Package ‘MPV’

March 30, 2021

Title Data Sets from Montgomery, Peck and Vining
Version 1.57
Author W.J. Braun and S. MacQueen
Description Most of this package consists of data sets from the textbook Introduction to Linear Regression Analysis (3rd ed), by Montgomery, Peck and Vining. Some additional data sets and functions are also included.
Maintainer W.J. Braun <john.braun@ubc.ca>
LazyLoad true
LazyData true
Depends R (>= 2.0.1), lattice, KernSmooth
ZipData no
License Unlimited
NeedsCompilation no
Repository CRAN
Date/Publication 2021-03-30 07:20:02 UTC

R topics documented:

BCCIPlot ................................................................. 3
BCLPBias ................................................................. 4
BiasVarPlot ............................................................... 5
BioOxyDemand .......................................................... 6
bp ................................................................. 7
cement ................................................................. 8
cigbutts ................................................................. 8
earthquake ............................................................. 9
fires ................................................................. 10
GANOVA ................................................................. 10
gasdata ................................................................. 11
GFplot ................................................................. 12
R topics documented:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRegplot</td>
<td>13</td>
</tr>
<tr>
<td>Juliet</td>
<td>15</td>
</tr>
<tr>
<td>lengthguesses</td>
<td>16</td>
</tr>
<tr>
<td>lesions</td>
<td>17</td>
</tr>
<tr>
<td>LPBias</td>
<td>18</td>
</tr>
<tr>
<td>motor</td>
<td>18</td>
</tr>
<tr>
<td>noisyimage</td>
<td>19</td>
</tr>
<tr>
<td>oldwash</td>
<td>20</td>
</tr>
<tr>
<td>p11.12</td>
<td>21</td>
</tr>
<tr>
<td>p11.15</td>
<td>22</td>
</tr>
<tr>
<td>p12.11</td>
<td>22</td>
</tr>
<tr>
<td>p12.12</td>
<td>23</td>
</tr>
<tr>
<td>p12.8</td>
<td>24</td>
</tr>
<tr>
<td>p13.1</td>
<td>25</td>
</tr>
<tr>
<td>p13.16</td>
<td>25</td>
</tr>
<tr>
<td>p13.2</td>
<td>26</td>
</tr>
<tr>
<td>p13.20</td>
<td>27</td>
</tr>
<tr>
<td>p13.3</td>
<td>27</td>
</tr>
<tr>
<td>p13.4</td>
<td>28</td>
</tr>
<tr>
<td>p13.5</td>
<td>29</td>
</tr>
<tr>
<td>p13.6</td>
<td>29</td>
</tr>
<tr>
<td>p13.7</td>
<td>30</td>
</tr>
<tr>
<td>p14.1</td>
<td>31</td>
</tr>
<tr>
<td>p14.2</td>
<td>31</td>
</tr>
<tr>
<td>p15.4</td>
<td>32</td>
</tr>
<tr>
<td>p2.10</td>
<td>33</td>
</tr>
<tr>
<td>p2.12</td>
<td>33</td>
</tr>
<tr>
<td>p2.13</td>
<td>34</td>
</tr>
<tr>
<td>p2.14</td>
<td>35</td>
</tr>
<tr>
<td>p2.15</td>
<td>36</td>
</tr>
<tr>
<td>p2.16</td>
<td>37</td>
</tr>
<tr>
<td>p2.7</td>
<td>38</td>
</tr>
<tr>
<td>p2.9</td>
<td>38</td>
</tr>
<tr>
<td>p4.18</td>
<td>39</td>
</tr>
<tr>
<td>p4.19</td>
<td>40</td>
</tr>
<tr>
<td>p4.20</td>
<td>41</td>
</tr>
<tr>
<td>p5.1</td>
<td>42</td>
</tr>
<tr>
<td>p5.10</td>
<td>42</td>
</tr>
<tr>
<td>p5.11</td>
<td>43</td>
</tr>
<tr>
<td>p5.2</td>
<td>44</td>
</tr>
<tr>
<td>p5.3</td>
<td>45</td>
</tr>
<tr>
<td>p5.4</td>
<td>45</td>
</tr>
<tr>
<td>p5.5</td>
<td>46</td>
</tr>
<tr>
<td>p7.1</td>
<td>47</td>
</tr>
<tr>
<td>p7.11</td>
<td>47</td>
</tr>
<tr>
<td>p7.15</td>
<td>48</td>
</tr>
<tr>
<td>p7.16</td>
<td>49</td>
</tr>
<tr>
<td>p7.19</td>
<td>49</td>
</tr>
</tbody>
</table>
Description

Graphs of confidence interval estimates for bias and standard deviation of bias-corrected local polynomial regression curve estimates.
**Usage**

BCCIPlot(data, k1=1, k2=2, h, h2, output, g, layout, incl.biasplot, plotdata)

**Arguments**

- **data**: A data frame, whose first column must be the explanatory variable and whose second column must be the response variable.
- **k1**: degree of local polynomial used in curve estimator.
- **k2**: degree of local polynomial used in bias estimator.
- **h**: bandwidth for regression estimator.
- **h2**: bandwidth for bias estimator.
- **output**: if TRUE, numeric output is printed to the console window.
- **g**: the target function, if known (for use in simulations).
- **layout**: if TRUE, a 2x1 layout of plots is sent to the graphics device.
- **incl.biasplot**: if TRUE, the confidence intervals for the bias of the uncorrected estimate are plotted.
- **plotdata**: if TRUE, the data points are plotted as a scatter plot.

**Value**

A list containing the confidence interval limits, pointwise estimates of bias, standard deviation of bias, curve estimate, standard deviation of curve estimate, and approximate confidence limits for curve estimates. Graphs of the curve estimate confidence limits and the bias confidence limits.

**Author(s)**

W. John Braun and Wenkai Ma

---

**BCLPBias**

*Bias for Bias-Corrected Local Polynomial Regression*

**Description**

Confidence interval estimates for bias in local polynomial regression.

**Usage**

BCLPBias(xy,k1,k2,h,h2,numgrid=401,alpha=.95)
**BiasVarPlot**

**Arguments**

- `xy` A data frame, whose first column must be the explanatory variable and whose second column must be the response variable.
- `k1` degree of local polynomial used in curve estimator.
- `k2` degree of local polynomial used in bias estimator.
- `h` bandwidth for regression estimator.
- `h2` bandwidth for bias estimator.
- `numgrid` number of grid points used in the curve estimator.
- `alpha` nominal confidence level.

**Value**

A list containing the confidence interval limits, pointwise estimates of bias, standard deviation of bias, curve estimate, standard deviation of curve estimate, and approximate confidence limits for curve estimates and corresponding bias-corrected estimates.

**Author(s)**

W. John Braun and Wenkai Ma

---

**Description**

Graphs of confidence interval estimates for bias and standard deviation of in local polynomial regression curve estimates.

**Usage**

`BiasVarPlot(data, k1=1, k2=2, h, h2, output=FALSE, g, layout=TRUE)`

**Arguments**

- `data` A data frame, whose first column must be the explanatory variable and whose second column must be the response variable.
- `k1` degree of local polynomial used in curve estimator.
- `k2` degree of local polynomial used in bias estimator.
- `h` bandwidth for regression estimator.
- `h2` bandwidth for bias estimator.
- `output` if true, numeric output is printed to the console window.
- `g` the target function, if known (for use in simulations).
- `layout` if true, a 2x1 layout of plots is sent to the graphics device.
**Value**

A list containing the confidence interval limits, pointwise estimates of bias, standard deviation of bias, curve estimate, standard deviation of curve estimate, and approximate confidence limits for curve estimates. Graphs of the curve estimate confidence limits and the bias confidence limits.

**Author(s)**

W. John Braun and Wenkai Ma

---

**BioOxyDemand**

**Biochemical Oxygen Demand**

---

**Description**

The BioOxyDemand data frame has 14 rows and 2 columns.

**Usage**

data(BioOxyDemand)

**Format**

This data frame contains the following columns:

- **x** a numeric vector
- **y** a numeric vector

**Source**


**Examples**

plot(BioOxyDemand)
summary(lm(y ~ x, data = BioOxyDemand))
Blood Pressure Measurements on a Single Adult Male

Description

Systolic and diastolic blood pressure measurement readings were taken on a 56-year-old male over a 39 day period, sometimes in the mornings (AM) and sometimes in the evening (PM). Varying number of replicate measurements were taken at each time point.

Usage

bp

Format

A data frame with 121 observations on the following 4 variables.

TimeofDay factor with levels AM and PM
Date numeric
Systolic numeric
Diastolic numeric

Examples

require(lattice)
xyplot(Date ~ Diastolic|TimeofDay, groups=cut(Systolic, c(0, 130, 140, 200)), data = bp, col=c(3, 1, 2), pch=16)
matplot(bp[, c(3, 4)], type="l", lwd=2, ylab="Pressure")
n <- nrow(bp)
abline(v=(1:n)[bp[,1]=="PM"]-.5, col="grey")
abline(v=(1:n)[bp[,1]=="PM"], col="grey")
abline(v=(1:n)[bp[,1]=="PM"]+.5, col="grey")
bp.stk <- stack(bp, c("Systolic", "Diastolic"))
bp.tmp <- rbind(bp[,1:2], bp[,1:2])
bp.stk <- cbind(bp.tmp, bp.stk)
names(bp.stk) <- c("TimeofDay", "Date", "Pressure", "Type")
reps <- NULL
for (j in rle(paste(bp.stk$Date, bp.stk$TimeofDay))$lengths) reps <- c(reps, (1:j))
bp.stk$Rep <- reps
xyplot(Pressure ~ I(Date+Rep/24)|TimeofDay, groups=Type, data = bp.stk, xlab="Date", pch=16)
### cement

**Cement Data**

**Description**

The cement data frame has 13 rows and 5 columns.

**Usage**

```r
data(cement)
```

**Format**

This data frame contains the following columns:

- `y` a numeric vector
- `x1` a numeric vector
- `x2` a numeric vector
- `x3` a numeric vector
- `x4` a numeric vector

**Source**


**Examples**

```r
data(cement)
pairs(cement)
```

### cigbutts

**Cigarette Butts**

**Description**

On a university campus there are a number of areas designated for smoking. Outside of those areas, smoking is not permitted. One of the smoking areas is towards the north end of the campus near some parking lots and a large walkway towards one of the residences. Along the walkway, cigarette butts are visible in the nearby grass. Numbers of cigarette butts were counted at various distances from the smoking area in 200x80 square-cm quadrats located just west of the walkway.

**Usage**

```r
data("cigbutts")
```
**earthquake**

**Format**

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

- **distance** distance from gazebo
- **count** observed number of butts

---

**earthquake**  
*Earthquakes Data*

**Description**

The earthquake data frame contains measurements of latitude, longitude, focal depth and magnitude for all earthquakes having magnitude greater than 5.8 between 1964 and 1985.

**Usage**

`earthquake`

**Format**

This data frame contains 2178 observations on the following columns:

- **depth** numeric vector of focal depths.
- **latitude** latitudinal coordinate.
- **longitude** longitudinal coordinate.
- **magnitude** numeric vector of magnitudes.

**Source**


**Examples**

`summary(earthquake)`
Micro-fires recorded in a lab setting

Description

Rate of spread measurements (inches/s) in each direction: East, West, North and South for each of 31 experimental runs at given slopes, measured over the given time period of each (measured in seconds).

Usage

fires

Format

A data frame with 31 observations on the following 7 variables.

Run numeric
Slope numeric: vertical rise divided by horizontal run, inclined from East to West
ROS_E numeric: rate of spread measured in easterly direction
ROS_W numeric: rate of spread measured in westerly direction
ROS_S numeric: rate of spread measured in southerly direction
ROS_N numeric: rate of spread measured in northerly direction
Time numeric

Source


GANOVA

Graphical ANOVA Plot

Description

Graphical analysis of one-way ANOVA data. It allows visualization of the usual F-test.

Usage

GANOVA(dataset, var.equal=TRUE, type="QQ", center=TRUE, shift=0)
Arguments

- **dataset**: A data frame, whose first column must be the factor variable and whose second column must be the response variable.
- **var.equal**: Logical: if TRUE, within-sample variances are assumed to be equal
- **type**: "QQ" or "hist"
- **center**: if TRUE, center and scale the means to match the scale of the errors
- **shift**: on the histogram, lift the points representing the means above the horizontal axis by this amount.

Value

A QQ-plot or a histogram and rugplot

Author(s)

W. John Braun and Sarah MacQueen

Source


---

**Description**

This data frame contains the average monthly volume of natural gas used in the furnace of a 1600 square foot house located in London, Ontario, for each month from 2006 until 2011. It also contains the average temperature for each month, and a measure of degree days. Insulation was added to the roof on one occasion, the walls were insulated on a second occasion, and the mid-efficiency furnace was replaced with a high-efficiency furnace on a third occasion.

**Usage**

data("gasdata")

**Format**

A data frame with 70 observations on the following 9 variables.

- **month**: numeric 1=January, 12=December
- **degreedays**: numeric, Celsius
- **cubicmetres**: total volume of gas used in a month
- **dailyusage**: average amount of gas used per day
- **temp**: average temperature in Celsius
year numeric
I1 indicator that roof insulation is present
I2 indicator that wall insulation is present
I3 indicator that high efficiency furnace is present

GFplot  Graphical F Plot for Significance in Regression

Description
This function analyzes regression data graphically. It allows visualization of the usual F-test for significance of regression.

Usage
GFplot(X, y, plotIt=TRUE, sortTrt=FALSE, type="hist", includeIntercept=TRUE, labels=FALSE)

Arguments
X The design matrix.
y A numeric vector containing the response.
plotIt Logical: if TRUE, a graph is drawn.
sortTrt Logical: if TRUE, an attempt is made at sorting the predictor effects in descending order.
type "QQ" or "hist"
includeIntercept Logical: if TRUE, the intercept effect is plotted; otherwise, it is omitted from the plot.
labels Logical: if TRUE, names of predictor variables are used as labels; otherwise, the design matrix column numbers are used as labels

Value
A QQ-plot or a histogram and rugplot, or a list if plotIt=FALSE

Author(s)
W. John Braun

Source
**GRegplot**  

**Graphical Regression Plot**

---

**Examples**

```r
# Example 1
X <- p4.18[,-4]
y <- p4.18[,4]
GFplot(X, y, type="hist", includeIntercept=FALSE)
title("Evidence of Regression in the Jojoba Oil Data")
# Example 2
set.seed(4571)
Z <- matrix(rnorm(400), ncol=10)
A <- matrix(rnorm(81), ncol=9)
simdata <- data.frame(Z[,1], crossprod(t(Z[,-1]),A))
names(simdata) <- c("y", paste("x", 1:9, sep=""))
GFplot(simdata[,-1], simdata[,1], type="hist", includeIntercept=FALSE)
title("Evidence of Regression in Simulated Data Set")
# Example 3
GFplot(table.b1[,-1], table.b1[,1], type="hist", includeIntercept=FALSE)
title("Evidence of Regression in NFL Data Set")
# An example where stepwise AIC selects the complement
# of the set of variables that are actually in the true model:
X <- pathoeg[,-10]
y <- pathoeg[,10]
par(mfrow=c(2,2))
GFplot(X, y)
GFplot(X, y, sortTrt=TRUE)
GFplot(X, y, type="QQ")
GFplot(X, y, sortTrt=TRUE, type="QQ")
X <- table.b1[,-1]  # NFL data
y <- table.b1[,1]
GFplot(X, y)
```

---

**GRegplot**  

**Graphical Regression Plot**

---

**Description**

This function analyzes regression data graphically. It allows visualization of the usual F-test for significance of regression.

**Usage**

```r
GRegplot(X, y, sortTrt=FALSE, includeIntercept=TRUE, type="hist")
```

**Arguments**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>The design matrix.</td>
</tr>
<tr>
<td>y</td>
<td>A numeric vector containing the response.</td>
</tr>
<tr>
<td>sortTrt</td>
<td>Logical: if TRUE, an attempt is made at sorting the predictor effects in descending order.</td>
</tr>
</tbody>
</table>
includeIntercept
Logical: if TRUE, the intercept effect is plotted; otherwise, it is omitted from the plot.

type
Character: hist, for histogram; dot, for stripchart

Value
A histogram or dotplot and rugplot

Author(s)
W. John Braun

Source

Examples

# Example 1
X <- p4.18[,-4]
y <- p4.18[,4]
GRegplot(X, y, includeIntercept=FALSE)
title("Evidence of Regression in the Jojoba Oil Data")

# Example 2
set.seed(4571)
Z <- matrix(rnorm(400), ncol=10)
A <- matrix(rnorm(81), ncol=9)
simdata <- data.frame(Z[,-1], crossprod(t(Z[,-1]),A))
names(simdata) <- c("y", paste("x", 1:9, sep=""))
GRegplot(simdata[,-1], simdata[,1], includeIntercept=FALSE)
title("Evidence of Regression in Simulated Data Set")

# Example 3
GRegplot(table.b1[,-1], table.b1[,1], includeIntercept=FALSE)
title("Evidence of Regression in NFL Data Set")

# An example where stepwise AIC selects the complement
# of the set of variables that are actually in the true model:
X <- pathoeg[,-10]
y <- pathoeg[,10]
par(mfrow=c(2,1))
GRegplot(X, y)
GRegplot(X, y, sortTrt=TRUE)
X <- table.b1[,-1] # NFL data
y <- table.b1[,1]
GRegplot(X, y)
Description

Juliet has 28 rows and 9 columns. The data is of the input and output of the Spirit Still "Juliet" from Endless Summer Distillery. It is suggested to split the data by the Batch factor for ease of use.

Usage

Juliet

Format

The data frame contains the following 9 columns.

- **Batch**: a Factor determining how many times the volume has been through the still.
- **Vo11**: Volume in litres, initial
- **P1**: Percent alcohol present, initial
- **LAA1**: Litres Absolute Alcohol initial, Vol11*P1
- **Vol2**: Volume in litres, final
- **P2**: Percent alcohol present, final
- **LAA2**: Litres Absolute Alcohol final, Vol2*P2
- **Yield**: Percent yield obtained, LAA2/LAA1
- **Date**: Character, Date of run

Details

The purpose of this information is to determine the optimal initial volume and percentage. The information is broken down by Batch. A batch factor 1 means that it is the first time the liquid has gone through the spirit still. The first run through the still should have the most loss due to the "heads" and "tails". Literature states that the first run through a spirit still should yield 70 percent. A batch factor 2 means that it is the second time the liquid has gone through the spirit still. A batch factor 3 means that it is the third time or more that the liquid has gone through the spirit still. Each subsequent distillation should result in a higher yield, never to exceed 95 percent.

Source

lengthguesses

Examples

summary(Juliet)

# Split apart the Batch factor for easier use.
juliet<-split(Juliet,Juliet$Batch)
juliet1<-juliet[1]
juliet2<-juliet[2]
juliet3<-juliet[3]

plot(LAA1~LAA2,data=Juliet)
plot(LAA1~LAA2,data=juliet1)

lengthguesses  Length Guesses Data

Description

The lengthguesses list consists of 2 numeric vectors, one giving the metric-converted length guesses (in feet) of an auditorium whose actual length (in meters) was 13.1m, and the other containing the length guesses of 69 others (in meters).

Usage

data(lengthguesses)

Format

This list contains the following columns:

imperial  a numeric vector of 69 student guesses as to the length of an auditorium using the imperial system, converted to meters.

metric   a numeric vector of 44 student guesses as to the length of an auditorium using the metric system.

Source


References


Examples

with(lengthguesses, t.test(imperial, metric))
Lesions in Rat Colons

Description

Numbers of aberrant crypt foci (ACF) in each of six cross-sectional regions of the colons of 66 rats subjected to varying doses of the carcinogen azoxymethane (AOM), sacrificed at 3 different times.

Usage

lesions

Format

This data frame contains the following columns:

T Incubation time factor, levels: 6, 12 and 18 weeks
INJ Number of injections
SECT Section of colon, a factor with levels 1 through 6, where 1 denotes the proximal end of the colon and 6 denotes the distal end
RAT Label for animal within a particular T-INJ factor level combination
ACF.Total Total number of ACF lesions in a section of a rat’s colon
ACF.total.mult Sum of ACF multiplicities for a section of a rat’s colon
id Identifier for each of the 66 rats.

Source

Ranjana P. Bird, University of Northern British Columbia, Prince George, Canada.

References


Examples

summary(lesions)
ACF.All <- aggregate(ACF.Total ~ id + INJ + T, FUN=sum, data = lesions)
lesions(glm <- glm(ACF.Total ~ INJ * T, data = ACF.All, family=poisson)
summary(lesions glm)
lesions.qp <- glm(ACF.Total ~ INJ * T, data = ACF.All, family=quasipoisson)
summary(lesions.qp)
lesions.noInt <- glm(ACF.Total ~ INJ + T, data = ACF.All, family=quasipoisson)
summary(lesions.noInt)
Description

Confidence interval estimates for bias in local polynomial regression.

Usage

LPBias(xy,k1,k2,h,h2,numgrid=401,alpha=.95)

Arguments

xy A data frame, whose first column must be the explanatory variable and whose second column must be the response variable.
k1 degree of local polynomial used in curve estimator.
k2 degree of local polynomial used in bias estimator.
h bandwidth for regression estimator.
h2 bandwidth for bias estimator.
numgrid number of gridpoints used in the curve estimator.
alpha nominal confidence level.

Value

A list containing the confidence interval limits, pointwise estimates of bias, standard deviation of bias, curve estimate, standard deviation of curve estimate, and approximate confidence limits for curve estimates.

Author(s)

W. John Braun and Wenkai Ma

Description

Noise measurements for 5 samples of motors, each sample based on a different brand of bearing.

Usage

data("motor")
Format

A data frame with 5 columns.

Brand 1  A numeric vector length 6
Brand 2  A numeric vector length 6
Brand 3  A numeric vector length 6
Brand 4  A numeric vector length 6
Brand 5  A numeric vector length 6

Source


Description

The noisyimage is a list. The third component is noisy version of the third component of tarimage.

Usage

data(noisyimage)

Format

This list contains the following elements:

x  a numeric vector having 101 elements.
y  a numeric vector having 101 elements.
xy a numeric matrix having 101 rows and columns

Examples

with(noisyimage, image(x, y, xy))
Description

The oldwash dataframe has 49 rows and 8 columns. The data are from the start up of a wash still considering the amount of time it takes to heat up to a specified temperature and possible influencing factors.

Usage

data("oldwash")

Format

A data frame with 49 observations on the following 8 variables.

- Date  character, the date of the run
- startT  degrees Celsius, numeric, initial temperature
- endT  degrees Celsius, numeric, final temperature
- time  in minutes, numeric, amount of time to reach final temperature
- Vol  in litres, numeric, amount of liquid in the tank (max 2000L)
- alc  numeric, the percentage of alcohol present in the liquid
- who  character, relates to the person who ran the still
- batch  factor with levels 1 = first time through, 2 = second time through

Details

The purpose of the wash still is to increase the percentage of alcohol and strip out unwanted particulate. It can take a long time to heat up and this can lead to problems in meeting production time limits.

Source

Charisse Woods, Endless Summer Distillery (2014)

Examples

oldwash.lm<-lm(log(time)~startT+endT+Vol+alc+who+batch,data=oldwash)
summary(oldwash.lm)
par(mfrow=c(2,2))
plot(oldwash.lm)

data2<-subset(oldwash,batch==2)
hist(data2$time)
data1<-subset(oldwash,batch==1)
hist(data1$time)
oldwash.lmc<-lm(time~startT+endT+Vol+alc+who+batch,data=data1)
summary(oldwash.lmc)
plot(oldwash.lmc)

oldwash.lmd<-lm(time~startT+endT+Vol+alc+who+batch,data=data2)
summary(oldwash.lmd)
plot(oldwash.lmd)

---

Data For Problem 11.12

Description

The p11.12 data frame has 19 observations on satellite cost.

Usage

data(p11.12)

Format

This data frame contains the following columns:

- **cost** first-unit satellite cost
- **x** weight of the electronics suite

Source


References

Simpson and Montgomery (1998)

Examples

data(p11.12)
attach(p11.12)
plot(cost~x)
detach(p11.12)
Data set for Problem 11-15

Description

The p11.15 data frame has 9 rows and 2 columns.

Usage

data(p11.15)

Format

This data frame contains the following columns:

- x  a numeric vector
- y  a numeric vector

Source


References

Ryan (1997), Stefanski (1991)

Examples

data(p11.15)
plot(p11.15)
attach(p11.15)
lines(lowess(x,y))
detach(p11.15)

Data Set for Problem 12-11

Description

The p12.11 data frame has 44 observations on the fraction of active chlorine in a chemical product as a function of time after manufacturing.

Usage

data(p12.11)
Format

This data frame contains the following columns:

- **xi** time
- **yi** available chlorine

Source


Examples

```
data(p12.11)
plot(p12.11)
lines(lowess(p12.11))
```

---

**p12.12**  
Data Set for Problem 12-12

Description

The `p12.12` data frame has 18 observations on an chemical experiment. A nonlinear model relating concentration to reaction time and temperature with an additive error is proposed to fit these data.

Usage

```
data(p12.12)
```

Format

This data frame contains the following columns:

- **x1** reaction time (in minutes)
- **x2** temperature (in degrees Celsius)
- **y** concentration (in grams/liter)

Source

Examples

data(p12.12)
attach(p12.12)
# fitting the linearized model
logy.lm <- lm(I(log(y))~I(log(x1))+I(log(x2)))
summary(logy.lm)
plot(logy.lm, which=1)  # checking the residuals

# fitting the nonlinear model
y.nls <- nls(y ~ theta1*I(x1^theta2)*I(x2^theta3), start=list(theta1=.95,
theta2=.76, theta3=.21))
summary(y.nls)
plot(resid(y.nls)~fitted(y.nls))  # checking the residuals

---

Data Set for Problem 12-8

Description

The p12.8 data frame has 14 rows and 2 columns.

Usage

data(p12.8)

Format

This data frame contains the following columns:

x  a numeric vector
y  a numeric vector

Source


Examples

data(p12.8)
Data Set for Problem 13-1

Description
The p13.1 data frame has 25 observation on the test-firing results for surface-to-air missiles.

Usage
data(p13.1)

Format
This data frame contains the following columns:
- x target speed (in Knots)
- y hit (=1) or miss (=0)

Source

Examples
data(p13.1)

Data Set for Problem 13-16

Description
The p13.16 data frame has 16 rows and 5 columns.

Usage
data(p13.16)

Format
This data frame contains the following columns:
- X1 a numeric vector
- X2 a numeric vector
- X3 a numeric vector
- X4 a numeric vector
- Y a numeric vector
Source


Examples

data(p13.16)

data(p13.2)
### Data Set for Problem 13-20

**Description**

The `p13.20` data frame has 30 rows and 2 columns.

**Usage**

```r
data(p13.20)
```

**Format**

This data frame contains the following columns:

- `yhat` a numeric vector
- `resdev` a numeric vector

**Source**


**Examples**

```r
data(p13.20)
```

---

### Data Set for Problem 13-3

**Description**

The `p13.3` data frame has 10 observations on the compressive strength of an alloy fastener used in aircraft construction.

**Usage**

```r
data(p13.3)
```

**Format**

This data frame contains the following columns:

- `x` load (in psi)
- `n` sample size
- `r` number failing
Data Set for Problem 13.4

Description

The p13.4 data frame has 11 observations on the effectiveness of a price discount coupon on the purchase of a two-litre beverage.

Usage

data(p13.4)

Format

This data frame contains the following columns:

- x  discount
- n  sample size
- r  number redeemed

Source


Examples

data(p13.4)
Data Set for Problem 13-5

Description
The p13.5 data frame has 20 observations on new automobile purchases.

Usage
data(p13.5)

Format
This data frame contains the following columns:
- **x1** income
- **x2** age of oldest vehicle
- **y** new purchase less than 6 months later (1=yes, 0=no)

Source

Examples
data(p13.5)

Data Set for Problem 13-6

Description
The p13.6 data frame has 15 observations on the number of failures of a particular type of valve in a processing unit.

Usage
data(p13.6)

Format
This data frame contains the following columns:
- **valve** type of valve
- **numfail** number of failures
- **months** months
Source


Examples

data(p13.6)

data(p13.7)

---

Description

The p13.7 data frame has 44 observations on the coal mines of the Appalachian region of western Virginia.

Usage

data(p13.7)

Format

This data frame contains the following columns:

- **y**: number of fractures in upper seams of coal mines
- **x1**: inner burden thickness (in feet), shortest distance between seam floor and the lower seam
- **x2**: percent extraction of the lower previously mined seam
- **x3**: lower seam height (in feet)
- **x4**: time that the mine has been in operation (in years)

Source


References

Myers (1990)

Examples

data(p13.7)
Data Set for Problem 14-1

Description
The p14.1 data frame has 15 rows and 3 columns.

Usage
data(p14.1)

Format
This data frame contains the following columns:

- **x** a numeric vector
- **y** a numeric vector
- **time** a numeric vector

Source

Examples
data(p14.1)

Data Set for Problem 14-2

Description
The p14.2 data frame has 18 rows and 3 columns.

Usage
data(p14.2)

Format
This data frame contains the following columns:

- **t** a numeric vector
- **xt** a numeric vector
- **yt** a numeric vector
**Source**


**Examples**

```r
data(p14.2)
```

---

**Data Set for Problem 15.4**

**Description**

The `p15.4` data frame has 40 rows and 4 columns.

**Usage**

```r
data(p15.4)
```

**Format**

This data frame contains the following columns:

- `x1` a numeric vector
- `x2` a numeric vector
- `y` a numeric vector
- `set` a factor with levels `e` and `p`

**Source**


**Examples**

```r
data(p15.4)
```
Data Set for Problem 2-10

Description
The p2.10 data frame has 26 observations on weight and systolic blood pressure for randomly selected males in the 25-30 age group.

Usage
data(p2.10)

Format
This data frame contains the following columns:

weight  in pounds
sysbp  systolic blood pressure

Source

Examples
data(p2.10)
attach(p2.10)
cor.test(weight, sysbp, method="pearson")  # tests rho=0
  # and computes 95% CI for rho
  # using Fisher's Z-transform

Data Set for Problem 2-12

Description
The p2.12 data frame has 12 observations on the number of pounds of steam used per month at a plant and the average monthly ambient temperature.

Usage
data(p2.12)
Format

This data frame contains the following columns:

- **temp** ambient temperature (in degrees F)
- **usage** usage (in thousands of pounds)

Source


Examples

```r
data(p2.12)
attach(p2.12)
usage.lm <- lm(usage ~ temp)
summary(usage.lm)
predict(usage.lm, newdata=data.frame(temp=58), interval="prediction")
detach(p2.12)
```

Description

The p2.13 data frame has 16 observations on the number of days the ozone levels exceeded 0.2 ppm in the South Coast Air Basin of California for the years 1976 through 1991. It is believed that these levels are related to temperature.

Usage

```r
data(p2.13)
```

Format

This data frame contains the following columns:

- **days** number of days ozone levels exceeded 0.2 ppm
- **index** a seasonal meteorological index giving the seasonal average 850 millibar temperature.

Source


References

Examples

```r
data(p2.13)
attach(p2.13)
plot(days~index, ylim=c(-20,130))
ozone.lm <- lm(days ~ index)
summary(ozone.lm)
# plots of confidence and prediction intervals:
ozone.conf <- predict(ozone.lm, interval="confidence")
lines(sort(index), ozone.conf[order(index),2], col="red")
lines(sort(index), ozone.conf[order(index),3], col="red")
ozone.pred <- predict(ozone.lm, interval="prediction")
lines(sort(index), ozone.pred[order(index),2], col="blue")
lines(sort(index), ozone.pred[order(index),3], col="blue")
detach(p2.13)
```

---

**Data Set for Problem 2-14**

**Description**

The p2.14 data frame has 8 observations on the molar ratio of sebacic acid and the intrinsic viscosity of copolyesters. One is interested in predicting viscosity from the sebacic acid ratio.

**Usage**

```r
data(p2.14)
```

**Format**

This data frame contains the following columns:

- **ratio**: molar ratio
- **visc**: viscosity

**Source**


**References**

Examples

data(p2.14)
attach(p2.14)
plot(p2.14, pch=16, ylim=c(0,1))
visc.lm <- lm(visc ~ ratio)
summary(visc.lm)
visc.conf <- predict(visc.lm, interval="confidence")
lines(ratio, visc.conf[,2], col="red")
lines(ratio, visc.conf[,3], col="red")
visc.pred <- predict(visc.lm, interval="prediction")
lines(ratio, visc.pred[,2], col="blue")
lines(ratio, visc.pred[,3], col="blue")
detach(p2.14)

Data Set for Problem 2.15

Description

The p2.15 data frame has 8 observations on the impact of temperature on the viscosity of toluene-tetralin blends. This particular data set deals with blends with a 0.4 molar fraction of toluene.

Usage

data(p2.15)

Format

This data frame contains the following columns:

- **temp** temperature (in degrees Celsius)
- **visc** viscosity (mPa s)

Source


References

Examples

```r
data(p2.15)
attach(p2.15)
plot(visc ~ temp, pch=16)
visc.lm <- lm(visc ~ temp)
plot(visc.lm, which=1)
detach(p2.15)
```

p2.16

*Data Set for Problem 2-16*

Description

The `p2.16` data frame has 33 observations on the pressure in a tank the volume of liquid.

Usage

```r
data(p2.16)
```

Format

This data frame contains the following columns:

- **volume**: volume of liquid
- **pressure**: pressure in the tank

Source


References


Examples

```r
data(p2.16)
attach(p2.16)
plot(pressure ~ volume, pch=16)
pressure.lm <- lm(pressure ~ volume)
plot(pressure.lm, which=1)
summary(pressure.lm)
detach(p2.16)
```
Data Set for Problem 2-7

Description

The p2.7 data frame has 20 observations on the purity of oxygen produced by a fractionation process. It is thought that oxygen purity is related to the percentage of hydrocarbons in the main condensor of the processing unit.

Usage

data(p2.7)

Format

This data frame contains the following columns:

- **purity** oxygen purity (percentage)
- **hydro** hydrocarbon (percentage)

Source


Examples

data(p2.7)
attach(p2.7)
purity.lm <- lm(purity ~ hydro)
summary(purity.lm)
# confidence interval for mean purity at 1% hydrocarbon:
predict(purity.lm,newdata=data.frame(hydro = 1.00),interval="confidence")
detach(p2.7)

Data Set for Problem 2-9

Description

The p2.9 data frame has 25 rows and 2 columns. See help on softdrink for details.

Usage

data(p2.9)
Format

This data frame contains the following columns:

- **y**: a numeric vector
- **x**: a numeric vector

Source


Examples

data(p2.9)

data(p4.1)

Description

The p4.18 data frame has 13 observations on an experiment to produce a synthetic analogue to jojoba oil.

Usage

data(p4.18)

Format

This data frame contains the following columns:

- **x1**: reaction temperature
- **x2**: initial amount of catalyst
- **x3**: pressure
- **y**: yield

Source


References

**Examples**

```r
data(p4.18)
y.lm <- lm(y ~ x1 + x2 + x3, data=p4.18)
summary(y.lm)
y.lm <- lm(y ~ x1, data=p4.18)
```

---

**p4.19 Data Set for Problem 4-19**

**Description**

The p4.19 data frame has 14 observations on a designed experiment studying the relationship between abrasion index for a tire tread compound and three factors.

**Usage**

```r
data(p4.19)
```

**Format**

This data frame contains the following columns:

- **x1** hydrated silica level
- **x2** silane coupling agent level
- **x3** sulfur level
- **y** abrasion index for a tire tread compound

**Source**


**References**


**Examples**

```r
data(p4.19)
attach(p4.19)
y.lm <- lm(y ~ x1 + x2 + x3)
summary(y.lm)
plot(y.lm, which=1)
y.lm <- lm(y ~ x1)
detach(p4.19)
```
Description

The p4.20 data frame has 26 observations on a designed experiment to determine the influence of five factors on the whiteness of rayon.

Usage

data(p4.20)

Format

This data frame contains the following columns:

- `acidtemp`  acid bath temperature
- `acidconc`  cascade acid concentration
- `watertemp`  water temperature
- `sulfconc`  sulfide concentration
- `amtbl`  amount of chlorine bleach
- `y`  a measure of the whiteness of rayon

Source


References


Examples

data(p4.20)
y.lm <- lm(y ~ acidtemp, data=p4.20)
summary(y.lm)
Data Set for Problem 5-1

Description
The p5.1 data frame has 8 observations on the impact of temperature on the viscosity of toluene-tetralin blends.

Usage
```r
data(p5.1)
```

Format
This data frame contains the following columns:

- `temp` temperature
- `visc` viscosity

Source

References

Examples
```r
data(p5.1)
plot(p5.1)
```

Data Set for Problem 5-10

Description
The p5.10 data frame has 27 observations on the effect of three factors on a printing machine’s ability to apply coloring inks on package labels.

Usage
```r
data(p5.10)
```
**Format**

This data frame contains the following columns:

- `x1` speed
- `x2` pressure
- `x3` distance
- `yi1` response 1
- `yi2` response 2
- `yi3` response 3
- `ybar.i` average response
- `si` standard deviation of the 3 responses

**Source**


**Examples**

```r
data(p5.10)
attach(p5.10)
y.lm <- lm(ybar.i ~ x1 + x2 + x3)
plot(y.lm, which=1)
detach(p5.10)
```

---

**Description**

The `p5.11` data frame has 8 observations on an experiment with a catapult.

**Usage**

```r
data(p5.11)
```

**Format**

This data frame contains the following columns:

- `x1` hook
- `x2` arm length
- `x3` start angle
- `x4` stop angle
- `yi1` response 1
- `yi2` response 2
- `yi3` response 3
Source


Examples

data(p5.11)
attach(p5.11)
ybar.i <- apply(p5.11[,5:7], 1, mean)
sd.i <- apply(p5.11[,5:7], 1, sd)
y.lm <- lm(ybar.i ~ x1 + x2 + x3 + x4)
plot(y.lm, which=1)
detach(p5.11)

Data Set for Problem 5.2

Description

The p5.2 data frame has 11 observations on the vapor pressure of water for various temperatures.

Usage

data(p5.2)

Format

This data frame contains the following columns:

- temp: temperature (K)
- vapor: vapor pressure (mm Hg)

Source


Examples

data(p5.2)
plot(p5.2)
Data Set for Problem 5.3

Description
The p5.3 data frame has 12 observations on the number of bacteria surviving in a canned food product and the number of minutes of exposure to 300 degree Fahrenheit heat.

Usage
data(p5.3)

Format
This data frame contains the following columns:
- **bact** number of surviving bacteria
- **min** number of minutes of exposure

Source

Examples
data(p5.3)
plot(bact~min, data=p5.3)

Data Set for Problem 5.4

Description
The p5.4 data frame has 8 observations on 2 variables.

Usage
data(p5.4)

Format
This data frame contains the following columns:
- **x** a numeric vector
- **y** a numeric vector
Source


Examples

data(p5.4)
plot(y ~ x, data=p5.4)

Data Set for Problem 5.5

Description

The p5.5 data frame has 14 observations on the average number of defects per 10000 bottles due to stones in the bottle wall and the number of weeks since the last furnace overhaul.

Usage

data(p5.5)

Format

This data frame contains the following columns:

- **defects** a numeric vector
- **weeks** a numeric vector

Source


Examples

data(p5.5)
defects.lm <- lm(defects~weeks, data=p5.5)
plot(defects.lm, which=1)
Data Set for Problem 7-1

Description
The p7.1 data frame has 10 observations on a predictor variable.

Usage
data(p7.1)

Format
This data frame contains the following columns:

- x  a numeric vector

Source

Examples
data(p7.1)
attach(p7.1)
x2 <- x^2
detach(p7.1)

Data Set for Problem 7-11

Description
The p7.11 data frame has 11 observations on production cost versus production lot size.

Usage
data(p7.11)

Format
This data frame contains the following columns:

- x  production lot size
- y  average production cost per unit
Source


Examples

data(p7.11)
plot(y ~ x, data=p7.11)

Description

The p7.15 data frame has 6 observations on vapor pressure of water at various temperatures.

Usage

data(p7.15)

Format

This data frame contains the following columns:

- y vapor pressure (mm Hg)
- x temperature (degrees Celsius)

Source


Examples

data(p7.15)
y.lm <- lm(y ~ x, data=p7.15)
plot(y ~ x, data=p7.15)
abline(coef(y.lm))
plot(y.lm, which=1)
Data Set for Problem 7-16

Description
The p7.16 data frame has 26 observations on the observed mole fraction solubility of a solute at a constant temperature.

Usage
data(p7.16)

Format
This data frame contains the following columns:
- y  negative logarithm of the mole fraction solubility
- x1  dispersion partial solubility
- x2  dipolar partial solubility
- x3  hydrogen bonding Hansen partial solubility

Source

References

Examples
  data(p7.16)
  pairs(p7.16)

Data Set for Problem 7-19

Description
The p7.19 data frame has 10 observations on the concentration of green liquor and paper machine speed from a kraft paper machine.

Usage
data(p7.19)
Format

This data frame contains the following columns:

- y  green liquor (g/l)
- x  paper machine speed (ft/min)

Source


References


Examples

data(p7.19)
y.lm <- lm(y ~ x + I(x^2), data=p7.19)
summary(y.lm)

<table>
<thead>
<tr>
<th>p7.2</th>
<th>Data Set for Problem 7.2</th>
</tr>
</thead>
</table>

Description

The p7.2 data frame has 10 observations on solid-fuel rocket propellant weight loss.

Usage

data(p7.2)

Format

This data frame contains the following columns:

- x  months since production
- y  weight loss (kg)

Source


Examples

data(p7.2)
y.lm <- lm(y ~ x + I(x^2), data=p7.2)
summary(y.lm)
plot(y ~ x, data=p7.2)
**Data Set for Problem 7-4**

**Description**

The p7.4 data frame has 12 observations on two variables.

**Usage**

```r
data(p7.4)
```

**Format**

This data frame contains the following columns:

- **x** a numeric vector
- **y** a numeric vector

**Source**


**Examples**

```r
data(p7.4)
y.lm <- lm(y ~ x + I(x^2), data = p7.4)
summary(y.lm)
```

---

**Data Set for Problem 7-6**

**Description**

The p7.6 data frame has 12 observations on softdrink carbonation.

**Usage**

```r
data(p7.6)
```

**Format**

This data frame contains the following columns:

- **y** carbonation
- **x1** temperature
- **x2** pressure
Source

Examples

data(p7.6)
y.lm <- lm(y ~ x1 + I(x1^2) + x2 + I(x2^2) + I(x1*x2), data=p7.6)
summary(y.lm)

p8.11
Data Set for Problem 8-11

Description
The p8.11 data frame has 25 observations on the tensile strength of synthetic fibre used for men’s shirts.

Usage
data(p8.11)

Format
This data frame contains the following columns:

  y  tensile strength
  percent  percentage of cotton

Source

References
Montgomery (2001)

Examples

data(p8.11)
y.lm <- lm(y ~ percent, data=p8.11)
model.matrix(y.lm)
**Data Set for Problem 8-3**

**Description**

The p8.3 data frame has 25 observations on delivery times taken by a vending machine route driver.

**Usage**

```r
data(p8.3)
```

**Format**

This data frame contains the following columns:

- **y**: delivery time (in minutes)
- **x1**: number of cases of product stocked
- **x2**: distance walked by route driver

**Source**


**Examples**

```r
data(p8.3)
pairs(p8.3)
```

---

**Data Set for Problem 9-10**

**Description**

The p9.10 data frame has 31 observations on the rut depth of asphalt pavements prepared under different conditions.

**Usage**

```r
data(p9.10)
```
Format

This data frame contains the following columns:

- y  change in rut depth/million wheel passes (log scale)
- x1  viscosity (log scale)
- x2  percentage of asphalt in surface course
- x3  percentage of asphalt in base course
- x4  indicator
- x5  percentage of fines in surface course
- x6  percentage of voids in surface course

Source


References

Gorman and Toman (1966)

Examples

data(p9.10)
pairs(p9.10)

---

pathoeg  Pathological Example

Description

Artificial regression data which causes stepwise regression with AIC to produce a highly non-parsimonious model. The true model used to simulate the data has only one real predictor (x8).

Usage

pathoeg

Format

This data frame contains the following columns:

- x1  a numeric vector
- x2  a numeric vector
- x3  a numeric vector
- x4  a numeric vector
x5 a numeric vector
x6 a numeric vector
x7 a numeric vector
x8 a numeric vector
x9 a numeric vector
y a numeric vector

<table>
<thead>
<tr>
<th>PRESS</th>
<th>PRESS statistic</th>
</tr>
</thead>
</table>

Description

Computation of Allen’s PRESS statistic for an lm object.

Usage

PRESS(x)

Arguments

x An lm object

Value

Allen’s PRESS statistic.

Author(s)

W.J. Braun

See Also

lm

Examples

data(p4.18)
attach(p4.18)
y.lm <- lm(y ~ x1 + I(x1^2))
PRESS(y.lm)
detach(p4.18)
qqANOVA  

**QQ Plot for Analysis of Variance**

**Description**

This function is used to display the weight of the evidence against null main effects in data coming from a 1 factor design, using a QQ plot. In practice this method is often called via the function GANOVA.

**Usage**

```
qqANOVA(x, y, plot.it = TRUE, xlab = deparse(substitute(x)), ylab = deparse(substitute(y)), ...)
```

**Arguments**

- `x` numeric vector of errors
- `y` numeric vector of scaled responses
- `plot.it` logical vector indicating whether to plot or not
- `xlab` character, x-axis label
- `ylab` character, y-axis label
- `...` any other arguments for the plot function

**Value**

A QQ plot is drawn.

**Author(s)**

W. John Braun

quadline  

**Quadratic Overlay**

**Description**

Overlays a quadratic curve to a fitted quadratic model.

**Usage**

```
quadline(lm.obj, ...)
```
**Qyplot**

**Arguments**

- `lm.obj` A `lm` object (a quadratic fit)
- `...` Other arguments to the `lines` function; e.g. `col`

**Value**

The function superimposes a quadratic curve onto an existing scatterplot.

**Author(s)**

W.J. Braun

**See Also**

`lm`

**Examples**

```r
data(p4.18)
attach(p4.18)
y.lm <- lm(y ~ x1 + I(x1^2))
plot(x1, y)
quadline(y.lm)
detach(p4.18)
```

**Description**

This function analyzes regression data graphically. It allows visualization of the usual F-test for significance of regression.

**Usage**

```r
Qyplot(X, y, plotIt=TRUE, sortTrt=FALSE, type="hist", includeIntercept=TRUE, labels=FALSE)
```

**Arguments**

- `X` The design matrix.
- `y` A numeric vector containing the response.
- `plotIt` Logical: if TRUE, a graph is drawn.
- `sortTrt` Logical: if TRUE, an attempt is made at sorting the predictor effects in descending order.
- `type` "QQ" or "hist"
includeIntercept
Logical: if TRUE, the intercept effect is plotted; otherwise, it is omitted from the plot.

labels
logical: if TRUE, names of predictor variables are used as labels; otherwise, the design matrix column numbers are used as labels

Value
A QQ-plot or a histogram and rugplot, or a list if plotIt=FALSE

Author(s)
W. John Braun

Source

Examples
# Example 1
X <- p4.18[,-4]
y <- p4.18[,4]
Qyplot(X, y, type="hist", includeIntercept=FALSE)
title("Evidence of Regression in the Jojoba Oil Data")

# Example 2
set.seed(4571)
Z <- matrix(rnorm(400), ncol=10)
A <- matrix(rnorm(81), ncol=9)
simdata <- data.frame(Z[,1], crossprod(t(Z[,1]),A))
names(simdata) <- c("y", paste("x", 1:9, sep=""))
Qyplot(simdata[,-1], simdata[,1], type="hist", includeIntercept=FALSE)
title("Evidence of Regression in Simulated Data Set")

# Example 3
Qyplot(table.b1[,-1], table.b1[,1], type="hist", includeIntercept=FALSE)
title("Evidence of Regression in NFL Data Set")

# An example where stepwise AIC selects the complement of the set of variables that are actually in the true model:
X <- pathoeg[,-10]
y <- pathoeg[,10]
par(mfrow=c(2,2))
Qyplot(X, y)
Qyplot(X, y, sortTrt=TRUE)
Qyplot(X, y, type="QQ")
Qyplot(X, y, sortTrt=TRUE, type="QQ")
X <- table.b1[,-1] # NFL data
y <- table.b1[,1]
Qyplot(X, y)
seismictimings  Seismic Timing Data

Description
The seismictimings data frame has 504 rows and 3 columns. Thickness of a layer of Alberta substratum as measured by several transects of geophones.

Usage
seismictimings

Format
This data frame contains the following columns:
x longitudinal coordinate of geophone.
y latitudinal coordinate of geophone.
z time for signal to pass through substratum.

Examples
plot(y ~ x, data = seismictimings)

softdrink  Softdrink Data

Description
The softdrink data frame has 25 rows and 3 columns.

Usage
data(softdrink)

Format
This data frame contains the following columns:
y a numeric vector
x1 a numeric vector
x2 a numeric vector

Source
Examples

data(softdrink)

-----------------------------
solar Solar Data

Description

The solar data frame has 29 rows and 6 columns.

Usage

data(solar)

Format

This data frame contains the following columns:

  total.heat.flux a numeric vector
  insolation a numeric vector
  focal.pt.east a numeric vector
  focal.pt.south a numeric vector
  focal.pt.north a numeric vector
  time.of.day a numeric vector

Source


Examples

data(solar)
**Stain Removal Data**

**Description**

Data on an experiment to remove ketchup stains from white cotton fabric by soaking the stained fabric in one of five substrates for one hour. Remaining stains were scored visually and subjectively according to a 6-point scale (0 = completely clean, 5 = no change). The *stain* data frame has 15 rows and 2 columns.

**Usage**

```r
data(stain)
```

**Format**

This data frame contains the following columns:

- **treatment** a factor
- **response** a numeric vector

**Examples**

```r
data(stain)
```

---

**Table B1**

**Description**

The *table.b1* data frame has 28 observations on National Football League 1976 Team Performance.

**Usage**

```r
data(table.b1)
```

**Format**

This data frame contains the following columns:

- **y** Games won in a 14 game season
- **x1** Rushing yards
- **x2** Passing yards
- **x3** Punting average (yards/punt)
**Description**

The `table.b10` data frame has 40 observations on kinematic viscosity of a certain solvent system.

**Usage**

```r
data(table.b10)
```

**Format**

This data frame contains the following columns:

- `x1` Ratio of 2-methoxyethanol to 1,2-dimethoxyethane
- `x2` Temperature (in degrees Celsius)
- `y` Kinematic viscosity (0.000001 m²/s)

---

**Source**

Source


References

Viscosimetric Studies on 2-Methoxyethanol + 1, 2-Dimethoxyethane Binary Mixtures from -10 to 80°C. Canadian Journal of Chemical Engineering, 75, 494-501.

Examples

data(table.b10)
attach(table.b10)
y.lm <- lm(y ~ x1 + x2)
summary(y.lm)
detach(table.b10)

Table B11

Description

The table.b11 data frame has 38 observations on the quality of Pinot Noir wine.

Usage

data(table.b11)

Format

This data frame contains the following columns:

- **Clarity** a numeric vector
- **Aroma** a numeric vector
- **Body** a numeric vector
- **Flavor** a numeric vector
- **Oakiness** a numeric vector
- **Quality** a numeric vector
- **Region** a numeric vector

Source

Examples

data(table.b11)
attach(table.b11)
Quality.lm <- lm(Quality ~ Clarity + Aroma + Body + Flavor + Oakiness +
factor(Region))
summary(Quality.lm)
detach(table.b11)

Description

The table.b12 data frame has 32 rows and 6 columns.

Usage

data(table.b12)

Format

This data frame contains the following columns:

- temp a numeric vector
- soaktime a numeric vector
- soakpct a numeric vector
- difftime a numeric vector
- diffpct a numeric vector
- pitch a numeric vector

Source


Examples

data(table.b12)
Table B13

Description
The table.b13 data frame has 40 rows and 7 columns.

Usage
data(table.b13)

Format
This data frame contains the following columns:

- y  a numeric vector
- x1 a numeric vector
- x2 a numeric vector
- x3 a numeric vector
- x4 a numeric vector
- x5 a numeric vector
- x6 a numeric vector

Source

Examples
data(table.b13)

Table B14

Description
The table.b14 data frame has 25 observations on the transient points of an electronic inverter.

Usage
data(table.b14)
**Format**

This data frame contains the following columns:

- **x1** width of the NMOS Device
- **x2** length of the NMOS Device
- **x3** width of the PMOS Device
- **x4** length of the PMOS Device
- **x5** a numeric vector
- **y** transient point of PMOS-NMOS Inverters

**Source**


**Examples**

```r
data(table.b14)
y.lm <- lm(y ~ x1 + x2 + x3 + x4, data=table.b14)
plot(y.lm, which=1)
```

**Description**

The `table.b2` data frame has 29 rows and 6 columns.

**Usage**

```r
data(table.b2)
```

**Format**

This data frame contains the following columns:

- **y** a numeric vector
- **x1** a numeric vector
- **x2** a numeric vector
- **x3** a numeric vector
- **x4** a numeric vector
- **x5** a numeric vector
Source

Examples

data(table.b2)

Description
The table.b3 data frame has observations on gasoline mileage performance for 32 different automobiles.

Usage

data(table.b3)

Format
This data frame contains the following columns:
y Miles/gallon
x1 Displacement (cubic in)
x2 Horsepower (ft-lb)
x3 Torque (ft-lb)
x4 Compression ratio
x5 Rear axle ratio
x6 Carburetor (barrels)
x7 No. of transmission speeds
x8 Overall length (in)
x9 Width (in)
x10 Weight (lb)
x11 Type of transmission (1=automatic, 0=manual)

Source

References
Motor Trend, 1975
### Examples
```r
data(table.b3)
attach(table.b3)
y.lm <- lm(y ~ x1 + x6)
summary(y.lm)
# testing for the significance of the regression:
y.null <- lm(y ~ 1)
anova(y.null, y.lm)
# 95% CI for mean gas mileage:
predict(y.lm, newdata=data.frame(x1=275, x6=2), interval="confidence")
# 95% PI for gas mileage:
predict(y.lm, newdata=data.frame(x1=275, x6=2), interval="prediction")
detach(table.b3)
```

### table.b4

#### Table B4

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>The <code>table.b4</code> data frame has 24 observations on property valuation.</td>
</tr>
</tbody>
</table>

#### Usage
```r
data(table.b4)
```

#### Format
This data frame contains the following columns:

- **y** sale price of the house (in thousands of dollars)
- **x1** taxes (in thousands of dollars)
- **x2** number of baths
- **x3** lot size (in thousands of square feet)
- **x4** living space (in thousands of square feet)
- **x5** number of garage stalls
- **x6** number of rooms
- **x7** number of bedrooms
- **x8** age of the home (in years)
- **x9** number of fireplaces

#### Source

Examples

```r
data(table.b4)
attach(table.b4)
y.lm <- lm(y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9)
summary(y.lm)
detach(y.lm)
```

Data Set for Table B5

The `table.b5` data frame has 27 observations on liquefaction.

Usage

data(table.b5)

Format

This data frame contains the following columns:

- `y` CO2
- `x1` Space time (in min)
- `x2` Temperature (in degrees Celsius)
- `x3` Percent solvation
- `x4` Oil yield (g/100g MAF)
- `x5` Coal total
- `x6` Solvent total
- `x7` Hydrogen consumption

Source


References

Examples

```r
data(table.b5)
attach(table.b5)
y.lm <- lm(y ~ x6 + x7)
summary(y.lm)
detach(table.b5)
```

---

**table.b6 Data Set for Table B6**

**Description**

The `table.b6` data frame has 28 observations on a tube-flow reactor.

**Usage**

```r
data(table.b6)
```

**Format**

This data frame contains the following columns:

- **y**: NbOCl₃ concentration (g-mol/l)
- **x1**: COCl₂ concentration (g-mol/l)
- **x2**: Space time (s)
- **x3**: Molar density (g-mol/l)
- **x4**: Mole fraction CO2

**Source**


**References**


**Examples**

```r
data(table.b6)
# Partial Solution to Problem 3.9
attach(table.b6)
y.lm <- lm(y ~ x1 + x4)
summary(y.lm)
detach(table.b6)
```
Description

The `table.b7` data frame has 16 observations on oil extraction from peanuts.

Usage

data(table.b7)

Format

This data frame contains the following columns:

- x1  CO2 pressure (bar)
- x2  CO2 temperature (in degrees Celsius)
- x3  peanut moisture (percent by weight)
- x4  CO2 flow rate (L/min)
- x5  peanut particle size (mm)
- y   total oil yield

Source


References


Examples

data(table.b7)
attach(table.b7)
# partial solution to Problem 3.11:
peanuts.lm <- lm(y ~ x1 + x2 + x3 + x4 + x5)
supply(peanuts.lm)
detach(table.b7)
Description

The table.b8 data frame has 36 observations on Clathrate formation.

Usage

data(table.b8)

Format

This data frame contains the following columns:

- x1  Amount of surfactant (mass percentage)
- x2  Time (min)
- y   Clathrate formation (mass percentage)

Source


References


Examples

data(table.b8)
attach(table.b8)
clathrate.lm <- lm(y ~ x1 + x2)
summary(clathrate.lm)
detach(table.b8)
Description

The `table.b9` data frame has 62 observations on an experimental pressure drop.

Usage

data(table.b9)

Format

This data frame contains the following columns:

- **x1** Superficial fluid velocity of the gas (cm/s)
- **x2** Kinematic viscosity
- **x3** Mesh opening (cm)
- **x4** Dimensionless number relating superficial fluid velocity of the gas to the superficial fluid velocity of the liquid
- **y** Dimensionless factor for the pressure drop through a bubble cap

Source


References


Examples

data(table.b9)
attach(table.b9)
# Partial Solution to Problem 3.13:
y.lm <- lm(y ~ x1 + x2 + x3 + x4)
summary(y.lm)
detach(table.b9)
tarimage  

**Description**

The tarimage is a list. Most of the values are 0, but there are small regions of 1’s.

**Usage**

data(tarimage)

**Format**

This list contains the following elements:

- **x** a numeric vector having 101 elements.
- **y** a numeric vector having 101 elements.
- **xy** a numeric matrix having 101 rows and columns

**Examples**

with(tarimage, image(x, y, xy))

tplot  

**Graphical t Test for Regression**

**Description**

This function analyzes regression data graphically. It allows visualization of the usual t-tests for individual regression coefficients.

**Usage**

tplot(X, y, plotIt=TRUE, type="hist", includeIntercept=TRUE)

**Arguments**

- **X** The design matrix.
- **y** A numeric vector containing the response.
- **plotIt** Logical: if TRUE, a graph is drawn.
- **type** "QQ" or "hist"
- **includeIntercept** Logical: if TRUE, the intercept effect is plotted; otherwise, it is omitted from the plot.
Value

A QQ-plot or a histogram and rugplot, or a list if plotIt=FALSE

Author(s)

W. John Braun

Examples

# Jojoba oil data set
X <- p4.18[, -4]
y <- p4.18[, 4]
tplot(X, y, type="hist", includeIntercept=FALSE)
title("Tests for Individual Coefficients in the Jojoba Oil Regression")

# Simulated data set where none of the predictors are in the true model:
set.seed(4571)
Z <- matrix(rnorm(400), ncol=10)
A <- matrix(rnorm(81), ncol=9)
simdata <- data.frame(Z[, 1], crossprod(t(Z[, -1]), A))
names(simdata) <- c("y", paste("x", 1:9, sep=""))
X <- simdata[, -1]
y <- simdata[, 1]
tplot(X, y, type="hist", includeIntercept=FALSE)
title("Tests for Individual Coefficients for the Simulated Data Set")

# NFL Data set:
X <- table.b1[, -1]
y <- table.b1[, 1]
tplot(X, y, type="hist", includeIntercept=FALSE)
title("Tests for Individual Coefficients for the NFL Data Set")

# Simulated Data set where x8 is the only predictor in the true model:
X <- pathoeg[, -10]
y <- pathoeg[, 10]
par(mfrow=c(2,2))
tplot(X, y)
tplot(X, y, type="QQ")

---

tree.sample                  Sample of Loblolly Pine Data

Description

A random sample of observations taken from the 'Loblolly' data frame, one per Seed.

Usage

data("tree.sample")
Format

A data frame with 12 observations on the following 2 variables.

height  tree heights (ft)
age  tree ages (yr)

Uplot

Description

This function graphically displays the coefficient multipliers used in the Regression Plot for the given predictor.

Usage

Uplot(X.qr, Xcolumn = 1, ...)

Arguments

X.qr  The design matrix or the QR decomposition of the design matrix.
Xcolumn  The column(s) of the design matrix under study; this can be either integer valued or a character string.
...
Additional arguments to barchart.

Value

A bar plot is displayed.

Author(s)

W. John Braun

Examples

# Jojoba oil data set
X <- p4.18[, -4]
Uplot(X, 1:4)
# NFL data set; see GFplot result first
X <- table.b1[, -1]
Uplot(X, c(2, 3, 9))
# In this example, x8 is the only predictor in
# the true model:
X <- pathoeg[, -10]
y <- pathoeg[, 10]
pathoeg.F <- GFplot(X, y, plotIt=FALSE)
Uplot(X, "x8")
Uplot(X, 9)  # same as above
The `windWin80` data frame has 366 observations on midnight and noon windspeed at the Winnipeg International Airport for the year 1980.

**Usage**
```
data(windWin80)
```

**Format**
This data frame contains the following columns:
- **h0** a numeric vector containing the wind speeds at midnight.
- **h12** a numeric vector containing the windspees at the following noon.

**Examples**
```
data(windWin80)
ts.plot(windWin80$h12^2)
```

The `Wpgtemp` data frame has 7671 observations on daily maximum temperatures at the Winnipeg International Airport for the years 1960 through 1980.

**Usage**
```
data(Wpgtemp)
```

**Format**
This data frame contains the following columns:
- **temperature** A numeric vector containing the temperatures in degrees Celsius
- **day** A numeric vector denoting the observation date in numbers of days after December 31, 1959
Source
Environment Canada

Examples
summary(Wpgtemp)
Index

* datasets
  BioOxyDemand, 6
  bp, 7
  cement, 8
  cigbutts, 8
  earthquake, 9
  fires, 10
  gasdata, 11
  Juliet, 15
  lengthguesses, 16
  lesions, 17
  motor, 18
  noisyimage, 19
  oldwash, 20
  p11.12, 21
  p11.15, 22
  p12.11, 22
  p12.12, 23
  p12.8, 24
  p13.1, 25
  p13.16, 25
  p13.2, 26
  p13.20, 27
  p13.3, 27
  p13.4, 28
  p13.5, 29
  p13.6, 29
  p13.7, 30
  p14.1, 31
  p14.2, 31
  p15.4, 32
  p2.10, 33
  p2.12, 33
  p2.13, 34
  p2.14, 35
  p2.15, 36
  p2.16, 37
  p2.7, 38
  p2.9, 38
  p4.18, 39
  p4.19, 40
  p4.20, 41
  p5.1, 42
  p5.10, 42
  p5.11, 43
  p5.2, 44
  p5.3, 45
  p5.4, 45
  p5.5, 46
  p7.1, 47
  p7.11, 47
  p7.15, 48
  p7.16, 49
  p7.19, 49
  p7.2, 50
  p7.4, 51
  p7.6, 51
  p8.11, 52
  p8.3, 53
  p9.10, 53
  pathoeg, 54
  seismictimings, 59
  softdrink, 59
  solar, 60
  stain, 61
  table.b1, 61
  table.b10, 62
  table.b11, 63
  table.b12, 64
  table.b13, 65
  table.b14, 65
  table.b2, 66
  table.b3, 67
  table.b4, 68
  table.b5, 69
  table.b6, 70
  table.b7, 71
  table.b8, 72
INDEX

* graphics
  BCCIPlot, 3
  BiasVarPlot, 5
  GANOVA, 10
  GFplot, 12
  GRegplot, 13
  qqANOVA, 56
  Qyplot, 57
  tplot, 74
  Uplot, 76

* models
  BCLPBias, 4
  LPBias, 18
  PRESS, 55
  quadline, 56

  BCCIPlot, 3
  BCLPBias, 4
  BiasVarPlot, 5
  BioOxyDemand, 6
  bp, 7
  cement, 8
  cigbutts, 8
  earthquake, 9
  fires, 10

  GANOVA, 10
  gasdata, 11
  GFplot, 12
  GRegplot, 13

  Juliet, 15
  lengthguesses, 16
  lesions, 17
  LPBias, 18
  motor, 18
  noisyimage, 19
  oldwash, 20

  table.b9, 73
  tarimage, 74
  tree.sample, 75
  windWin80, 77
  Wpgtemp, 77

  p11.12, 21
  p11.15, 22
  p12.11, 22
  p12.12, 23
  p12.8, 24
  p13.1, 25
  p13.16, 25
  p13.2, 26
  p13.20, 27
  p13.3, 27
  p13.4, 28
  p13.5, 29
  p13.6, 29
  p13.7, 30
  p14.1, 31
  p14.2, 31
  p15.4, 32
  p2.10, 33
  p2.12, 33
  p2.13, 34
  p2.14, 35
  p2.15, 36
  p2.16, 37
  p2.7, 38
  p2.9, 38
  p4.18, 39
  p4.19, 40
  p4.20, 41
  p5.1, 42
  p5.10, 42
  p5.11, 43
  p5.2, 44
  p5.3, 45
  p5.4, 45
  p5.5, 46
  p7.1, 47
  p7.11, 47
  p7.15, 48
  p7.16, 49
  p7.19, 49
  p7.2, 50
  p7.4, 51
  p7.6, 51
  p8.11, 52
  p8.3, 53
  p9.10, 53
  pathoeg, 54
  PRESS, 55
INDEX

qqANOVA, 56
quadline, 56
Qyplot, 57

seismictimings, 59
softdrink, 59
solar, 60
stain, 61

table.b1, 61
table.b10, 62
table.b11, 63
table.b12, 64
table.b13, 65
table.b14, 65
table.b2, 66
table.b3, 67
table.b4, 68
table.b5, 69
table.b6, 70
table.b7, 71
table.b8, 72
table.b9, 73
tarimage, 19, 74
tplot, 74
tree.sample, 75

Uplot, 76

windWin80, 77
Wpgtemp, 77