Package ‘ACSWR’

September 5, 2015

Type Package
Title A Companion Package for the Book "A Course in Statistics with R"
Version 1.0
Date 2015-09-05
Author Prabhanjan Tattar
Maintainer Prabhanjan Tattar <prabhanjannt@gmail.com>
Imports MASS
License GPL-2
Repository CRAN
NeedsCompilation no
Date/Publication 2015-09-05 17:50:53

R topics documented:

ACSWR-package .............................................................. 3
abrasion_index ............................................................. 4
adjectives ................................................................. 5
atombomb ................................................................. 6
battery ................................................................. 5
Binom_Sim ............................................................. 7
bottling ................................................................. 8
bs ................................................................. 9
bs1 ................................................................. 10
caesareans ............................................................. 11
calcium .............................................................. 12
cardata .............................................................. 13
chdage .............................................................. 14
chemicaldata ......................................................... 15
chest .............................................................. 16
<table>
<thead>
<tr>
<th>R topics documented:</th>
</tr>
</thead>
<tbody>
<tr>
<td>cloud ..........................</td>
</tr>
<tr>
<td>cork ................................</td>
</tr>
<tr>
<td>cs ..................................</td>
</tr>
<tr>
<td>depression ........................</td>
</tr>
<tr>
<td>Disease ................................</td>
</tr>
<tr>
<td>Ehrenfest ................................</td>
</tr>
<tr>
<td>flight ................................</td>
</tr>
<tr>
<td>Geom_Sim ................................</td>
</tr>
<tr>
<td>girder ................................</td>
</tr>
<tr>
<td>hardness ................................</td>
</tr>
<tr>
<td>hearing ................................</td>
</tr>
<tr>
<td>hw .....................................</td>
</tr>
<tr>
<td>insurance ................................</td>
</tr>
<tr>
<td>intensity ................................</td>
</tr>
<tr>
<td>kurtcoeff ................................</td>
</tr>
<tr>
<td>life ...................................</td>
</tr>
<tr>
<td>lowbwt ................................</td>
</tr>
<tr>
<td>LRNormal2Mean ........................</td>
</tr>
<tr>
<td>LRNormalMean_KV ......................</td>
</tr>
<tr>
<td>LRNormalMean_UV ......................</td>
</tr>
<tr>
<td>LRNormalVariance_UM ..................</td>
</tr>
<tr>
<td>lval ..................................</td>
</tr>
<tr>
<td>memory ................................</td>
</tr>
<tr>
<td>mfp ..................................</td>
</tr>
<tr>
<td>MPbinomial ..............................</td>
</tr>
<tr>
<td>MPNormal ...............................</td>
</tr>
<tr>
<td>MPoisson ................................</td>
</tr>
<tr>
<td>msteptpm ................................</td>
</tr>
<tr>
<td>Mucociliary ................................</td>
</tr>
<tr>
<td>nerve ..................................</td>
</tr>
<tr>
<td>ns ....................................</td>
</tr>
<tr>
<td>olson ..................................</td>
</tr>
<tr>
<td>pareto_density ........................</td>
</tr>
<tr>
<td>pareto_quantile .........................</td>
</tr>
<tr>
<td>Poisson_Sim .............................</td>
</tr>
<tr>
<td>powertestplot ............................</td>
</tr>
<tr>
<td>ps ....................................</td>
</tr>
<tr>
<td>pw ....................................</td>
</tr>
<tr>
<td>QH_CI ..................................</td>
</tr>
<tr>
<td>reaction ................................</td>
</tr>
<tr>
<td>resistant_line ..........................</td>
</tr>
<tr>
<td>rocket ................................</td>
</tr>
<tr>
<td>rocket_Graeco .............................</td>
</tr>
<tr>
<td>rootstock ................................</td>
</tr>
<tr>
<td>sample ................................</td>
</tr>
<tr>
<td>sheishu ................................</td>
</tr>
<tr>
<td>shelf_stock ................................</td>
</tr>
<tr>
<td>siegel.tukey ..............................</td>
</tr>
</tbody>
</table>
Description

"A Course in Statistics with R" has been designed to meet the requirements of masters students.

Details

<table>
<thead>
<tr>
<th>Package</th>
<th>ACSWR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Package</td>
</tr>
<tr>
<td>Version</td>
<td>1.0</td>
</tr>
<tr>
<td>Date</td>
<td>2015-08-19</td>
</tr>
<tr>
<td>License</td>
<td>GPL-2</td>
</tr>
</tbody>
</table>
Author(s)
Prabhanjan Tattar
Maintainer: Prabhanjan Tattar <prabhanjannt@gmail.com>

References

Examples
```r
hist(rnorm(100))
```

---

**abrasion_index**

*Abrasion Index for the Tire Tread*

Description
To understand the relationship between the abrasion index for the tire tread, the output y, as a linear function of the hydrated silica level x1, silane coupling agent level x2 and the sulfur level x3, Derringer and Suich (1980) collected data on 14 observation points.

Usage
```r
data("abrasion_index")
```

Format
A data frame with 14 observations on the following 4 variables.

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

References

Examples
```r
data(abrasion_index)
ailm <- lm(y~x1+x2+x3, data=abrasion_index)
pairs(abrasion_index)
```
adjectives

A Dataset for Factor Analysis

Description
The data set is obtained from Rencher (2002). Here, a 12-year old girl rates 7 of her acquaintances on a differential grade of 1-9 for five adjectives kind, intelligent, happy, likable, and just.

Usage
data(adjectives)

Format
A data frame with 7 observations on the following 6 variables.

People a factor with levels FATHER FSM1a FSM2 FSM3 MSMB SISTER TEACHER
Kind a numeric vector
Intelligent a numeric vector
Happy a numeric vector
Likeable a numeric vector
Just a numeric vector

References

Examples
data(adjectives)
adjectivescor <- cor(adjectives[,1])
round(adjectivescor,3)
ad_j_eig <- eigen(adjectivescor)
cumsum(adj_eig$values)/sum(adj_eig$values)
ad_j_eig$vectors[,1:2]
loadings1 <- adj_eig$vectors[,1]*sqrt(adj_eig$values[1])
loadings2 <- adj_eig$vectors[,2]*sqrt(adj_eig$values[2])
cbind(loadings1,loadings2)
communalities <- (adj_eig$vectors[,1]*sqrt(adj_eig$values[1]))^2 +
(adj_eig$vectors[,2]*sqrt(adj_eig$values[2]))^2
round(communalities,3)
specific_variances <- 1-communalities
round(specific_variances,3)
var_acc_factors <- adj_eig$values
round(var_acc_factors,3)
prop_var <- adj_eig$values/sum(adj_eig$values)
round(prop_var,3)
cum_prop <- cumsum(adj_eig$values)/sum(adj_eig$values)
round(cum_prop,3)
Description

Gore, et al. (2006) consider the frequencies of cancer deaths of Japanese atomic bomb survivors by extent of exposure, years after exposure, etc. This data set has appeared in the journal "Statistical Sleuth".

Usage

data("atombomb")

Format

A data frame with 84 observations on the following 4 variables.

- Radians: Extent of exposure to the radian levels
- Count.Type: the type of count At Risk Death Count
- Count.Age.Group: age group with levels '0-7' '12-15' '16-19' '20-23' '24-27' '28-41' '42-57'
- Frequency: the count of deaths

References


Examples

data(atombomb)
atombombxtabs <- xtabs(Frequency~Radians+Count.Type+Count.Age.Group,data=atombomb)
atombombxtabs

Description

An experiment where the life of a battery is modeled as a function of the extreme variations in temperature of three levels 15, 70, and 1250 Fahrenheit and three type of plate material. Here, the engineer has no control on the temperature variations once the device leaves the factory. Thus, the task of the engineer is to investigate two major problems: (i) The effect of material type and temperature on the life of the device, and (ii) Finding the type of material which has least variation among the varying temperature levels. For each combination of the temperature and material, 4 replications of the life of battery are tested.
Binom_Sim

Usage

data(battery)

Format

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

Life battery life
Material the type of plate material
Temperature three extreme variations of temperature

Source


Examples

data(battery)
names(battery) <- c("L","M","T")
battery$M <- as.factor(battery$M)
battery$T <- as.factor(battery$T)
battery.aov <- aov(L~M+T,data=battery)
model.matrix(battery.aov)
summary(battery.aov)

---

Binom_Sim Simulation for Binomial Distribution

Description

A simple function to understand the algorithm to simulate psuedo-observations from binomial distribution. It is an implementation of the algorithm given in Section 11.3.1. This function is not an alternative to the rbinom function.

Usage

Binom_Sim(size, p, N)

Arguments

size Size of the binomial distribution
p Denotes the probability of success
N The number of observations required from b(n,p)

Note

This function is to simply explain the algorithm described in the text. For efficient results, the user should use the rbinom function.
A Three Factorial Experiment for Bottling Data

Description

The height of the fills in the soft drink bottle is required to be as consistent as possible and it is controlled through three factors: (i) the percent carbonation of the drink, (ii) the operating pressure in the filler, and (iii) the line speed which is the number of bottles filled per minute. The first factor variable of the percent of carbonation is available at three levels of 10, 12, and 14, the operating pressure is at 25 and 30 psi units, while the line speed are at 200 and 250 bottles per minute. Two complete replicates are available for each combination of the three factor levels, that is, 24 total number of observations. In this experiment, the deviation from the required height level is measured.

Usage

data(bottling)

Format

A data frame with 24 observations on the following 4 variables.

- Deviation : deviation from required height level
- Carbonation : the percent carbonation of the drink
- Pressure : the operating pressure in the filler
- Speed : the number of bottles filled per minute

Examples

data(bottling)
summary(bottling.aov <- aov(Deviation~.^3,bottling))
# Equivalent way
summary(aov(Deviation~ Carbonation + Pressure + Speed+ (Carbonation*Pressure)+ (Carbonation*Speed)+(Pressure*Speed)+(Carbonation*Speed*Pressure),data=bottling))
Simulated Sample from Binomial Distribution

Description

The data set is used to understand the sampling variation of the score function. The simulated data is available in Pawitan (2001).

Usage

data(bs)

Format

A data frame with 10 observations on the following 20 variables.

Sample.1  a numeric vector
Sample.2  a numeric vector
Sample.3  a numeric vector
Sample.4  a numeric vector
Sample.5  a numeric vector
Sample.6  a numeric vector
Sample.7  a numeric vector
Sample.8  a numeric vector
Sample.9  a numeric vector
Sample.10 a numeric vector
Sample.11 a numeric vector
Sample.12 a numeric vector
Sample.13 a numeric vector
Sample.14 a numeric vector
Sample.15 a numeric vector
Sample.16 a numeric vector
Sample.17 a numeric vector
Sample.18 a numeric vector
Sample.19 a numeric vector
Sample.20 a numeric vector

Source

References


Examples

data(bs)
n <- 10
sample_means <- colMeans(bs)
binomial_score_fn <- function(p,xbar)
  n*(xbar-10*p)/(p*(1-p))
p <- seq(from=0,to=1,by=.02)
plot(p,sapply(p,binomial_score_fn,xbar=sample_means[1]),"l",xlab=expression(p),
ylab=expression(S(p))
title(main="C: Score Function Plot of Binomial Model")
for(i in 2:20) lines(p,sapply(p, 
  binomial_score_fn,xbar=sample_means[i]),"l")
abline(v=4)
abline(h=0)

bs1  British Doctors Smoking and Coronary Heart Disease

Description

The problem is to investigate the impact of smoking tobacco among British doctors, refer Example 9.2.1 of Dobson. In the year 1951, a survey was sent across among all the British doctors asking them whether they smoked tobacco and their age group Age_Group. The data also collects the person-years Person_Years of the doctors in the respective age group. A follow-up after ten years reveals the number of deaths Deaths, the smoking group indicator Smoker_Cat.

Usage

data(bs1)

Format

A data frame with 10 observations on the following 9 variables.

Age_Group  a factor variable of age group with levels 35-44 45-54 55-64 65-74 75-84
Age_Cat  slightly re-coded to extract variables with Age_Cat taking values 1-5 respectively for the age groups 35-44, 45-54, 55-64, 65-74, and 75-84
Age_Square  square of the variable Age_Cat
Smoker_Cat  the smoking group indicator NO YES
Smoke_Age  a numeric vector
Smoke_Age takes the Age_Cat values for the smokers group and 0 for the non-smokers
Deaths  a follow-up after ten years revealing the number of deaths
Person_Years  the number of deaths standardized to 100000
Deaths_Per_Lakh_Years  a numeric vector
The Cesarean Cases

Description

An increasing concern has been the number of cesarean deliveries, especially in the private hospitals. Here, we know the number of births, the type of hospital (private or Government hospital), and the number of cesareans. We would like to model the number of cesareans as a function of the number of births and the type of hospital. A Poisson regression model is fitted for this data set.

Usage

data(caesareans)

Format

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Births</td>
<td>total number of births</td>
</tr>
<tr>
<td>Hospital_Type</td>
<td>type of hospital, private or government</td>
</tr>
<tr>
<td>Caesareans</td>
<td>number of cesareans</td>
</tr>
</tbody>
</table>

Source

**Examples**

```r
data(caesareans)
names(caesareans)
cae_pois <- glm(Caesareans~Hospital_Type+Births, data=caesareans, family='poisson')
summary(cae_pois)
```

---

**calcium**

*The Calcium in Soil*

**Description**

Kramer and Jensen (1969) collected data on three variables at 10 different locations. The variables of interest are available calcium in the soil, \( y_1 \), exchangeable soil calcium, \( y_2 \), and turnip green calcium, \( y_3 \). The hypothesis of interest is whether the mean vector is \([15.0 \ 6.0 \ 2.85]\).

**Usage**

```r
data(calcium)
```

**Format**

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

- **Location.Number** a numeric vector
- \( y_1 \) a numeric vector
- \( y_2 \) a numeric vector
- \( y_3 \) a numeric vector

**Source**


**References**


**Examples**

```r
data(calcium)
n <- nrow(calcium)
meanx <- colMeans(calcium[, -1])
varx <- var(calcium[, -1])
mu0 <- c(15.0, 6.0, 2.85)
t2 <- n*t(meanx-mu0)
t2
```
**cardata**

---

**Car Data**

---

**Description**

The data is used to show the effectiveness of Chernoff faces.

**Usage**

data(cardata)

**Format**

A data frame with 74 observations on the following 14 variables:

- **Model** various car models
- **P** Price
- **M** Mileage (in miles per gallon)
- **R78** Repair record 1978
- **R77** Repair record 1977
- **H** Headroom (in inches)
- **R** Rear seat clearance
- **Tr** Trunk space
- **W** Weight (in pound)
- **L** Length (in inches)
- **T** Turning diameter
- **D** Displacement (in cubic inches)
- **G** Gear ratio for high gear
- **C** Company headquarter

**Examples**

data(cardata)
pairs(cardata)
Coronary Heart Disease

Description
A well known explanation of the heart disease is that as the age increases, the risk of coronary heart disease also increase. The current data set and the example may be found in Chapter 1 of Hosmer and Lemeshow (1990-2013).

Usage
data(chdage)

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

ID patient ID
AGE age of the patient
CHD Coronary Heart Disease indicator

Source
Hosmer and Lemeshow (1990-2013).

References

Examples
data(chdage)
plot(chdage$AGE, chdage$CHD, xlab="AGE", ylab="CHD Indicator", main="Scatter plot for CHD Data")
agegrp <- cut(chdage$AGE, c(19, 29, 34, 39, 44, 49, 54, 59, 69), include.lowest=TRUE, labels=c(25, seq(31.5, 56.5, 5), 64.5))
mp <- c(25, seq(31.5, 56.5, 5), 64.5) # mid-points
chd_percent <- prop.table(table(agegrp, chdage$CHD), 1)[,2]
points(mp, chd_percent,"1", col="red")
**Description**

This data set is used to illustrate the concept of canonical correlations. Here, temperature, concentration, and time have influence on three yield variables, namely outputs, while the percentage of unchanged starting material, the percentage converted to the desired product, and the percentage of unwanted by-product form another set of related variables.

**Usage**

```r
data(chemicaldata)
```

**Format**

A data frame with 19 observations on the following 6 variables.

- y1  the percentage of unchanged starting material
- y2  the percentage converted to the desired product
- y3  the percentage of unwanted by-product
- x1  temperature
- x2  concentration
- x3  time

**Source**


**References**


**Examples**

```r
data(chemicaldata)
names(chemicaldata)
chemicaldata$x12 <- chemicaldata$x1*chemicaldata$x2;
chemicaldata$x13 <- chemicaldata$x1*chemicaldata$x3;
chemicaldata$x23 <- chemicaldata$x2*chemicaldata$x3
chemicaldata$x1sq <- chemicaldata$x1^2;
chemicaldata$x2sq <- chemicaldata$x2^2;
chemicaldata$x3sq <- chemicaldata$x3^2
S_Total <- cov(chemicaldata)
cancor_xy <- sqrt(eigen(solve(S_Total[1:3, 1:3])%*%S_Total[1:3, 4:12]%*%solve(S_Total[4:12, 1:3]))%*%S_Total[4:12, 1:3]$values)
cancor_xy
cancor(chemicaldata[,1:3],chemicaldata[,4:12])
```
### The Militiamen’s Chest Dataset

**Description**

Militia means an army composed of ordinary citizens and not of professional soldiers. This data set is available in an 1846 book published by the Belgian statistician Adolphe Quetelet, and the data is believed to have been collected some thirty years before that.

**Usage**

```r
data(chest)
```

**Format**

A data frame with 16 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>chest</th>
<th>chest width measured in inches</th>
</tr>
</thead>
<tbody>
<tr>
<td>count</td>
<td>the corresponding frequencies</td>
</tr>
</tbody>
</table>

**References**


**Examples**

```r
data(chest)
attach(chest)
names(chest)
militiamen <- rep(Chest,Count)
length(militiamen)
bins <- seq(33,48)
bins
bin.mids <- (bins[-1]+bins[-length(bins)])/2
par(mfrow=c(1,2))
h <- hist(militiamen, breaks = bins, xlab= "Chest Measurements (Inches)", main= "A: Histogram for the Militiamen")
h$counts <- sqrt(h$counts)
plot(h,xlab= "Chest Measurements (Inches)",ylab= "ROOT FREQUENCY", main= "B: Rootogram for the Militiamen")
```
The Cloud Seeding Data

Description

Chambers, et al. (1983), page 381, contains the cloud seeding data set. Rainfall in acre-feet for 52 clouds are measured, 50% of which have natural rain (control group) whereas the others are seeded. We need to visually compare whether seeding the clouds lead to increase in rainfall in acre-feet.

Usage
data(cloud)

Format

A data frame with 26 observations on the following 2 variables.

Control  Rainfall in acre-feet for 26 clouds are measured which had natural rain, that is, control group
Seeded   Rainfall in acre-feet for 26 clouds are measured which had seeded rain

References


Examples
data(cloud)
stem(log(cloud$Seeded),scale=1)
stem(log(cloud$Control),scale=1)

The Cork Dataset

Description

Thickness of cork borings in four directions of North, South, East, and West are measured for 28 trees. The problem here is to examine if the bark deposit is same in all the directions.

Usage
data(cork)
Format
A data frame with 28 observations on the following 4 variables.
North  thickness of cork boring in the North direction
East   thickness of cork boring in the East direction
South  thickness of cork boring in the South direction
West   thickness of cork boring in the West direction

References

Examples
```r
data(cork)
corkcent <- cork*0
corkcent[,1] <- cork[,1]-mean(cork[,1])
corkcent[,2] <- cork[,2]-mean(cork[,2])
corkcent[,3] <- cork[,3]-mean(cork[,3])
corkcent[,4] <- cork[,4]-mean(cork[,4])
corkcentsvd <- svd(corkcent)
t(corkcentsvd$u)%*%corkcentsvd$u
t(corkcentsvd$v)%*%corkcentsvd$v
round(corkcentsvd$u,%*% diag(corkcentsvd$d) %*% t(corkcentsvd$v),2)
round(corkcent,2)
corkcentsvd$d
```

---

Random Samples from Cauchy Distribution

Description
The data set is used to understand the sampling variation of the score function. The simulated data is available in Pawitan (2001).

Usage
data(cs)

Format
A data frame with 10 observations on the following 20 variables.
Sample.1  a numeric vector
Sample.2  a numeric vector
Sample.3  a numeric vector
Sample.4  a numeric vector
Sample.5 a numeric vector
Sample.6 a numeric vector
Sample.7 a numeric vector
Sample.8 a numeric vector
Sample.9 a numeric vector
Sample.10 a numeric vector
Sample.11 a numeric vector
Sample.12 a numeric vector
Sample.13 a numeric vector
Sample.14 a numeric vector
Sample.15 a numeric vector
Sample.16 a numeric vector
Sample.17 a numeric vector
Sample.18 a numeric vector
Sample.19 a numeric vector
Sample.20 a numeric vector

References


Examples

data(cs)
n <- 10
cauhcy_score_fn <- function(mu,x)
    sum(2*(x-mu)/(1+(x-mu)^2))
mu <- seq(from=-15,to=20,by=0.5)
plot(mu,sapply(mu,cauchy_score_fn,x=cs[,1]),"l",xlab=expression(S(mu)),
ylab=expression(S(mu)),ylim=c(-10,10))
title(main="D: Score Function Plot of Cauchy Model")
for(i in 2:20) lines(mu,sapply(mu,
    cauchy_score_fn,x=cs[,1]),"l")
abline(v=4)
abline(h=0)
**depression**  
*The Hamilton Depression Scale Factor*

**Description**

Hamilton depression scale factor IV is a measurement of mixed anxiety and depression and it is named after its inventor. In a double-blind experiment, this scale factor is obtained for 9 patients on their entry in a study, denoted by X. Post a tranquilizer T, the scale factor IV is again obtained for the same set of patients, which is denoted by Y. Here, an improvement due to tranquilizer T corresponds to a reduction in factor IV values.

**Usage**

```r
data(depression)
```

**Format**

A data frame with 9 observations on the following 3 variables.

- `Patient_No` Patient ID
- `X` measurement of depression at entry in a study
- `Y` measurement of depression post a tranquilizer

**References**


**Examples**

```r
data(depression)  
attach(depression)  
names(depression)  
wilcox.test(Y-X, alternative = "less")  
wilcox.test(Y-X, alternative = "less", exact=FALSE, correct=FALSE)
```

---

**Disease**

*Disease Outbreak Study*

**Description**

The purpose of this health study is investigation of an epidemic outbreak due to mosquitoes. A random sample from two sectors of the city among the individuals has been tested to determine if the individual had contracted the disease forming the binary outcome.
Usage

data(Disease)

Format

A data frame with 98 observations on the following 5 variables.

- x1 age
- x2 socioeconomic status of three categories between x2 and x3
- x3 socioeconomic status of three categories between x2 and x3
- x4 sector of the city
- y if the individual had contracted the disease forming the binary outcome

References


Examples

data(Disease)
DO_LR <- glm(y~., data=Disease, family='binomial')
LR_Residuals <- data.frame(Y = Disease$y, Fitted = fitted(DO_LR),
Hatvalues = hatvalues(DO_LR), Response = residuals(DO_LR,"response"), Deviance = residuals(DO_LR,"deviance"), Pearson = residuals(DO_LR,"pearson"),
Pearson_Standardized = residuals(DO_LR,"pearson")/sqrt(1-hatvalues(DO_LR)))
LR_Residuals

---

Ehrenfest

Generate transition probability matrix of Ehrenfest model

Description

The Ehrenfest model is an interesting example of a Markov chain. Though the probabilities in decimals are not as interesting as expressed in fractions, the function will help the reader generate the transition probability matrices of 2n balls among two urns.

Usage

Ehrenfest(n)

Arguments

n 2n will be the number of balls in the urns.
Details

In this experiment there are \( i \) balls in Urn I, and remaining \( 2n-i \) balls in Urn II. Then at any instance, the probability of selecting a ball from Urn I and placing it in Urn II is \( \frac{i}{2n} \), and the other way of placing a ball from Urn II to Urn I is \( \frac{2n-i}{2n} \). At each instant we let the number \( i \) of balls in the Urn I to be the state of the system. Thus, the state space is \( S = 0, 1, 2, \ldots, 2n \). Then we can pass from state \( i \) only to either of the states \( i-1 \) or \( i+1 \). Here, \( S = 0, 1, \ldots, 2n \).

Author(s)

Prabhanjan N. Tattar

Examples

Ehrenfest(2)
EhrenFest(3)

Description

Injuries in airflights, road accidents, etc, are instances of rare occurrences which are appropriately modeled by a Poisson distribution. Two models, before and after transformation, are fit and it is checked if the transformation led to a reduction to the variance.

Usage

data(flight)

Format

A data frame with 9 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Injury_Incidents</th>
<th>Count of injury incidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total_Flights</td>
<td>Total number of flights</td>
</tr>
</tbody>
</table>

References


Examples

data(flight)
names(flight)
injurylm <- lm(Injury_Incidents~Total_Flights,data=flight)
injurysqtlm <- lm(sqrt(Injury_Incidents)~Total_Flights,data=flight)
summary(injurylm)
summary(injurysqtlm)
Geom_Sim

Simulation for Geometric Distribution

Description
A simple function to understand the algorithm to simulate (pseudorandom) observations from binomial distribution. It is an implementation of the algorithm given in Section 11.3.1 "Simulation from Discrete Distributions", and as such the function is not an alternative to the "rgeom" function.

Usage
Geom_Sim(p, n)

Arguments
- p probability of success
- n number of pseudorandom observations required

Details
To simulate a random number from geometric RV, we make use of the algorithm described in the book.

Author(s)
Prabhanjan N. Tattar

See Also
rgeom

Examples
mean(Geom_Sim(0.01, 10))

girder

Strength Data Set of a Girder Experiment

Description
The shear strength of steel plate girders need to be modeled as a function of the four methods and nine girders.

Usage
data(girder)
Format

A data frame with 9 observations on the following 5 variables.

Girder  The row names, varying from S1.1 to S4.2, represent the nine type of girders, S1.1 S1.2 S2.1 S2.2 S3.1 S3.2 S4.1 S4.2 S5.1
Aarau  one of the four methods of preparation of the steel plates
Karisruhe  one of the four methods of preparation of the steel plates
Lehigh  one of the four methods of preparation of the steel plates
Cardiff  one of the four methods of preparation of the steel plates

References


Examples

data(girder)
girder
boxplot(girder[,2:5])

<table>
<thead>
<tr>
<th>hardness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardness and a Block Experiment</td>
</tr>
</tbody>
</table>

Description

Four types of tip are used which form the blocks in this experiment. The variable of interest is the hardness which further depends on the type of metal coupon. For each type of the tip, the hardness is observed for 4 different types the metal coupon.

Usage

data(hardness)

Format

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

Tip_Type  Four types of tip which form the blocks
Test_Coupon  Four different types of metal coupons
Hardness  Hardness of the coupon

References

**Examples**

```r
data(hardness)
hardness$Tip_Type <- as.factor(hardness$Tip_Type)
hardness$Test_Coupon <- as.factor(hardness$Test_Coupon)
hardness_aov <- aov(Hardness~Tip_Type+Test_Coupon, data=hardness)
summary(hardness_aov)
```

---

**Description**

A study was carried in the Eastman Kodak Company which involved the measurement of hearing loss. Such studies are called as audiometric study. This data set contains 100 males, each aged 39, who had no indication of noise exposure or hearing disorders. Here, the individual is exposed to a signal of a given frequency with an increasing intensity till the signal is perceived.

**Usage**

```r
data(hearing)
```

**Format**

A data frame with 100 observations on the following 9 variables.

- **SL_No** Serial Number
- **L500** Observation for 500Hz in the left ear
- **L1000** Observation for 1000Hz in the left ear
- **L2000** Observation for 2000Hz in the left ear
- **L4000** Observation for 4000Hz in the left ear
- **R500** Observation for 500Hz in the right ear
- **R1000** Observation for 1000Hz in the right ear
- **R2000** Observation for 2000Hz in the right ear
- **R4000** Observation for 4000Hz in the right ear

**References**


**Examples**

```r
data(hearing)
round(cor(hearing[,1]),2)
round(cov(hearing[,1]),2)
hearing.pc <- princomp(hearing[,1])
screepplot(hearing.pc,main="B: Scree Plot for Hearing Loss Data")
```
**Description**

The data set highlights the importance of handling covariance when such information is available. If the covariance is not incorporated, hypothesis testing may lead to entirely different conclusions.

**Usage**

```r
data(hw)
```

**Format**

A data frame with 20 observations on the following 2 variables.

- **Height**  the height of an individual
- **Weight** the weight of an individual

**References**


**Examples**

```r
data(hw)
sigma0 <- matrix(c(20, 100, 100, 1000), nrow=2)
sigma <- var(hw)
v <- nrow(hw)-1
p <- ncol(hw)
u <- v*(log(det(sigma0))-log(det(sigma)) + sum(diag(sigma%*%solve(sigma0)))-p)
u1 <- (1- (1/(6*v-1))*(2*p+1 - 2/(p+1)))^u
u;u1;qchisq(1-0.05,p*(p+1)/2)
```

---

**Insurance**

**Description**

Montgomery (2005), page 42, describes this data set in which the number of days taken by the company to process and settle the claims of employee health insurance customers. The data is recorded for the number of days for settlement from the first to fortieth claim.

**Usage**

```r
data(insurance)
```
*intensity*

**Format**

A data frame with 40 observations on the following 2 variables.

- Claim  Claim number
- Days   Days to settle the claim amount

**References**


**Examples**

data(insurance)
plot(insurance$Claim, insurance$Days,"l", xlab="Claim Sequence", ylab="Time to Settle the Claim")
title("B: Run Chart for Insurance Claim Settlement")

---

*Blocking for Intensity Data Set*

**Description**

The intent of this experiment is to help the engineer in improving the ability of detecting targets on a radar system. The two variables chosen which are believed to have the most impact on the detecting abilities of the radar system are marked as the amount of the background noise and the type of filter on the screen.

**Usage**

data(intensity)

**Format**

A data frame with 24 observations on the following 4 variables.

- Intensity intensity of targets
- Operator different operators who form the blocks 1 2 3 4
- Filter two types of filter 1 2
- Ground the type of background noise high low medium

**References**


**Examples**

data(intensity)
intensity.aov <- aov(Intensity~Ground*Filter+Error(Operator),intensity)
summary(intensity.aov)
intensity.aov
kurtcoeff  

**Coefficient of Kurtosis**

---

**Description**

A simple function to obtain the coefficient of kurtosis on numeric variables.

**Usage**

kurtcoeff(x)

**Arguments**

- x: the numeric vector for which the coefficient of kurtosis is required

**Details**

A straight-forward implementation of the formula is given here. A complete function "kurtosis" is available in the "e1071" package.

**Author(s)**

Prabhanjan N. Tattar

**See Also**

e1071::kurtosis

---

**life**  

**Life Expectancies**

---

**Description**

This data set consists of life expectancy in years by country, age, and sex.

**Usage**

data(life)
The Low-Birth Weight Problem

Description

Low birth weight of new-born infants is a serious concern. If the weight of the new-born is less than 2500 grams, we consider that instance as a low-birth weight case. A study was carried out at Baystate Medical Center in Springfield, Massachusetts.

Usage

data(lowbwt)

Format

A data frame with 189 observations on the following 10 variables.

LOW  Low Birth Weight
AGE  Age of Mother
LWT  Weight of Mother at Last Menstrual Period
RACE  Race 1 2 3

References


Examples

data(life)
factanal(life,factors=1)$PVAL
factanal(life,factors=2)$PVAL
factanal(life,factors=3)

Format

A data frame with 31 observations on the following 8 variables.

m0  life expectancy for males at age 0
m25 life expectancy for males at age 25
m50 life expectancy for males at age 50
m75 life expectancy for males at age 75
w0  life expectancy for females at age 0
w25 life expectancy for females at age 25
w50 life expectancy for females at age 50
w75 life expectancy for females at age 75
SMOKE  Smoking Status During Pregnancy
PTL   History of Premature Labor
HT    History of Hypertension
UI    Presence of Uterine Irritability
FTV   Number of Physician Visits During the First Trimester
BWT   Birth Weight

References

Examples

data(lowbwt)
lowglm <- glm(LOW~AGE+LWT+RACE+FTV, data=lowbwt, family='binomial')
lowglm$coefficients

________________________________________________________________
<table>
<thead>
<tr>
<th>LRNormal2Mean</th>
<th>Likelihood Ratio Test for Equality of Means when Variance Unknown</th>
</tr>
</thead>
</table>

Description
This function sets up the likelihood ratio test for equality of means when the variance term is unknown. Refer Chapter 7 for more details.

Usage
LRNormal2Mean(x, y, alpha)

Arguments
x      Observations from Population 1
y      Observations from Population 2
alpha  Size alpha test

Details
Likelihood ratio test is setup through this function. For more details, refer Chapter 7 of the book.

Author(s)
Prabhanjan N. Tattar

See Also
t.test
Examples

```r
lisa <- c(234.26, 237.18, 238.16, 259.53, 242.76, 237.81, 250.95, 277.83)
mike <- c(187.73, 206.08, 176.71, 213.69, 224.34, 235.24)
LRNormal2Mean(mike, lisa, 0.05)
```

---

**Description**

Likelihood ratio test for equality of mean when the variance is known for a sample from normal distribution is setup here. For details, refer Chapter 7 of the book.

**Usage**

```r
LRNormalMean_KV(x, mu0, alpha, sigma)
```

**Arguments**

- `x` : the variable of interest
- `mu0` : the mean of interest
- `alpha` : size of the LR test
- `sigma` : value of the known standard deviation

**Author(s)**

Prabhanjan N. Tattar

**See Also**

`t.test`

**Examples**

```r
data(hw)
LRNormalMean_KV(hw$Height, mu0=70, alpha=0.05, sigma=sqrt(20))
```
**Likelihood ratio test for mean when variance is unknown**

**Description**

Likelihood ratio test for mean when variance is unknown for a sample from normal distribution is set up here.

**Usage**

```r
LRNormalMean_UV(x, mu0, alpha)
```

**Arguments**

- **x**: the variable of interest
- **mu0**: the mean value of interest
- **alpha**: size of the LR test

**Author(s)**

Prabhanjan N. Tattar

**See Also**

LRNormalMean_KV

---

**Likelihood ratio test for the variance of normal distribution with mean is unknown**

**Description**

This function returns the LR test for the variance of normal distribution with the mean being unknown. Refer Chapter 7 for more details.

**Usage**

```r
LRNormalVariance_UM(x, sigma0, alpha)
```

**Arguments**

- **x**: the vector of sample values
- **sigma0**: the standard deviation of interest under the hypothesis
- **alpha**: the required level of significance
**lval**

**Author(s)**
Prabhanjan Tattar

**Examples**

```r
LRNormalVariance_UM(rnorm(20), 1, 0.05)
```

---

**Description**
This function is adapted from Prof. Jim Albert’s "LearnEDA" package. It returns the letter values as discussed in Chapter 4.

**Usage**

```r
lval(x, na.rm = TRUE)
```

**Arguments**

- `x` the variable of interest
- `na.rm` the default setting removes the missing values

**Author(s)**
Prabhanjan Tattar

**See Also**
LearnEDA

---

**memory**

**Memory Recall Times**

---

**Description**
A test had been conducted with the purpose of investigating if people recollect pleasant memories associated with a word earlier than some unpleasant memory related with the same word. The word is flashed on the screen and the time an individual takes to respond via keyboard is recorded for both type of the memories.

**Usage**

```r
data(memory)
```
Format

A data frame with 20 observations on the following 2 variables.

\texttt{pleasant.memory}  time to recollect pleasant memory
\texttt{unpleasant.memory}  time to recollect unpleasant memory

References

Dunn, and Master. (1982). Obtained from

Examples

data(memory)
lapply(memory, fiveunum)
lapply(memory, mad)
lapply(memory, IQR)

\begin{verbatim}
mfp

Psychological Tests for Males and Females

Description

A psychological study consisting of four tests was conducted on males and females group and the results were noted. Since the four tests are correlated and each one is noted for all the individuals, one is interested to know if the mean vector of the test scores is same across the gender group.

Usage

data(mfp)

Format

A data frame with 32 observations on the following 8 variables.

\texttt{M_y1}  pictorial inconsistencies for males
\texttt{M_y2}  paper form board test for males
\texttt{M_y3}  tool recognition test for males
\texttt{M_y4}  vocabulary test for males
\texttt{F_y1}  pictorial inconsistencies for females
\texttt{F_y2}  paper form board test for females
\texttt{F_y3}  tool recognition test for females
\texttt{F_y4}  vocabulary test for females
Examples

data(mfp)
males <- mfp[1:4]; females <- mfp[5:8]
nm <- nrow(males); nf <- nrow(females)
p <- 4; k <- 2
vm <- nm-1; vf <- nf-1
meanm <- colMeans(males); meanf <- colMeans(females)
sigmam <- var(males); sigmaf <- var(females)
sigmapl <- (1/(nm+nf-2))*(nm-1)*sigmam*(nf-1)*sigmaf
ln_M <- .5*(vm*log(det(sigmam)) + vf*log(det(sigmaf))) - .5*(vm+vf)*log(det(sigmapl))
extact_test <- -2*ln_M # the Exact Test
effect_test

MPbinomial  Most Powerful Binomial Test

Description

The function returns the level alpha MP test for the testing the hypothesis H:p=p_0 against K:p=p_1
as ensured by the application of Neyman-Pearson lemma.

Usage

MPbinomial(Hp, Kp, alpha, n)

Arguments

Hp the value of p under hypothesis H
Kp the value of p under hypothesis K
alpha size of the test
n sample size

Author(s)

Prabhanjan N. Tattar

See Also

binom.test
MPNormal  

*Most Powerful Test for Normal Distribution*

**Description**

The most powerful test for a sample from normal distribution is given here. The test is obtained by an application of the Neyman-Pearson lemma.

**Usage**

```r
MPNormal(mu0, mu1, sigma, n, alpha)
```

**Arguments**

- `mu0`: mean under hypothesis H
- `mu1`: mean under hypothesis K
- `sigma`: standard deviation
- `n`: sample size
- `alpha`: size of the test

**Author(s)**

Prabhanjan N. Tattar

**See Also**

t.test

MPPoisson  

*Most Powerful Test for Poisson Distribution*

**Description**

The most powerful test for a sample from Poisson distribution is given here. The test is obtained by an application of the Neyman-Pearson lemma.

**Usage**

```r
MPPoisson(Hlambda, Klambda, alpha, n)
```

**Arguments**

- `Hlambda`: parameter under hypothesis H
- `Klambda`: parameter under hypothesis K
- `alpha`: size of the MP test
- `n`: sample size
**msteptpm**

**Author(s)**
Prabhanjan N. Tattar

---

**m-step Transition Probability Matrix Computation**

**Description**

The m-step transition probability matrix computation is provided in this function. The equation is based on the well-known "Chapman-Kolmogorov equation".

**Usage**
msteptpm(TPM, m)

**Arguments**

TPM a transition probability matrix

m the m step required

**Author(s)**
Prabhanjan N. Tattar

**Examples**

```r
EF2 <- Ehrenfest(2)
msteptpm(as.matrix(EF2), 4)
```

---

**Mucociliary**

**Mucociliary Clearance**

**Description**

Table 6.1 of Hollander and Wolfe (1999) lists the data for Half-Time of Mucociliary Clearance. We need to test if the time across various treatments is equal or not.

**Usage**
data(Mucociliary)

**Format**

A data frame with 14 observations on the following 2 variables.

Treatment treatment levels Asbestosis Normal Subjects Obstructive Airways Disease

Time half-time of mucociliary clearance
References


Examples

data(Mucociliary)
Mucociliary$Rank <- rank(Mucociliary$Time)
aggregate(Mucociliary$Rank, by=list(Mucociliary$Treatment), sum)
kruskal.test(Time~treatment, data=Mucociliary)

nerve

The Nerve Data

Description

The Nerve data set has been popularized by Cox and Lewis (1966). In this experiment 799 waiting times are recorded for successive pulses along a nerve fiber.

Usage

data(nerve)

Format

The format is: num [1:799] 0.21 0.03 0.05 0.11 0.59 0.06 0.18 0.55 0.37 0.09 ...

Source


Examples

data(nerve)
nerve_ecdf <- ecdf(nerve)
knots(nerve_ecdf) # Returns the jump points of the edf
summary(nerve_ecdf) # the usual R summaries
nerve_ecdf(nerve) # returns the percentiles at the data points
Simulated Sample from Normal Distribution

Description

The data set is used to understand the sampling variation of the score function. The simulated data is available in Pawitan (2001).

Usage

data(ns)

Format

A data frame with 10 observations on the following 20 variables.

Sample.1 a numeric vector
Sample.2 a numeric vector
Sample.3 a numeric vector
Sample.4 a numeric vector
Sample.5 a numeric vector
Sample.6 a numeric vector
Sample.7 a numeric vector
Sample.8 a numeric vector
Sample.9 a numeric vector
Sample.10 a numeric vector
Sample.11 a numeric vector
Sample.12 a numeric vector
Sample.13 a numeric vector
Sample.14 a numeric vector
Sample.15 a numeric vector
Sample.16 a numeric vector
Sample.17 a numeric vector
Sample.18 a numeric vector
Sample.19 a numeric vector
Sample.20 a numeric vector

Source

References


Examples

```r
library(stats4)
data(ns)
x <- ns[,1]
nlogl <- function(mean,sd) { -sum(dnorm(x,mean=mean,sd=sd,log=TRUE)) }
norm_mle <- mle(nlogl,start=list(mean=median(x),sd=IQR(x)),nobs=length(x))
summary(norm_mle)
```

olson

The Olson Heart Lung Dataset

Description

We need to determine the effect of the number of revolutions per minute (rpm) of the rotary pump head of an Olson heart-lung pump on the fluid flow rate liters_minute. The rpm’s are replicated at 50, 75, 100, 125, and 150 levels with respective frequencies 5, 3, 5, 2, and 5. The fluid flow rate is measured in liters per minute.

Usage

data(olson)

Format

A data frame with 20 observations on the following 4 variables.

- **observation**: observation number
- **rpm**: rpm levels at 50, 75, 100, 125, and 150
- **level**: the rpm levels
- **liters_minute**: liters per minute

References


Examples

```r
data(olson)
par(mfrow=c(2,2))
plot(olson$rpm,olson$liters_minute,xlim=c(25,175),xlab="RPM",ylab="Flow Rate",main="Scatter Plot")
boxplot(olson$liters_minute~rpm,data=olson,main="Box Plots")
aggregate(olson$liters_minute,by=list(olson$rpm),mean)
olson_crd <- aov(liters_minute ~ as.factor(rpm), data=olson)
```
Description

A simple function is given here which returns the density function values for a Pareto RV. A more efficient implementation is obtainable in the function "dpareto" from the "VGAM" package.

Usage

pareto_density(x, scale, shape)

Arguments

x the x value
scale the scale parameter of Pareto RV
shape the shape parameter of Pareto RV

Author(s)

Prabhanjan N. Tattar

See Also

VGAM::dpareto

Examples

```r
m <- 9184
n <- 103
b <- 10000
K <- 10
theta <- seq(1000,20000,500)
plot(theta,as.numeric(sapply(theta,pareto_density,scale=b,shape=K)),"l",
     xlab=expression(theta),ylab="The Posterior Density")
(n+1)*m/n
```
**Pareto Quantile**

**Description**

A simple function is given here which returns the quantiles for a Pareto RV. A more efficient implementation is obtainable in the function "qpareto" from the "VGAM" package.

**Usage**

```r
pareto_quantile(p, scale, shape)
```

**Arguments**

- `p`: the percentiles required
- `scale`: scale of Pareto RV
- `shape`: shape of Pareto RV

**Author(s)**

Prabhanjan N. Tattar

**See Also**

VGAM::qpareto

**Examples**

```r
pareto_quantile(c(0.05, 0.95), scale=10000, shape=10)
```

---

**Poisson Sim**

**Simulation for Poisson Distribution**

**Description**

A simple function to understand the algorithm to simulate (psuedo-)observations from binomial distribution. It is an implementation of the algorithm given in Section 11.3.1 "Simulation from Discrete Distributions". This function is not an alternative to the "rpois" function.

**Usage**

```r
Poisson_Sim(lambda, n)
```
powertestplot

**Arguments**

- `lambda`: rate of the Poisson RV
- `n`: required number of pseudo-observations

**Author(s)**

Prabhanjan N. Tattar

**See Also**

rpois

**Examples**

```r
set.seed(123)
mean(Poisson_Sim(4,1000))
```

---

**powertestplot**  
* A Function to Plot the Power of a UMP Test for Normal Distribution

**Description**

A simple function for obtaining the plot of power of UMP test.

**Usage**

```r
powertestplot(mu0, sigma, n, alpha)
```

**Arguments**

- `mu0`: the value of mean  
- `sigma`: standard deviation  
- `n`: sample size  
- `alpha`: size of the test

**Author(s)**

Prabhanjan N. Tattar

**See Also**

- t.test
Examples

```
UMPNormal <- function(mu0, sigma, n, alpha) {
    qnorm(alpha) * sigma / sqrt(n) + mu0
}

UMPNormal(mu0=0, sigma=1, n=1, alpha=0.5)

corrected <- function(mu0, sigma, n, alpha) {
    mu0seq <- seq(mu0 - 3 * sigma, mu0 + 3 * sigma, (6 * sigma) / 50)
    betamu <- pnorm(sqrt(n) * (mu0seq - mu0) / sigma - qnorm(1 - alpha))
    plot(mu0seq, betamu, main = expression(h:mu > mu0 vs k:mu < mu0),
         xlab = expression(mu), ylab = expression("Power of UMP Test"),
         abline(h = alpha)
    }

corrected(mu0 = 0, sigma = 1, n = 10, alpha = 0.05)
```

Description

The data set is used to understand the sampling variation of the score function. The simulated data is available in Pawitan (2001).

Usage

```
data(ps)
```

Format

A data frame with 10 observations on the following 20 variables.

Sample.1 a numeric vector
Sample.2 a numeric vector
Sample.3 a numeric vector
Sample.4 a numeric vector
Sample.5 a numeric vector
Sample.6 a numeric vector
Sample.7 a numeric vector
Sample.8 a numeric vector
Sample.9 a numeric vector
Sample.10 a numeric vector
Sample.11 a numeric vector
Sample.12 a numeric vector
Sample.13 a numeric vector
Sample.14 a numeric vector
Sample.15 a numeric vector
Sample.16 a numeric vector
Sample.17 a numeric vector
Sample.18 a numeric vector
Sample.19 a numeric vector
Sample.20 a numeric vector

Source

References

Examples
```r
data(ps)
n <- 10
sample_means <- colMeans(ps)
poisson_score_fn <- function(theta,xbar) n*(xbar-theta)/theta
theta <- seq(from=2,to=8,by=0.2)
plot(theta,sapply(theta,poisson_score_fn,xbar=sample_means[1]),"l",xlab=expression(lambda),ylab=expression(S(lambda)),ylim=c(-5,15))
title(main="B: Score Function Plot of the Poisson Model")
for(i in 2:20)
  lines(theta,sapply(theta,poisson_score_fn,xbar=sample_means[i]),"l")
abline(v=4)
abline(h=0)
```
The Linguistic Probe Word Analysis

Description

Probe words are used to test the recall ability of words in various linguistic contexts. In this experiment the response time to five different probe words are recorded for 11 individuals. The interest in the experiment is to examine if the response times to the different words are independent or not.

Usage

data(pw)

Format

A data frame with 11 observations on the following 6 variables.

Subject.Number a numeric vector
y1 a numeric vector
y2 a numeric vector
y3 a numeric vector
y4 a numeric vector
y5 a numeric vector

References


Examples

data(pw)
sigma <- var(pw[2:6])
p <- ncol(pw)-1; v <- nrow(pw)-1
u <- p*p*(det(sigma))/(sum(diag(sigma)))*p
u1 <- -(v-(2*p^2+p+2)/(6*p))*log(u)
u;u1
## QH_CI

### Quesenberry-Hurst Simultaneous Confidence Interval

#### Description

Quesenberry and Hurst (1964) have obtained the "simultaneous confidence intervals" for the vector of success in a multinomial distribution.

#### Usage

\[
\text{QH_CI}(x, \alpha)
\]

#### Arguments

- **x**: a numeric vector
- **alpha**: as in 100 (1-\(\alpha\))

#### Author(s)

Prabhanjan N. Tattar

#### See Also

prop.test

## reaction

### Chemical Reaction Experiment

#### Description

For a chemical reaction experiment, the blocks arise due to the Batch number, Catalyst of different types form the treatments, and the reaction time is the output. Due to a restriction, all the catalysts cannot be analysed within each batch and hence we need to look at the BIBD model.

#### Usage

\[
\text{data("reaction")}
\]

#### Format

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

- **Catalyst**: different types forming the treatments
- **Batch**: batch number
- **Reaction**: reaction time
Examples

```r
data(reaction)
```

Description

"Resistant Line" is an important EDA way of fitting a regression model. The function here develops the discussion in Section 4.5.1 Resistant Line. An alternative for this function is available in "rline" function of the "LearnEDA" package.

Usage

```r
resistant_line(x, y, iterations)
```

Arguments

- `x`: the covariate or independent vector
- `y`: the dependent variate
- `iterations`: the required number of iterations

Author(s)

Prabhanjan N. Tattar

References


See Also

LearnEDA::rline
rocket

**Rocket Propellant**

**Description**

Five different formulations of a rocket propellant x1 may be used in an aircrew escape systems on the observed burning rate Y. Here, each of the formulation is prepared by mixing from a batch of raw materials x2 which can support only five formulations required for the purpose of testing.

**Usage**

```r
data(rocket)
```

**Format**

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

- `y` burning rate
- `batch` raw materials batch
- `op` experience of the operator
- `treat` formulation type of the propellant A B C D E

**References**


**Examples**

```r
data(rocket)
matrix(rocket$treat,nrow=5)
par(mfrow=c(1,3))
plot(y~factor(op)+factor(batch)+treat,rocket)
rocket_aov <- aov(y~factor(op)+factor(batch)+treat,rocket)
```

**rocket_Graeco**

**Rocket Propellant Example Extended**

**Description**

In continuation of Example 13.4.7 of the Rocket Propellant data, we now have the added blocking factor in test assemblies.

**Usage**

```r
data(rocket_Graeco)
```
**Format**

A data frame with 25 observations on the following 5 variables.

- **y** burning rate
- **batch** raw materials batch
- **op** experience of the operator
- **treat** formulation type of the propellant A B C D E
- **assembly** test assemblies a b c d e

**References**


**Examples**

```r
data(rocket_graeco)
plot(y~op+batch+treat+assembly,rocket_graeco)
rocket.glsd.aov <- aov(y~factor(op)+factor(batch)+treat+assembly,rocket_graeco)
summary(rocket.glsd.aov)
```

---

**rootstock**

*Apple of Different Rootstock*

**Description**

The goal is to test if the mean vector of the four variables is same across 6 stratas of the experiment.

**Usage**

```r
data(rootstock)
```

**Format**

A data frame with 48 observations on the following 5 variables.

- **rootstock** Six different rootstocks
- **y1** trunk girth at 4 years
- **y2** extension growth at 4 years
- **y3** trunk girth at 15 years
- **y4** weight of tree above ground at 15 years

**References**

Examples

data(rootstock)
attach(rootstock)
rs <- rootstock[,1]
rs <- factor(rs, ordered=is.ordered(rs)) # Too important a step
root.manova <- manova(cbind(y1,y2,y3,y4)~rs)
summary(root.manova, test = "Wilks")

Description

In the data set sample, we have data from five different probability distributions. Histograms are used to intuitively understand the underlying probability model.

Usage

data(sample)

Format

A data frame with 100 observations on the following 5 variables.

Sample_1  A sample 1
Sample_2  A sample 2
Sample_3  A sample 3
Sample_4  A sample 4
Sample_5  A sample 5

Examples

data(sample)
layout(matrix(c(1,1,2,2,3,3,0,4,4,5,5,0), 2, 6, byrow=TRUE),respect=FALSE)
matrix(c(1,1,2,2,3,3,0,4,4,5,5,0), 2, 6, byrow=TRUE)
hist(sample[,1],main="Histogram of Sample 1",xlab="sample1", ylab="frequency")
hist(sample[,2],main="Histogram of Sample 2",xlab="sample2", ylab="frequency")
hist(sample[,3],main="Histogram of Sample 3",xlab="sample3", ylab="frequency")
hist(sample[,4],main="Histogram of Sample 4",xlab="sample4", ylab="frequency")
hist(sample[,5],main="Histogram of Sample 5",xlab="sample5", ylab="frequency")
The Seishu Wine Study

Description

The odor and taste of wines are recorded in a study. It is believed that the variables such as the pH concentration, alcohol content, total sugar, etc, explain the odor and taste of the wine.

Usage

data(sheishu)

Format

A data frame with 30 observations on the following 10 variables.

Taste  taste
Odor   odor
pH     pH concentration
Acidity_1  Acidity 1
Acidity_2  Acidity 2
Sake_meter Sake meter
Direct_reducing_sugar Direct reducing sugar
Total_sugar Total sugar
Alcohol  type of alcohol
Formyl_nitrogen Formyl nitrogen

References


Examples

data(sheishu)
noc <- c(2,3,3,2)
nov <- 10
v <- nrow(sheishu)-1
vareishu <- var(sheishu)
s11 <- var(sheishu[1:2,1:2])
s22 <- var(sheishu[3:5,3:5])
s33 <- var(sheishu[6:8,6:8])
s44 <- var(sheishu[9:10,9:10])
u <- det(vareishu)/(det(s11)*det(s22)*det(s33)*det(s44))
a2 <- nov^2 - sum(noc^2)
a3 <- nov^3 - sum(noc^3)
f <- a2/2
The Shelf-Stocking Data

Description
A merchandiser stocks soft-drink on a shelf as a multiple number of the number of cases. The time required to put the cases in the shelves is recorded as a response. Clearly, if there are no cases to be stocked, it is natural that the time to put them on the shelf will be 0.

Usage
data("shelf_stock")

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

- Time  time required to put the cases in the shelves
- Cases_Stocked  number of cases

Examples
data(shelf_stock)
sslm <- lm(Time ~ Cases_Stocked -1, data=shelf_stock)

Siegel-Tukey Nonparametric Test

Description
This function provided an implementation of the nonparametric discussed in "Section 8.5.3 The Siegel-Tukey Test".

Usage
siegel.tukey(x, y)

Arguments
- x  Values from Sample 1
- y  Values from Sample 2
Details

For more details, refer Section 8.5.3 The Siegel-Tukey Test.

Author(s)

Prabhanjan N. Tattar

Examples

```r
x <- c(0.028, 0.029, 0.011, -0.030, 0.017, -0.012, -0.027, -0.018, 0.022, -0.023)
y <- c(-0.002, 0.016, 0.005, -0.001, 0.000, 0.008, -0.005, -0.009, 0.001, -0.019)
siegel.tukey(x, y)
```

**skewcoeff**

A simple and straightforward function to compute the coefficient of skewness

Description

The function is fairly easy to follow.

Usage

```r
skewcoeff(x)
```

Arguments

x variable of interest

Author(s)

Prabhanjan N. Tattar

See Also

e1071::skewness
**Description**

A cooked data tailor made for the use of scatter plots towards understanding correlations.

**Usage**

```r
data(somesamples)
```

**Format**

A data frame with 200 observations on the following 12 variables.

- `x1` x of Sample 1
- `y1` y of Sample 1
- `x2` x of Sample 2
- `y2` y of Sample 2
- `x3` x of Sample 3
- `y3` y of Sample 3
- `x4` x of Sample 4
- `y4` y of Sample 4
- `x5` x of Sample 5
- `y5` y of Sample 5
- `x6` x of Sample 6
- `y6` y of Sample 6

**Examples**

```r
data(somesamples)
attach(somesamples)
par(mfrow=c(2,3))
plot(x1,y1,main="Sample I",xlim=c(-4,4),ylim=c(-4,4))
plot(x2,y2,main="Sample II",xlim=c(-4,4),ylim=c(-4,4))
plot(x3,y3,main="Sample III",xlim=c(-4,4),ylim=c(-4,4))
plot(x4,y4,main="Sample IV",xlim=c(-4,4),ylim=c(-4,4))
plot(x5,y5,main="Sample V",xlim=c(-4,4),ylim=c(-4,4))
plot(x6,y6,main="Sample VI",xlim=c(-4,4),ylim=c(-4,4))
```
Understanding Strength of Paper with a Three Factorial Experiment

Description
The strength of a paper depends on three variables: (i) the percentage of hardwood concentration in the raw pulp, (ii) the vat pressure, and (iii) the cooking time of the pulp. The hardwood concentration is tested at three levels of 2, 4, and 8 percentage, the vat pressure at 400, 500, and 650, while the cooking time is at 3 and 4 hours. For each combination of these three factor variables, 2 observations are available, and thus a total of $3 \times 3 \times 2 \times 2 = 36$ observations. The goal of the study is investigation of the impact of the three factor variables on the strength of the paper, and the presence of interaction effect, if any.

Usage
data(SP)

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

- Hardwood: a factor with levels 2 4 8
- Pressure: a factor with levels 400 500 650
- Cooking_Time: a factor with levels 3 4
- Strength: a numeric vector

References

Examples
data(SP)
summary(SP.aov <- aov(Strength~.^3,SP))

stationdistTPM  
A function which will return the stationary distribution of an ergodic Markov chain

Description
This function returns the stationary distribution of an ergodic Markov chain. For details, refer Chapter 11 of the book.
The Board Stiffness Dataset

Description

Four measures of stiffness of 30 boards are available. The first measure of stiffness is obtained by sending a shock wave down the board, the second measure is obtained by vibrating the board, and remaining are obtained from static tests.

Usage

data(stiff)

Format

A data frame with 30 observations on the following 4 variables.

x1 first measure of stiffness is obtained by sending a shock wave down the board
x2 second measure is obtained by vibrating the board
x3 third measure is obtained by a static test
x4 fourth measure is obtained by a static test
References


Examples

```r
data(stiff)
colMeans(stiff)
var(stiff)
pairs(stiff)
```

<table>
<thead>
<tr>
<th>ST_Ordered</th>
<th>Simulating Random Observations from an Arbitrary Distribution</th>
</tr>
</thead>
</table>

Description

An implementation of the algorithm for simulation of observations from an arbitrary discrete distribution is provided here.

Usage

```r
ST_Ordered(N, x, p_x)
```

Arguments

- **N**: number of required random observations
- **x**: the possible values of the RV
- **p_x**: the probability vector associated with x

Author(s)

Prabhanjan N. Tattar

See Also

sample

Examples

```r
N <- 1e4
x <- 1:10
p_x <- c(0.05, 0.17, 0.02, 0.14, 0.11, 0.06, 0.05, 0.04, 0.17, 0.19)
table(ST_Ordered(N, x, p_x))
```
ST_Unordered

Simulating Random Observations from an Arbitrary Distribution (ordered probabilities)

Description
Simulation observations from an arbitrary discrete distribution with probabilities arranged in descending/ascending order.

Usage
ST_Unordered(N, x, p_x)

Arguments
N number of required random observations
x the possible values of the RV
p_x the probability vector associated with x

Author(s)
Prabhanjan N. Tattar

See Also
sample

Examples
N <- 1e2
x <- 1:10
p_x <- c(0.05, 0.17, 0.02, 0.14, 0.11, 0.06, 0.05, 0.04, 0.17, 0.19)
ST_Unordered(N, x, p_x)

swiss

Forged Swiss Bank Notes

Description
The swiss data set consists of measurements on the width of bottom margin and image diagonal length for forged and real notes. Histogram smoothing method is applied to understand the width of bottom margins for the forged notes.

Usage
data(swiss)
Format

A data frame with 100 observations on the following 4 variables.

Bottforg  bottom margin of forged notes
Diagforg  diagonal margin of forged notes
Bottreal  bottom margin of real notes
Diagreal  diagonal margin of real notes

References


Examples

data(swiss)
par(mfrow=c(1,3))
hist(swiss$Bottforg,breaks=28,probability=TRUE,col=0,ylim=c(0, .5),
   xlab="Margin width (mm)",ylab="Density")
hist(swiss$Bottforg,breaks=12,probability=TRUE,col=0,ylim=c(0, .5),
   xlab="Margin width (mm)",ylab="Density")
hist(swiss$Bottforg,breaks=6,probability=TRUE,col=0,ylim=c(0, .5),
   xlab="Margin width (mm)",ylab="Density")

The Toluca Company Labour Hours against Lot Size

Description

The Toluca Company manufactures equipment related to refrigerator. The company, in respect of a particular component of a refrigerator, has data on the labor hours required for the component in various lot sizes. Using this data, the officials wanted to find the optimum lot size for producing this part.

Usage

data("tc")

Format

A data frame with 25 observations on the following 2 variables.

Lot_Size  size of the lot
Labour_Hours  the labor hours required

References

**Description**

An engineer wants to find out if the cotton weight percentage in a synthetic fiber effects the tensile strength. Towards this, the cotton weight percentage is fixed at 5 different levels of 15, 20, 25, 30, and 35. Each level of the percentage is assigned 5 experimental units and the tensile strength is measured on each of them. The randomization is specified in the Run_Number column. The goal of the engineer is to investigate if the tensile strength is same across the cotton weight percentage.

**Usage**

data(tensile)

**Format**

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

- Test_Sequence  the test sequence
- Run_Number  the run number
- CWP  cotton weight percentage
- Tensile_Strength  the tensile strength

**References**


**Examples**

data(tensile)
tensile$CWP <- as.factor(tensile$CWP)
tensile_aov <- aov(Tensile_Strength~CWP, data=tensile)
summary(tensile_aov)
model.matrix(tensile_aov)
### testtpm

**A transition probability matrix**

**Description**

A transition probability matrix for understanding Markov chains.

**Usage**

```
data(testtpm)
```

**Format**

A matrix of transition probability matrix.

- A transitions probabilities from State A
- B transitions probabilities from State B
- C transitions probabilities from State C
- D transitions probabilities from State D
- E transitions probabilities from State E
- F transitions probabilities from State F

**Examples**

```
data(testtpm)
```

### testtpm2

**A matrix of transition probability matrix, second example**

**Description**

A transition probability matrix for understanding Markov chains.

**Usage**

```
data(testtpm2)
```

**Format**

A matrix of transition probability matrix.

- A transitions probabilities from State A
- B transitions probabilities from State B
- C transitions probabilities from State C
- D transitions probabilities from State D
- E transitions probabilities from State E
- F transitions probabilities from State F
Examples
data(testtpm3)

Description
A matrix of transition probability matrix, third example

Usage
data(testtpm3)

Format
A data frame with 7 observations on the following 7 variables.
A transitions probabilities from State A
B transitions probabilities from State B
C transitions probabilities from State C
D transitions probabilities from State D
E transitions probabilities from State E
F transitions probabilities from State F
G transitions probabilities from State G

Examples
data(testtpm3)

---

Trm

Trimmed Mean

Description
The trimean can be viewed as the average of median and average of the lower and upper quartiles. A fairly simply function is defined here.

Usage
Trm(x)

Arguments
x variable of interest
**Author(s)**
Prabhanjan N. Tattar

**See Also**
TMH, mean, median

---

**TMH**
*Trimean based on hinges instead of quartiles*

**Description**
The trimean is modified and defined based on hinges instead of the quartiles.

**Usage**

```
T MH(x)
```

**Arguments**

- `x` variable of interest

**Author(s)**
Prabhanjan N. Tattar

**See Also**
TM

---

**UMPExponential**
*Uniformly Most Powerful Test for Exponential Distribution*

**Description**
A function is defined here which will return the uniformly most powerful test for exponential distribution. The function needs a simple use of the "qgamma" function.

**Usage**

```
UMPExponential(theta0, n, alpha)
```

**Arguments**

- `theta0` the parameter of interest
- `n` the sample size
- `alpha` level of the UMP test
**UMPNormal**

*Uniformly Most Powerful Test for Normal Distribution*

**Description**

The "UMPNormal" function returns the critical points required for the UMP test for a sample from normal distribution. The standard deviation is assumed to be known.

**Usage**

```
UMPNormal(mu0, sigma, n, alpha)
```

**Arguments**

- `mu0` the value of mean of interest
- `sigma` standard deviation
- `n` sample size
- `alpha` size of the UMP test

**Author(s)**

Prabhanjan N. Tattar

---

**UMPUniform**

*Uniformly Most Powerful Test for Uniform Sample*

**Description**

A simple and straightforward function for obtaining the UMP test for a random sample from uniform distribution.

**Usage**

```
UMPUniform(theta0, n, alpha)
```

**Arguments**

- `theta0` the parameter value of interest
- `n` the sample size
- `alpha` the size of the required UMP test

**Author(s)**

Prabhanjan N. Tattar
Author(s)
Prabhanjan N. Tattar

Examples
\texttt{UMPUniform(0.6,10,0.05)}

\begin{tabular}{ll}
\texttt{usc} & \textit{US Crime Data} \\
\end{tabular}

Description
Data is available on the crime rates across 47 states in USA, and we have additional information on 13 more explanatory variables.

Usage
data(usc)

Format
A data frame with 47 observations on the following 14 variables.

\texttt{R} Crime rate - the number of offenses known to the police per 1,000,000 population
\texttt{Age} Age distribution - the number of males aged 14 to 24 years per 1000 of total state population
\texttt{S} Binary variable distinguishing southern states (S = 1) from the rest
\texttt{Ed} Educational level - mean number of years of schooling times 10 of the population 25 years old and over
\texttt{Ex0} Police expenditure - per capita expenditure on police protection by state and local governments in 1960
\texttt{Ex1} Police expenditure - as Ex0, but for 1959
\texttt{LF} Labour force participation rate per 1000 civilian urban males in the age group 14 to 24 years
\texttt{M} Number of males per 1000 females
\texttt{N} State population size in hundred thousands
\texttt{NW} Number of non-whites per 1000
\texttt{U1} Unemployment rate of urban males per 1000 in the age group 14 to 24 years
\texttt{U2} Unemployment rate of urban males per 1000 in the age group 35 to 39 years
\texttt{W} Wealth, as measured by the median value of transferable goods and assets. or family income (unit 10 dollars)
\texttt{X} Income inequality: the number of families per 1000 earning below one half of the median income

References
The Box-Cox Transformation for Viscosity Dataset

Description
The goal of this study is to find the impact of temperature on the viscosity of toluence-tetralin blends.

Usage
data(viscos)

Format
A data frame with 8 observations on the following 2 variables.

Temperature  temperature
Viscosity  viscosity of toluence-tetralin blends

References

Examples
data(viscos)
names(viscos)
viscoslm <- lm(Viscosity~Temperature, data=viscos)
vonNeumann  

*von Neumann Random Number Generator*

**Description**

The "vonNeumann" function implements the von Neumann random generator as detailed in Section 11.2.

**Usage**

`vonNeumann(x, n)`

**Arguments**

- `x`  
  the initial seed
- `n`  
  number of required observations

**Author(s)**

Prabhanjan N. Tattar

**Examples**

- `vonNeumann(x=11, n=10)`
- `vonNeumann(x=675248, n=10)`
- `vonNeumann(x=8653, n=100)`

waterquality  

*Testing for Physico-chemical Properties of Water in 4 Cities*

**Description**

Water samples from four cities are collected and their physico-chemical properties for ten variables, such as pH, Conductivity, Dissolution, etc., are measured. We would then like to test if the properties are same across the four cities and in which case a same water treatment approach can be adopted across the cities.

**Usage**

`data(waterquality)`
Format

A data frame with 63 observations on the following 10 variables.

City  four cities City1 City2 City3 City4
pH   the pH concentration
Conductivity water conductivity
Dissolution water dissolution
Alkalinity  alkalinity of the water sample
Hardness  water hardness
Calcium.Hardness  calcium hardness of the water
Magnesium.Hardness  magnesium hardness of the water
Chlorides  chloride content
Sulphates  sulphate content

References


Examples

data(waterquality)

Description

The Wilson confidence interval for a sample from binomial distribution is a complex formula. This function helps the reader in easily obtaining the required confidence interval as discussed and detailed in Section 16.5.

Usage

WilsonCI(x, n, alpha)

Arguments

x   the number of successes
n   the number of trials
alpha  the confidence interval size

Author(s)

Prabhanjan N. Tattar
Examples

WilsonCI(x=10658, n=15000, alpha=0.05)
prop.test(x=10658, n=15000)$conf.int

ww.test

Wald-Wolfowitz Nonparametric Test

Description

The "ww.test" function is an implementation of the famous Wald-Wolfowitz nonparametric test as discussed in Section 8.5.

Usage

ww.test(x, y)

Arguments

x values from sample 1
y values from sample 2

Author(s)

Prabhanjan N. Tattar

x_bimodal

Understanding kernel smoothing through a simulated dataset

Description

This is a simulated dataset with two modes at -2 and 2 and we have 400 observations.

Usage

data(x_bimodal)

Format

The format is: num [1:400] -4.68 -4.19 -4.05 -4.04 -4.02 ...
Youden and Beale's Data on Lesions of Half-Leaves of Tobacco Plant

Description

A simple and innovative design is often priceless. Youden and Beale (1934) sought to find the effect of two preparations of virus on tobacco plants. One half of a tobacco leaf was rubbed with cheesecloth soaked in one preparation of the virus extract and the second half was rubbed with the other virus extract. This experiment was replicated on just eight leaves, and the number of lesions on each half leaf was recorded.

Usage

data(yb)

Format

A data frame with 8 observations on the following 2 variables.

Preparation_1 a numeric vector
Preparation_2 a numeric vector
References


Examples

data(yb)
summary(yb)
quantile(yb$Preparation_1,seq(0,1,.1)) # here seq gives 0, .1, .2, ..., 1
quantile(yb$Preparation_2,seq(0,1,.1))
fivenum(yb$Preparation_1)
fivenum(yb$Preparation_2)
Index

*Topic ANOVA
  SP, 56
*Topic Arbitrary discrete distribution
  ST_Ordered, 58
  ST_Unordered, 59
*Topic Balanced incomplete block design
  reaction, 47
*Topic Binomial Distribution
  Binom_Sim, 7
*Topic Box-Cox transformation
  viscos, 67
*Topic Cauchy random samples, score function
  cs, 18
*Topic EDA summary
  TM, 63
  TMH, 64
*Topic EDA resistant_line, 48
*Topic Ehrenfest
  Ehrenfest, 21
*Topic Ergodic Markov chain
  stationdistTPM, 56
*Topic Exponential Distribution
  UMPExponential, 64
*Topic Geometric distribution
  Geom_Sim, 23
*Topic Graeco Latin square design
  rocket_Graeco, 49
*Topic Kruskal-Wallis test
  Mucociliary, 37
*Topic Letter values
  lval, 33
*Topic Likelihood Ratio Test
  LRNormalMean_KV, 31
  LRNormalMean_UV, 32
*Topic Likelihood Ratio test
  LRNormal2Mean, 30
*Topic Likelihood ratio test
  LRNormalVariance_UW, 32
*Topic Manova
  waterquality, 68
*Topic Markov chain
  Ehrenfest, 21
  testtpm, 62
  testtpm2, 62
  testtpm3, 63
*Topic Most Powerful Test
  MPNormal, 36
  MPoisson, 36
*Topic Most powerful test
  MPbinomial, 35
*Topic Nonparametric Test
  siegel.tukey, 53
*Topic Normal Distribution
  LRNormal2Mean, 30
  LRNormalMean_KV, 31
  LRNormalMean_UV, 32
  MPNormal, 36
*Topic Normal distribution
  UMPNormal, 65
*Topic Pareto Density
  pareto_density, 41
*Topic Poisson Distribution
  MPoisson, 36
  Poisson_Sim, 42
*Topic Poisson regression model
  caesareans, 11
*Topic Poisson regression
  bs1, 10
*Topic Quantile of Pareto RV
  pareto_quantile, 42
*Topic Quesenberry-Hurst Simultaneous Confidence Interval
  QH_CI, 47
*Topic Resistant Line
resistant_line, 48
*Topic Siegel-Tukey
  siegel.tukey, 53
*Topic Simulation unordered probabilities
  ST_Ordered, 58
  ST_Unordered, 59
*Topic Simulation
  Binom_Sim, 7
  Geom_Sim, 23
  Poisson_Sim, 42
*Topic Statistics
  ACSWR-package, 3
*Topic Testing for sphericity
  pw, 46
*Topic Transition Probability Matrix
  msteptpm, 37
*Topic Trimean based on hinges
  TMH, 64
*Topic Trimean
  TM, 63
*Topic Uniform Distribution
  UMUniform, 65
*Topic Uniformly Most Powerful Test
  UMPExponential, 64
  UMPNormal, 65
  UMUniform, 65
*Topic Wald-Wolfowitz
  wwNtest, 70
*Topic Wilcoxon test, Hamilton depression scale
  depression, 20
*Topic Wilson confidence interval
  WilsonCI, 69
*Topic binomial distribution
  mpbinomial, 35
  WilsonCI, 69
*Topic block experiment
  hardness, 24
*Topic blocked design
  intensity, 27
*Topic canonical correlation
  chemicaldata, 15
*Topic categorical data analysis
  atombomb, 6
*Topic completely randomized design
 olson, 40
  tensile, 61
*Topic datasets
  abrasion_index, 4
*Topic descriptive statistics
  kurtcoeff, 28
*Topic empirical distribution function
  nerve, 38
*Topic exploratory data analysis
  lval, 33
  yb, 71
*Topic factor analysis
  adjectives, 5
  life, 28
*Topic factorial experiment
  battery, 6
*Topic fivenum, IQR
  memory, 33
*Topic histogram smoothing
  swiss, 59
*Topic histogram
  sample, 51
*Topic hypothesis testing for equality of covariance matrices
  mfp, 34
*Topic kernel smoothing
  x_bimodal, 70
*Topic kurtosis
  kurtcoeff, 28
*Topic latin square design
  rocket, 49
*Topic linear regression model
  shelf_stock, 53
  tc, 60
*Topic logistic regression, model selection
  lowbwt, 29
*Topic logistic regression
  chdage, 14
*Topic m-step
  msteptpm, 37
*Topic manova
  rootstock, 50
*Topic maximum likelihood estimation
  ns, 39
*Topic maximum likelihood estimator
  ps, 44
*Topic median polish
  girder, 23
*Topic multiple linear regression model
usc, 66
*Topic multivariate dataset, singular value decomposition, cork dataset
cork, 17
*Topic multivariate hypothesis testing
calcium, 12
*Topic multivariate summaries
stiff, 57
*Topic nonparametric test
ww.test, 70
*Topic normal distribution
LRNormalVariance_UM, 32
*Topic power plot
powertestplot, 43
*Topic principal component analysis
hearing, 25
*Topic random generator
vonNeumann, 68
*Topic residual analysis, logistic regression
Disease, 20
*Topic rootogram, militiamen
chest, 16
*Topic scatterplot
somesamples, 55
*Topic score function, binomial distribution
bs, 9
*Topic skewness coefficient
skewcoeff, 54
*Topic stationary distribution
stationdistTPM, 56
*Topic stem-and-leaf plot, cloud seeding experiment
cloud, 17
*Topic summary statistics
kurtcoeff, 28
*Topic testing independence of subvectors
sheishu, 52
*Topic testing mean vector
hw, 26
*Topic three factorial experiment
bottling, 8
*Topic time series plot
insurance, 26
*Topic transformation, variance reduction
flight, 22
*Topic visualizing multivariate data
cardata, 13
*Topic von Neumann
vonNeumann, 68
abrasion_index, 4
ACSWR (ACSWR-package), 3
ACSWR-package, 3
adjectives, 5
atombomb, 6
battery, 6
Binom_Sim, 7
bottling, 8
bs, 9
bs1, 10
caesareans, 11
calcium, 12
cardata, 13
chdage, 14
chemicaldata, 15
chest, 16
cloud, 17
cork, 17
cs, 18
depression, 20
Disease, 20
Ehrenfest, 21
flight, 22
Geom_Sim, 23
girder, 23
hardness, 24
hearing, 25
hw, 26
insurance, 26
intensity, 27
kurtcoeff, 28
life, 28