**Author(s)**

J.M. Curran

---

`intPlot`*Modified Interactions Plot for Two-way Analysis of Variance*

---

**Description**

This is a modified version of the function `interactionPlots` from the `s20x` library which produces greyscale plots.

Displays data with intervals for each combination of the two factors and shows the mean differences between levels of the first factor for each level of the second factor. Note that there should be more than one observation for each combination of factors.

**Usage**

```
intPlot(y, ...)
## Default S3 method:
  intPlot(y,
          fac1 = NULL,
          fac2 = NULL,
          xlab = NULL,
          xlab2 = NULL,
          ylab = NULL,
          data.order = TRUE,
          exlim = 0.1,
          jitter = 0.02,
          conf.level = 0.95,
          interval.type = "tukey",
          pooled = TRUE,
          tick.length = 0.1,
          interval.distance = 0.2,
          col.width = 2/3,
          xlab.distance = 0.1,
          xlen = 1.5,
          ylen = 1,
          ...)
## S3 method for class 'formula'
  intPlot(y,
          data,
          xlab = NULL,
          xlab2 = NULL,
          ylab = NULL,
          data.order = TRUE,
          exlim=0.1,
          jitter=0.02,
          conf.level=0.95,
          interval.type = "tukey",
          pooled = TRUE,
```

```

tick.length = 0.1,
interval.distance = 0.2,
col.width = 2/3,
xlab.distance = 0.1,
xlen=1.5,
ylen = 1,
...)
```

### Arguments

y	either a formula of the form: $y \sim \text{fac1} + \text{fac2}$ where y is the response and fac1 and fac2 are the two explanatory variables used as factors, or a single response vector
fac1	if 'y' is a vector, then fac1 contains the levels of factor 1 which correspond to the y value
fac2	if 'y' is a vector, then fac1 contains the levels of factor 2 which correspond to the y value
data	an optional data frame containing the variables in the model.
xlab	an optional label for the x-axis. If not specified the name of fac1 will be used.
xlab2	an optional label for the lines. If not specified the name of fac2 will be used.
ylab	An optional label for the y-axis. If not specified the name of y will be used.
data.order	if TRUE the levels of fac1 and fac2 will be set to unique(fac1) and unique(fac2) respectively.
exlim	provide extra limits.
jitter	the amount of horizontal jitter to show in the plot. The actual jitter is determined as the function is called, and will likely be different each time the function is used.
conf.level	confidence level of the intervals.
interval.type	four options for intervals appearing on plot: "tukey", "hsd", "lsd" or "ci".
pooled	two options: pooled or unpooled standard deviation used for plotted intervals.
tick.length	size of tick, in inches.
interval.distance	distance, as a fraction of the column width, between the points and interval. This is in addition to the extra space allocated for the jitter.
col.width	width of a factor 'column', as a fraction of the space between the centres of two columns.
xlab.distance	distance of x-axis labels from bottom of plot, as a fraction of the overall height of the plot.
xlen, ylen	xxx
...	optional arguments.

### Examples

```

library(s20x)
data(mtcars)
intPlot(wt~vs+gear, mtcars)
```

---

`jauhari1.df`*Pellet pattern data 1*

---

**Description**

This data comes from a pilot experiment, which fired a sequence of 10 shots through a set of six paper targets that were equally spaced from 3 ft to 18 ft and recorded the size of the pellet pattern at 18 ft. They then repeated the experiment with five of the targets removed and a single target at 18 ft. The third variable indicate with and without intermediate targets.

**Usage**

```
data(jauhari1.df)
```

**References**

M. Jauhari, S. M. Chatterjee, and P. K. Ghosh. *Statistical treatment of pellet dispersion data for estimating range of firing*. Journal of Forensic Sciences, 17(1):141-149, 1972.

---

`jauhari2.df`*Pellet pattern data 2*

---

**Description**

This data comes from a pilot experiment by Jauhari et al. They fired a sequence of 10 shots through a set of six paper targets that were equally spaced from 3 ft to 18 ft and recorded the size of the pellet pattern at 18 ft.

**Usage**

```
data(jauhari2.df)
```

**References**

M. Jauhari, S. M. Chatterjee, and P. K. Ghosh. *Statistical treatment of pellet dispersion data for estimating range of firing*. Journal of Forensic Sciences, 17(1):141-149, 1972.

---

`kent.df`*Chest deflection tolerance*

---

**Description**

In this data set are the results of 93 human cadaver crash tests. The tests were used in the development of thoracic injury risk functions with consideration of age and restraint condition. The data can be used with logistic regression models by recoding the variable fracture into  $<6$  and  $\geq 6$ , or with a Poisson/quasi-Poisson/negative binomial GLM.

**Usage**

```
data(kent.df)
```

**Value**

A data frame with ... variables

**Author(s)**

Kent, R, and Petrie, J.

**References**

Kent, R. and Petrie, J., Chest deflection tolerance to blunt anterior loading is sensitive to age but not load distribution , Forensic Science International 149(2004):2-3 p.121-128.

**Examples**

```
data(kent.df)

##recode the response fracture to minor injury (<6 rib fractures) and
##severe injury (>=6 rib fractures)
##kent.df = within(kent.df, {
##           injury = factor(ifelse(fractures<6,'minor','severe'),
##                             levels = c('severe','minor'))})

## fit a binomial GLM

## kent.fit = glm(injury~cmax*
```



---

kerckring.df	<i>Kerckring data</i>
--------------	-----------------------

---

**Description**

This data was analyzed the process of Kerckring on the occipital bone of numbers black and white perinates. The process of Kerckring is a projection of bone occasionally observed emerging from the inferior margin of the supraoccipital portion of the occipital squamous at the midline. Its status was regarded as either absent or present.

**Usage**

```
data(kerckring.df)
```

**References**

S. M. Weinberg, D. A. Putz, M. P. Mooney, and M. I. Siegel. *Evaluation of non-metric variation in the crania of black and white perinates*. Forensic Science International, 151(2-3):177-185, July 2005.

---

last	<i>Return the last element of a vector.</i>
------	---

---

**Description**

Return the last element of a vector

**Usage**

```
last(x)
```

**Arguments**

x                      a vector

**Author(s)**

James Curran

**Examples**

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

---

`liver.df`*Liver data*

---

**Description**

The data record information on the presence or absence of extramedullary haematopoiesis (EMH) symptoms in 51 liver of sudden infant death (SIDS) and 102 non-SIDS cases.

**Usage**

```
data(liver.df)
```

**References**

K. Toro, M.Hubay, and E. Keller. *Extramedullary haematopoiesis in liver of sudden infant death cases*. Forensic Science International, 170(1):15-19, July 2007.

---

`morphine.df`*Morphine concentration in heart and perpipheal blood samples*

---

**Description**

Data was compiled from 126 morphine-involved cases investigated by the Office of the Chief Medical Examiner, State of Maryland, USA. An investigation was conducted into whether comparison of morphine concentrations from a central and peripheral site could be used to determine whether a morphine death was acute or delayed. Fifty cases were identified as 'acute' because the urine free morphine concentration by radioimmunoassay (RIA) was less than 25 ng/mL; 76 cases were classified as 'random' because they had a urine morphine concentration greater than 25 ng/mL by RIA. The average heart blood to peripheral blood morphine concentration ratio in the acute deaths was 1.40. The average heart blood to peripheral blood morphine concentration ratio in the random deaths was 1.18. Because there was considerable overlap between the two groups of data, the authors conclude that it was not possible to predict 'acute' opiate intoxication deaths versus 'delayed' deaths when the only information available is heart and peripheral blood free morphine concentrations.

**Usage**

```
data(morphine.df)
```

**Author(s)**

Levine et al.

**References**

B. Levine, D. Green-Johnson, K.A. Moore, D. Fowler, A. Jenkins, Assessment of the acuteness of heroin deaths from the analysis of multiple blood specimens, *Science & Justice*, Volume 42, Issue 1, January 2002, Pages 17-20.

---

nasal.spline.df      *Nasal spine data*

---

**Description**

This data was analyzed the projection of the anterior nasal spine on the maxillae of numbers black and white perinates. The anterior nasal spine was assessed by its forward projection away from the frontal plane of the anterior maxillary surface. This was best assessed from the lateral and/or superior perspective. Projection was described as slight, moderate or pronounced.

**Usage**

data(nasal.spline.df)

**References**

S. M. Weinberg, D. A. Putz, M. P. Mooney, and M. I. Siegel. *Evaluation of non-metric variation in the crania of black and white perinates*. *Forensic Science International*, 151(2-3):177-185, July 2005.

---

newton.df      *Glass strata data*

---

**Description**

This data has refractive index (RI) measurements for 30 fragments on each of the five strata - float surface (FS), near float surface (NFS), bulk (B), near anti-float (NAFS), and anti-float (AFS).

**Usage**

data(newton.df)

**References**

A. W. N. Newton, J. M. Curran, C. M. Triggs, and J. S. Buckleton. *The consequences of potentially differing distributions of the refractive indices of glass fragments from control and recovered sources*. *Forensic Science International*, 140(2?3):185-193, March 2004.

---

nzglass.df	<i>New Zealand glass RI data</i>
------------	----------------------------------

---

**Description**

This data has refractive index (RI) measurements made on glass fragments recovered in New Zealand case work.

**Usage**

```
data(nzglass.df)
```

**Author(s)**

John Buckleton, Sally Coulson, Tony Gummer, Angus Newton, Gerhard Wevers, Kevan Walsh (and ESR)

---

pairsDAFS	<i>Modified Pairwise Scatter Plots with Histograms and Correlations</i>
-----------	---

---

**Description**

This is a modified version of the `pairs20x` function from the `s20x` library which produces greyscale plots.

Plots pairwise scatter plots with histograms and correlations for the data frame.

**Usage**

```
pairsDAFS(x, ...)
```

**Arguments**

x	a data frame.
...	optional arguments which are passed to the generic <code>pairs</code> function.

**Value**

Returns the plots.

**Examples**

```
##peruvian indians  
library(s20x)  
data(peru.df)  
pairsDAFS(peru.df)
```

---

palatal.df	<i>Palatal arch shape data</i>
------------	--------------------------------

---

**Description**

This data was analyzed the shape of the palatal arch on the maxillae of numbers black and white perinates. Palatal arch shape was determined by the curve formed by the inner alveolar margin. In order to determine shape, the separate maxillary halves were joined together at the midline. If the inner alveolar margins were observed to converge gradually along a continuous arc towards the midline, the palate was considered parabolic. On the other hand, if the inner alveolar margins were parallel to each other without converging until anterior of the premolars, the palate was considered hyperbolic. If the palate fell somewhere in between, it was considered intermediate.

**Usage**

```
data(palatal.df)
```

**References**

S. M. Weinberg, D. A. Putz, M. P. Mooney, and M. I. Siegel. *Evaluation of non-metric variation in the crania of black and white perinates*. Forensic Science International, 151(2-3):177-185, July 2005.

---

plotBoth	<i>Produce postscript and pdf images simulataneously</i>
----------	--

---

**Description**

Produce simulataneous postscript and pdf images using a user defined plot function.

**Usage**

```
plotBoth(plotfn, filename, control = plotBoth.control(), ...)
```

**Arguments**

plotfn	A function containing the plotting commands
filename	The name of the output file - .eps and .pdf will be appended to the postscript and pdf images respectively
control	The results of plotBoth.control controls font embedding and plotting.
...	Any additional arguments that need to be fed to plotfn

**Author(s)**

J.M. Curran

**See Also**

plotBoth.control

**Examples**

```
## Not run: plotBoth(function(){
                    plot(rnorm(100),rnorm(100))}, 'test')
## End(Not run)
```

---

plotBoth.control	<i>Control over plotBoth function</i>
------------------	---------------------------------------

---

**Description**

Gives user control over font embedding and plot generation.

**Usage**

```
plotBoth.control(genPlots = .genPlots, embedF = .embedF, embedFoptions = .embedFoptions)
```

**Arguments**

genPlots	TRUE or FALSE depending on whether plotBoth should produce a plot
embedF	TRUE or FALSE depending on whether font embedding is desired
embedFoptions	A string containing all the font embedding options to be sent to Ghostscript

**Value**

A list containing the values of the variables genPlots, embedF, and embedFoptions

**Author(s)**

J Curran

---

plotRegDiagPlots	<i>Plot some standard regression diagnostic plots</i>
------------------	---

---

**Description**

Produces 3 plots in a 2 x 2 array. The plots are a plot of residuals versus predicted values (pred-res plot), a histogram of the residuals with a normal distribution superimposed, and a normal QQ-plot of the residuals.

**Usage**

```
plotRegDiagPlots(fit)
```

**Arguments**

fit	A lm object
-----	-------------

**Author(s)**

J Curran

**See Also**

plot.lm

**Examples**

```
x = runif(100)
y = 3*x+2+dnorm(100)
fit = lm(y~x)
plotRegDiagPlots(fit)
```

---

printANODEVTable	<i>Print an analysis of deviance (ANODEV) table for a GLM</i>
------------------	---

---

**Description**

Formats the column headings an table for a GLM using the input of xtable.

**Usage**

```
printANODEVTable(xtbl, sanitize.text.function = function(x){x},
                 test = NULL, ...)
```

**Arguments**

`xtbl`            The output of `xtable` applied to the summary output of a normal GLM.  
`sanitize.text.function`  
                   Don't change this  
`test`            "Chisq" or "F" depending on the desired test  
`...`            Extra arguments to be fed to `print.xtable`

**Value**

A LaTeX formatted ANODEV table

**Author(s)**

J Curran

**See Also**

`print.xtable`

---

`printANOVATable`            *Print a ANOVA table for a Normal GLM*

---

**Description**

Formats the column headings an ANOVA table for a normal GLM using the input of `xtable`.

**Usage**

```
printANOVATable(xtbl, sanitize.text.function = function(x){x}, ...)
```

**Arguments**

`xtbl`            The output of `xtable` applied to the summary output of a normal GLM.  
`sanitize.text.function`  
                   Don't change this  
`...`            Extra arguments to be fed to `print.xtable`

**Value**

A LaTeX formatted ANOVA table

**Author(s)**

J Curran

**See Also**

`print.xtable`



---

printCI	<i>Produce a formatted confidence interval</i>
---------	--

---

**Description**

Takes a vector of length two and prints out a confidence interval in a user specified format.

**Usage**

```
printCI(x, fmt)
```

**Arguments**

x	A vector of length two containing the lower and upper bounds of a confidence interval
fmt	A format string to be used by <code>sprintf</code> . Note that this format string is repeated twice in the code for each bound of the confidence interval.

**Value**

A string containing the formatted CI

**Author(s)**

James Curran

**See Also**

`sprintf`

**Examples**

```
n = 100
x = rnorm(n)
mx = mean(x)
se = sd(x)/sqrt(n)
ci = mx + qnorm(c(0.025,0.975))*se
printCI(ci, '%5.2f')
```

---

`printRegTable`                    *Print a regression table*

---

**Description**

Formats the column headings a regression table using the input of `xtable`.

**Usage**

```
printRegTable(xtbl, sanitize.text.function = function(x){x},
              test = 't', ...)
```

**Arguments**

<code>xtbl</code>	The output of <code>xtable</code> applied to the summary output of a normal GLM.
<code>sanitize.text.function</code>	Don't change this
<code>test</code>	"t" or "z" depending on the desired test
<code>...</code>	Extra arguments to be fed to <code>print.xtable</code>

**Value**

A LaTeX formatted regression table

**Author(s)**

J Curran

**See Also**

`print.xtable`

---

`pvalue`                            *Formats an P-value figure for LaTeX*

---

**Description**

Produces a math-mode formatted string for a P-value.

**Usage**

```
pvalue(p, fmt = "$P = %6.4f$")
```

**Arguments**

p	The P-value
fmt	A formatting string for LaTeX

**Value**

A string that will be correctly formatted by LaTeX

**Author(s)**

J Curran

**See Also**

Sweave

**Examples**

```
p = 0.04
pvalue(p)
```

---

ri.calibration.df      *Combined calibration data*

---

**Description**

This data is combined from two calibration experiments which are conducted by Dr. Grzegorz Zadora and Bennett. The factor owner has two levels, RB for Rachel Bennett, and GZ for Grzegorz Zadora. The calibration of the instrument was established using standard glasses (Locke Scientific) set B1-B12 (RI=1.52912-1.520226).

**Usage**

```
data(ri.calibration.df)
```

**References**

Dr. Grzegorz Zadora from the Institute for Forensic Research in Krakow, Poland. R. L. Bennett. *Aspects of the analysis and interpretation of glass trace evidence*. Master's thesis, Department of Chemistry, University of Waikato, 2002.

---

ri.calibration2.df      *GRIM2 calibration data*

---

### Description

This data comes from a 2nd generation Glass Refractive Index Measurer (GRIM2) calibration experiment. The calibration of the instrument was established using standard glasses (Locke Scientific) set B1-B12 (RI=1.52912-1.520226). Each of the twelve reference sample glasses B1-B12 was measured five times.

### Usage

```
data(ri.calibration2.df)
```

### References

Dr. Grzegorz Zadora from the Institute for Forensic Research in Krakow, Poland.

---

Rsq      *Formats an R-squared figure for LaTeX*

---

### Description

Produces a math-mode formatted string for R-squared.

### Usage

```
Rsq(R, fmt = "$R^2 = %6.4f$", adj = FALSE)
```

### Arguments

R	The R-squared value
fmt	A formatting string for LaTeX
adj	If TRUE appends 'adjusted' to the format string

### Value

A string that will be correctly formatted by LaTeX

### Author(s)

J Curran

### See Also

Sweave

**Examples**

```
r = 0.99
Rsqr(r)
```

salting1.df

*Salting out effects in forensic blood alcohol determination***Description**

Blood alcohol measurements determined by headspace gas chromatography have been challenged on the grounds that the presence of the preservative sodium fluoride in blood samples artificially increases headspace alcohol concentrations due to a salting out effect. Blood samples containing varying amounts of ethanol and sodium fluoride (NaF) were tested using semi-automated headspace gas chromatography with n-propyl alcohol as the internal standard to assess the validity of this challenge. Miller et al found, in fact, that under these test conditions the measured alcohol levels are systematically depressed as the amount of sodium fluoride in the blood sample increases.

Blood was drawn from each of six subjects near the time of estimated peak blood alcohol concentration. Each subject had three blood samples taken to which 0, 5, and 10mg/mL of NaF were added. The blood alcohol concentration for each tube was determined twice for each tube

**Usage**

```
data(salting1.df)
```

**Format**

A data frame containing four variables

[,1]	subject	numeric factor	subject identifier 1-6
[,2]	tube	numeric factor	tube 1,2,3 for each subject
[,3]	rep	numeric factor	levels 1,2 indicating replicate measurement
[,4]	NaF	numeric factor	the level of sodium fluoride added in mg/mL
[,5]	conc	numeric	alcohol concentration in g/100mL

**Details**

Note that the blocking and treatment factors in this data frame are numeric. Therefore, to use them as such will require the use of factor or ordered.

**Author(s)**

B. A. Miller et al.

**References**

B.A. Miller, S.M. Day, T.E. Vasquez, F.M. Evans, Absence of salting out effects in forensic blood alcohol determination at various concentrations of sodium fluoride using semi-automated headspace gas chromatography, Science & Justice, Volume 44, Issue 2, April 2004, Pages 73-76.

---

 salting2.df
 

---



---

*Salting out effects in forensic blood alcohol determination*


---

### Description

Blood alcohol measurements determined by headspace gas chromatography have been challenged on the grounds that the presence of the preservative sodium fluoride in blood samples artificially increases headspace alcohol concentrations due to a salting out effect. Blood samples containing varying amounts of ethanol and sodium fluoride (NaF) were tested using semi-automated headspace gas chromatography with n-propyl alcohol as the internal standard to assess the validity of this challenge. Miller et al found, in fact, that under these test conditions the measured alcohol levels are systematically depressed as the amount of sodium fluoride in the blood sample increases.

Blood was drawn from each of four subjects at two time points, first near the time of estimated peak blood alcohol concentration and then approximately 1.5 hours later. Samples were initially analyzed with NaF at manufacturer's levels (ca. 10 mg/mL).

### Usage

```
data(salting2.df)
```

### Format

A data frame containing four variables

[,1]	subject	numeric factor	subject identifier 1-4
[,2]	time	numeric factor	time sample taken 0 or 1.5h
[,3]	NaF	numeric factor	the level of sodium fluoride added in mg/mL
[,4]	EtOH	numeric	alcohol concentration in g/100mL

### Details

Note that the blocking and treatment factors in this data frame are numeric. Therefore, to use them as such will require the use of factor or ordered.

### Author(s)

B. A. Miller et al.

### References

B.A. Miller, S.M. Day, T.E. Vasquez, F.M. Evans, Absence of salting out effects in forensic blood alcohol determination at various concentrations of sodium fluoride using semi-automated headspace gas chromatography, *Science & Justice*, Volume 44, Issue 2, April 2004, Pages 73-76.

---

sex.age.df	<i>Victims data</i>
------------	---------------------

---

**Description**

This data is collected gender and age of victims in certain crimes. The age of the victims had been classified as 0-10 and 11+.

**Usage**

```
data(sex.age.df)
```

**References**

C. G. G. Aitken, T. Connolly, A. Gammernan, G. Zhang, D. Bailey, R. Gordon, and R. Oldfield. *Statistical modelling in specific case analysis*. Science & Justice, 36(4):245-255, October 1996.

---

shadeDens	<i>Add a shaded region to a pdf plot</i>
-----------	--

---

**Description**

Useful for shading regions of interest (critical regions perhaps) on a theoretical pdf to illustrate concepts such as P-values.

**Usage**

```
shadeDens(x0, x1, dens, col = "lightgrey", n.points = 200,
          lty = 1, ...)
```

**Arguments**

x0	A starting x-value for the region to be shaded
x1	An ending x-value for the region to be shaded
dens	A function that calculates the pdf
col	A color for the shaded region
n.points	The number of points to calculate the pdf at over the interval [x0,x1]
lty	Line type
...	Additional arguments to be fed to dens such as degrees of freedom, mean or standard deviation

**Details**

Adds a filled polygon to an existing pdf plot.

**Author(s)**

J Curran

**Examples**

```
x = seq(-4.5,4.5,by = 0.01)
plot(x, dnorm(x), type = 'l')
x0 = qnorm(0.975)
x1 = 4.5
shadeDens(x0, x1, dnorm)
```

---

`shotgun.df`*Shotgun data*

---

**Description**

In order to test the validity of range-of-fire estimates obtained by the application of regression analysis to shotgun pellet patterns, a blind study was conducted in which questioned pellet patterns were fired at randomly selected ranges between 3.0 and 15.2 m (10 and 50 ft) with two different 12-gauge shotguns. each firing a different type of buckshot cartridge. Test firings at known ranges were also conducted with the same weapons and ammunition.

**Usage**`data(shotgun.df)`**Format**

A data frame with 3 variables:

[,1]	range	numeric	the range in feet of the firing
[,2]	sqrt.area	numeric	the square root of the area of the smallest rectangle that would enclose the pellet pattern
[,3]	model	factor	the model of shotgun used in the experiment

**Author(s)**

J.M. Curran

**References**

Rowe, W.F. and Hanson, S.R. (1985) Range-of-fire estimates from regression analysis applied to the spreads of shotgun pellet patterns: Results of a blind study, *Forensic Science International*, 28(3-4): 239-250.



---

 squamous.df

*Occipital squamous bone data*


---

**Description**

This data has been recorded the shape of the occipital squamous bone as narrow, equal, or greater for different races.

**Usage**

```
data(squamous.df)
```

**References**

S. M. Weinberg, D. A. Putz, M. P. Mooney, and M. I. Siegel. *Evaluation of non-metric variation in the crania of black and white perinates*. *Forensic Science International*, 151(2-3):177-185, July 2005.

---

 StangleBook

*Stangle a set of files*


---

**Description**

Stangle more than one file at once. This is a helper function to Stangle all the chapters in Data Analysis for Forensic Scientists (DAFS) at once. I have put it into the library for completeness and in the hope that someone else might find it useful.

**Usage**

```
StangleBook(idx = 0:6,
            fileList = paste(paste("Chapter", idx, "/", sep = ""),
                            paste("ch", idx, ".rnw", sep = ""),
                            sep = ""))
```

**Arguments**

idx	There seven chapters in Data Analysis for Forensic Scientists. Corresponding to each chapter is a folder named ChapterI where I is a number from 0 to 6, and an Sweave file named chI.rnw. This parameter lets the user select which subset of files need to be run through Stangle
fileList	This parameter may be altered to suit the structure of your book. By default it is set to create a list of files that have the same name and directory structure as DAFS. The function iterates over a set of file names specified by this parameter.

**Value**

No values are returned

**Author(s)**

James Curran

**See Also**

Stangle

---

SweaveBook

*Sweave a set of files*

---

**Description**

Sweave more than one file at once. This is a helper function to Sweave all the chapters in Data Analysis for Forensic Scientists (DAFS) at once. I have put it into the library for completeness and in the hope that someone else might find it useful.

**Usage**

```
SweaveBook(idx = 0:6,  
           fileList = paste(paste("Chapter", idx, "/", sep = ""),  
                           paste("ch", idx, ".rnw", sep = ""),  
                           sep = ""))
```

**Arguments**

<code>idx</code>	There seven chapters in Data Analysis for Forensic Scientists. Corresponding to each chapter is a folder named ChapterI where I is a number from 0 to 6, and an Sweave file named chI.rnw. This parameter lets the user select which subset of files need to be run through Sweave
<code>fileList</code>	This parameter may be altered to suit the structure of your book. By default it is set to create a list of files that have the same name and directory structure as DAFS. The function iterates over a set of file names specified by this parameter.

**Value**

No values are returned

**Author(s)**

James Curran

**See Also**

Sweave

---

`tryptase.df`*Tryptase data*

---

**Description**

This data has tryptase concentrations, which was measured in blood from the femoral vein in 60 deaths: 39 control cases who died rapidly (within minutes) from natural causes (sudden cardiac death and acute aortic dissection), 16 with death caused by prolonged asphyxia (traumatic compression of the chest and suffocation due to body position or smothering), and five anaphylactic deaths. In 44 of these cases, tryptase was measured in both heart (Tryp.cor) and femoral blood (Tryp.fem).

**Usage**`data(tryptase.df)`**References**

E. Edston, O. Eriksson, M. V. Hage. *Mast cell tryptase in postmortem serum - reference values and confounders*. International Journal of Legal Medicine, 121(4):275-280, May 2006

---

`velocity.df`*Calibration data*

---

**Description**

This data is about calibration between gun powder loading and speed.

**Usage**`data(velocity.df)`**References**

S. C. K. Wong. *The effects of projectile properties on glass backscatter: A statistical analysis*. Master's thesis, Department of Chemistry, University of Auckland, 2007. Forensic Science.

vhcdt.df

*Vitreous Humour Carbohydrate Deficient Transferrin***Description**

This data has the carbohydrate-deficient transferrin concentration in vitreous humour (VH-CDT) in 21 alcoholics and 7 non-alcoholics.

**Usage**

```
data(vhcdt.df)
```

**Format**

The data consist of a data frame with 28 observations on 7 variables.

age	alc	vhcdt1	vhcdt2	vhtf	td	th	[,1]	age	integer	age in years
							[,2]	alc	factor	levels (Y, N)
							[,3]	vhcdt1	double	vitreous humour carbohydrate deficient transferrin concentration (micrograms per liter)
							[,4]	vhcdt2	double	vitreous humour carbohydrate deficient transferrin concentration (micrograms per liter)
							[,5]	vhtf	double	vitreous humour transferrin (micrograms per liter)
							[,6]	td	integer	time interval to autopsy since found dead
							[,7]	td	integer	time interval to autopsy since found alive

vhcdt1 had a detection limit of 5 micrograms per liter. Observations below this level were coded as 2.5. Similarly in the second assay a detection limit of 2.5 micrograms per liter was used. Observations below this threshold were coded as 1.125.

**Author(s)**

Berkowicz, A. et al.

**References**

Berkowicz, A., Wallerstedt, S., Wall, K. and Denison, H., Analysis of carbohydrate-deficient transferrin (CDT) in vitreous humour as a forensic tool for detection of alcohol misuse, *Forensic Science International* 137:(2003) 119-124.

---

`wong.df`*Glass fragments data*

---

**Description**

This data comes from a glass fragments experiment, count the number of glass fragments on the ground that were recovered after a window pane was shot with a handgun. The projectile velocity was controlled by altering the amount of gunpowder added to each bullet. The hardness (as measured on the Rockwell scale of hardness) of each projectile was altered by changing the amount of antimony (Sb) added to the projectile lead during casting. The profile of the projectile was changed by using a round-nose (RN) or wad-cutter (WC) mold. A full factorial design was used to allocate combinations of the factors (velocity, hardness, and profile) to the experimental units (shots). There were four velocity levels, three hardness levels, and two profile levels.

**Usage**`data(wong.df)`**References**

S. C. K. Wong. *The effects of projectile properties on glass backscatter: A statistical analysis*. Master's thesis, Department of Chemistry, University of Auckland, 2007. Forensic Science.

# Index

## \* **Book**

plotBoth, [21](#)  
plotBoth.control, [22](#)  
plotRegDiagPlots, [23](#)  
printANODEVTable, [23](#)  
printANOVATable, [24](#)  
printCI, [25](#)  
printRegTable, [26](#)  
pvalue, [26](#)  
Rsq, [28](#)

## \* **datasets**

abaz.df, [3](#)  
abduct.age.df, [4](#)  
anneal.df, [4](#)  
anscombe.df, [5](#)  
bennett.df, [5](#)  
bottle.df, [6](#)  
casework.df, [6](#)  
cck196.df, [7](#)  
cck45.df, [7](#)  
docexam.df, [8](#)  
dpd.df, [8](#)  
fiber.color.df, [9](#)  
gc.df, [11](#)  
ghb.df, [11](#)  
gustafson.df, [12](#)  
jauhari1.df, [15](#)  
jauhari2.df, [15](#)  
kent.df, [16](#)  
kerckring.df, [17](#)  
liver.df, [18](#)  
morphine.df, [18](#)  
nasal.spline.df, [19](#)  
newton.df, [19](#)  
nzglass.df, [20](#)  
palatal.df, [21](#)  
ri.calibration.df, [27](#)  
ri.calibration2.df, [28](#)  
salting1.df, [29](#)

salting2.df, [30](#)  
sex.age.df, [31](#)  
shotgun.df, [32](#)  
squamous.df, [33](#)  
trypase.df, [35](#)  
velocity.df, [35](#)  
vhcdt.df, [36](#)  
wong.df, [37](#)

## \* **helper**

StangleBook, [33](#)  
SweaveBook, [34](#)

## \* **hplot**

intPlot, [13](#)  
pairsDAFS, [20](#)  
shadeDens, [31](#)

abaz.df, [3](#)  
abduct.age.df, [4](#)  
anneal.df, [4](#)  
anscombe.df, [5](#)

bennett.df, [5](#)  
bottle.df, [6](#)

casework.df, [6](#)  
cck196.df, [7](#)  
cck45.df, [7](#)

docexam.df, [8](#)  
dpd.df, [8](#)

fiber.color.df, [9](#)  
fitPvalue, [9](#)  
fmtST (formatScientificTeX), [10](#)  
formatScientificTeX, [10](#)

gc.df, [11](#)  
ghb.df, [11](#)  
gustafson.df, [12](#)

halfnorm, [12](#)

intPlot, 13

jauhari1.df, 15  
jauhari2.df, 15

kent.df, 16  
kerckring.df, 17

last, 17  
liver.df, 18

morphine.df, 18

nasal.spline.df, 19  
newton.df, 19  
nzglass.df, 20

pairsDAFS, 20  
palatal.df, 21  
plotBoth, 21  
plotBoth.control, 22  
plotRegDiagPlots, 23  
printANODEVTable, 23  
printANOVATable, 24  
printCI, 25  
printRegTable, 26  
pvalue, 26

ri.calibration.df, 27  
ri.calibration2.df, 28  
Rsq, 28

salting1.df, 29  
salting2.df, 30  
sex.age.df, 31  
shadeDens, 31  
shotgun.df, 32  
sprintf, 11  
squamous.df, 33  
StangleBook, 33  
summary.lm, 10  
SweaveBook, 34

tryptase.df, 35

velocity.df, 35  
vhcdt.df, 36

wong.df, 37