Package ‘HSAUR3’

October 12, 2022

Title A Handbook of Statistical Analyses Using R (3rd Edition)
Date 2022-04-25
Version 1.0-13
Author Torsten Hothorn and Brian S. Everitt
Maintainer Torsten Hothorn <Torsten.Hothorn@R-project.org>
Description Functions, data sets, analyses and examples from the third edition of the book "A Handbook of Statistical Analyses Using R" (Torsten Hothorn and Brian S. Everitt, Chapman & Hall/CRC, 2014). The first chapter of the book, which is entitled "An Introduction to R", is completely included in this package, for all other chapters, a vignette containing all data analyses is available. In addition, Sweave source code for slides of selected chapters is included in this package (see HSAUR3/inst/slides). The publishers web page is 'http://www.crcpress.com/product/isbn/9781482204582'.
Depends R (>= 3.0.0), tools
Suggests boot (>= 1.3-9), lattice (>= 0.20-23), MASS (>= 7.3-29), mgcv (>= 1.7-27), rpart (>= 4.1-4), survival (>= 2.37-4), ape (>= 3.0-11), coin (>= 1.1-3), flexmix (>= 2.3-11), Formula (>= 1.1-1), gamair (>= 0.0.8), gamlss.data (>= 4.2.6), gee (>= 4.13-18), KernSmooth (>= 2.23-10), lme4 (>= 1.0-5), maps (>= 2.3-6), maptools (>= 0.8-27), mboost (>= 2.2-3), mclust (>= 4.2), mlbench (>= 2.1-1), mice (>= 2.18), multicomp (>= 1.3-1), mvtnorm (>= 0.9-9996), partykit (>= 0.8-0), quantreg (>= 5.05), randomForest (>= 4.6-7), rmeta (>= 2.16), sandwich (>= 2.3-0), scatterplot3d (>= 0.3-34), sp (>= 1.0-14), TH.data (>= 1.0-2), vcd (>= 1.3-1), wordcloud (>= 2.4), HSAUR2
LazyData yes
License GPL-2
Encoding UTF-8
NeedsCompilation no
Repository CRAN
Date/Publication 2022-04-26 04:52:12 UTC
R topics documented:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>agefat</td>
<td>3</td>
</tr>
<tr>
<td>aspirin</td>
<td>4</td>
</tr>
<tr>
<td>backpain</td>
<td>5</td>
</tr>
<tr>
<td>BCG</td>
<td>6</td>
</tr>
<tr>
<td>birds</td>
<td>7</td>
</tr>
<tr>
<td>birthdeathrates</td>
<td>8</td>
</tr>
<tr>
<td>bladdercancer</td>
<td>8</td>
</tr>
<tr>
<td>bp</td>
<td>9</td>
</tr>
<tr>
<td>BtheB</td>
<td>10</td>
</tr>
<tr>
<td>CHFLS</td>
<td>11</td>
</tr>
<tr>
<td>clouds</td>
<td>13</td>
</tr>
<tr>
<td>CYGOB1</td>
<td>15</td>
</tr>
<tr>
<td>EFT</td>
<td>16</td>
</tr>
<tr>
<td>epilepsy</td>
<td>16</td>
</tr>
<tr>
<td>Forbes2000</td>
<td>17</td>
</tr>
<tr>
<td>foster</td>
<td>18</td>
</tr>
<tr>
<td>gardenflowers</td>
<td>19</td>
</tr>
<tr>
<td>GHQ</td>
<td>19</td>
</tr>
<tr>
<td>heptathlon</td>
<td>20</td>
</tr>
<tr>
<td>household</td>
<td>22</td>
</tr>
<tr>
<td>HSAURtable</td>
<td>23</td>
</tr>
<tr>
<td>Lanza</td>
<td>24</td>
</tr>
<tr>
<td>mastectomy</td>
<td>25</td>
</tr>
<tr>
<td>men1500m</td>
<td>26</td>
</tr>
<tr>
<td>meteo</td>
<td>26</td>
</tr>
<tr>
<td>orallesions</td>
<td>27</td>
</tr>
<tr>
<td>phosphate</td>
<td>28</td>
</tr>
<tr>
<td>pistonrings</td>
<td>29</td>
</tr>
<tr>
<td>planets</td>
<td>29</td>
</tr>
<tr>
<td>plasma</td>
<td>30</td>
</tr>
<tr>
<td>polyps</td>
<td>31</td>
</tr>
<tr>
<td>polyps3</td>
<td>32</td>
</tr>
<tr>
<td>pottery</td>
<td>33</td>
</tr>
<tr>
<td>rearrests</td>
<td>34</td>
</tr>
<tr>
<td>respiratory</td>
<td>35</td>
</tr>
<tr>
<td>roomwidth</td>
<td>36</td>
</tr>
<tr>
<td>schizophrenia</td>
<td>37</td>
</tr>
<tr>
<td>schizophrenia2</td>
<td>38</td>
</tr>
<tr>
<td>schooldays</td>
<td>39</td>
</tr>
<tr>
<td>skulls</td>
<td>40</td>
</tr>
<tr>
<td>smoking</td>
<td>41</td>
</tr>
<tr>
<td>Smoking_DollHill1950</td>
<td>42</td>
</tr>
<tr>
<td>Smoking_Mueller1940</td>
<td>42</td>
</tr>
<tr>
<td>Smoking_SchainerScoeniger1944</td>
<td>43</td>
</tr>
<tr>
<td>Smoking_Wassink1945</td>
<td>44</td>
</tr>
<tr>
<td>students</td>
<td>44</td>
</tr>
</tbody>
</table>
**Description**

Age and body fat percentage of 25 normal adults.

**Usage**

data("agefat")

**Format**

A data frame with 25 observations on the following 3 variables.

- **age**: the age of the subject.
- **fat**: the body fat percentage.
- **gender**: a factor with levels female and male.

**Details**

The data come from a study investigating a new methods of measuring body composition (see Mazess et al, 1984), and give the body fat percentage (percent fat), age and gender for 25 normal adults aged between 23 and 61 years. The questions of interest are how are age and percent fat related, and is there any evidence that the relationship is different for males and females.

**Source**

Examples

```r
data("agefat", package = "HSAUR3")
plot(fat ~ age, data = agefat)
```

---

**aspirin**  
**Aspirin Data**

**Description**

Efficacy of Aspirin in preventing death after a myocardial infarct.

**Usage**

```r
data("aspirin")
```

**Format**

A data frame with 7 observations on the following 4 variables.

- **dp** number of deaths after placebo.
- **tp** total number subjects treated with placebo.
- **da** number of deaths after Aspirin.
- **ta** total number of subjects treated with Aspirin.

**Details**

The data were collected for a meta-analysis of the effectiveness of Aspirin (versus placebo) in preventing death after a myocardial infarction.

**Source**


**Examples**

```r
data("aspirin", package = "HSAUR3")
aspirin
```
Description
A case-control study to investigate whether driving a car is a risk factor for low back pain resulting from acute herniated lumbar intervertebral discs (AHLID).

Usage
data("backpain")

Format
A data frame with 434 observations on the following 4 variables.

- **ID**  a factor which identifies matched pairs.
- **status**  a factor with levels case and control.
- **driver**  a factor with levels no and yes.
- **suburban**  a factor with levels no and yes indicating a suburban resident.

Details
These data arise from a study reported in Kelsey and Hardy (1975) which was designed to investigate whether driving a car is a risk factor for low back pain resulting from acute herniated lumbar intervertebral discs (AHLID). A case-control study was used with cases selected from people who had recently had X-rays taken of the lower back and had been diagnosed as having AHLID. The controls were taken from patients admitted to the same hospital as a case with a condition unrelated to the spine. Further matching was made on age and sex and a total of 217 matched pairs were recruited, consisting of 89 female pairs and 128 male pairs.

Source

Examples
```r
data("backpain", package = "HSAUR3")
summary(backpain)
```
**BCG Vaccine Data**

**Description**

A meta-analysis on the efficacy of BCG vaccination against tuberculosis (TB).

**Usage**

data("BCG")

**Format**

A data frame with 13 observations on the following 7 variables.

- **Study** an identifier of the study.
- **BCGTB** the number of subjects suffering from TB after a BCG vaccination.
- **BCGVacc** the number of subjects with BCG vaccination.
- **NoVaccTB** the number of subjects suffering from TB without BCG vaccination.
- **NoVacc** the total number of subjects without BCG vaccination.
- **Latitude** geographic position of the place the study was undertaken.
- **Year** the year the study was undertaken.

**Details**

Bacille Calmette Guerin (BCG) is the most widely used vaccination in the world. Developed in the 1930s and made of a live, weakened strain of Mycobacterium bovis, the BCG is the only vaccination available against tuberculosis today. Colditz et al. (1994) report data from 13 clinical trials of BCG vaccine each investigating its efficacy in the treatment of tuberculosis. The number of subjects suffering from TB with or without BCG vaccination are given here. In addition, the data contains the values of two other variables for each study, namely, the geographic latitude of the place where the study was undertaken and the year of publication. These two variables will be used to investigate and perhaps explain any heterogeneity among the studies.

**Source**

Example

data("BCG", package = "HSAUR3")

### sort studies w.r.t. sample size
BCG <- BCG[order(rowSums(BCG[,2:5])),]

### to long format
BCGlong <- with(BCG, data.frame(Freq = c(BCGTB, BCGVacc - BCGTB,
NoVaccTB, NoVacc - NoVaccTB),
infected = rep(rep(factor(c("yes", "no")),
rep(nrow(BCG), 2), 2),
vaccinated = rep(factor(c("yes", "no")),
rep(nrow(BCG) * 2, 2),
study = rep(factor(Study, levels = as.character(Study)),
4)))

### doubledecker plot
library("vcd")
doubledecker(xtabs(Freq ~ study + vaccinated + infected,
data = BCGlong))

---

**birds**

**Birds in Paramo Vegetation Data**

**Description**

The data were originally derived from a study which investigated numbers of bird species in isolated islands of paramo vegetation in the northern Andes.

**Usage**

data(birds)

**Format**

A data frame with 14 observations on the following 5 variables.

- N number of species
- AR area of island in thousands of square km
- EL elevation in thousands of m
- Dec distance from Ecuador in km
- DNI distance to the nearest island in km

**Source**

birthdeathrates  
*Birth and Death Rates Data*

**Description**

Birth and death rates for 69 countries.

**Usage**

```r
data("birthdeathrates")
```

**Format**

A data frame with 69 observations on the following 2 variables.
- `birth`  birth rate.
- `death`  death rate.

**Source**


**Examples**

```r
data("birthdeathrates", package = "HSAUR3")
plot(birthdeathrates)
```

---

bladdercancer  
*Bladder Cancer Data*

**Description**

Data arise from 31 male patients who have been treated for superficial bladder cancer, and give the number of recurrent tumours during a particular time after the removal of the primary tumour, along with the size of the original tumour.

**Usage**

```r
data("bladdercancer")
```

**Format**

A data frame with 31 observations on the following 3 variables.
- `time`  the duration.
- `tumorsize`  a factor with levels <=3cm and >3cm.
- `number`  number of recurrent tumours.
Details

The aim is the estimate the effect of size of tumour on the number of recurrent tumours.

Source


Examples

data("bladdercancer", package = "HSAUR3")
mosaicplot(xtabs(~ number + tumorsize, data = bladdercancer))

---

bp  
Lowering Blood Pressure Data  

Description

Lowering a patient’s blood pressure during surgery, using a hypotensive drug.

Usage

data(bp)

Format

A data frame with 53 observations on the following 3 variables.

- logdose: the logarithm (base 10) of the dose of drug in milligrams
- bloodp: average systolic blood pressure achieved while the drug was being administered
- recovtime: time in minutes before the patient’s systolic blood pressure returned to 100mm of mercury

Source

J. D. Robertson and P. Armitage (1959) Comparison of Two Hypotensive Agents, Anaesthesia, 14(1), 53–64
Beat the Blues Data

Description

Data from a clinical trial of an interactive multimedia program called ‘Beat the Blues’.

Usage

data("BtheB")

Format

A data frame with 100 observations of 100 patients on the following 8 variables.

- **drug**: did the patient take anti-depressant drugs (No or Yes).
- **length**: the length of the current episode of depression, a factor with levels <6m (less than six months) and >6m (more than six months).
- **treatment**: treatment group, a factor with levels TAU (treatment as usual) and BtheB (Beat the Blues).
- **bdi.pre**: Beck Depression Inventory II before treatment.
- **bdi.2m**: Beck Depression Inventory II after two months.
- **bdi.3m**: Beck Depression Inventory II after one month follow-up.
- **bdi.5m**: Beck Depression Inventory II after three months follow-up.
- **bdi.8m**: Beck Depression Inventory II after six months follow-up.

Details

Longitudinal data from a clinical trial of an interactive, multimedia program known as "Beat the Blues" designed to deliver cognitive behavioural therapy to depressed patients via a computer terminal. Patients with depression recruited in primary care were randomised to either the Beating the Blues program, or to "Treatment as Usual (TAU)".

Note that the data are stored in the wide form, i.e., repeated measurements are represented by additional columns in the data frame.

Source

Examples

```r
data("BtheB", package = "HSAUR3")
layout(matrix(1:2, nrow = 1))
ylim <- range(BtheB[,grep("bdi", names(BtheB))], na.rm = TRUE)
boxplot(subset(BtheB, treatment == "TAU")[,grep("bdi", names(BtheB))],
    main = "Treated as usual", ylab = "BDI",
    xlab = "Time (in months)", names = c(0, 2, 3, 5, 8), ylim = ylim)
boxplot(subset(BtheB, treatment == "BtheB")[,grep("bdi", names(BtheB))],
    main = "Beat the Blues", ylab = "BDI", xlab = "Time (in months)",
    names = c(0, 2, 3, 5, 8), ylim = ylim)
```

CHFLS

Chinese Health and Family Life Survey

Description

The Chinese Health and Family Life Survey sampled 60 villages and urban neighborhoods chosen in such a way as to represent the full geographical and socioeconomic range of contemporary China.

Usage

```r
data("CHFLS")
```

Format

A data frame with 1534 observations on the following 10 variables.

- **R_region**: a factor with levels Coastal South, Coastal East, Inlands, North, Northeast, Central West.
- **R_age**: age of the responding woman.
- **R_edu**: education level of the responding woman, an ordered factor with levels Never attended school < Elementary school < Junior high school < Senior high school < Junior college < University.
- **R_income**: monthly income of the responding woman.
- **R_health**: self-reported health status, an ordered factor with levels Poor < Not good < Fair < Good < Excellent.
- **R_height**: height of the responding woman.
- **R_happy**: self-reported happiness of the responding woman, an ordered factor with levels Very unhappy < Not too happy < Somewhat happy < Very happy.
- **A_height**: height of the responding woman’s partner.
- **A_edu**: level of education of the woman’s partner, an ordered factor with levels Never attended school < Elementary school < Junior high school < Senior high school < Junior college < University.
- **A_income**: monthly income of the woman’s partner.
Details
Contemporary China is on the leading edge of a sexual revolution, with tremendous regional and
generational differences that provide unparalleled natural experiments for analysis of the antecedents
as a collaborative research project of the Universities of Chicago, Beijing, and North Carolina, pro-
vides a baseline from which to anticipate and track future changes. Specifically, this study produces
a baseline set of results on sexual behavior and disease patterns, using a nationally representative
probability sample. The Chinese Health and Family Life Survey sampled 60 villages and urban
neighborhoods chosen in such a way as to represent the full geographical and socioeconomic range
of contemporary China excluding Hong Kong and Tibet. Eighty-three individuals were chosen at
random for each location from official registers of adults aged between 20 and 64 years to target
a sample of 5000 individuals in total. Here, we restrict our attention to women with current male
partners for whom no information was missing, leading to a sample of 1534 women. The data have
been extracted as given in the example section.

Source
https://sscs.uchicago.edu

References
William L. Parish, Edward O. Laumann, Myron S. Cohen, Suiming Pan, Heyi Zheng, Irving Hoffman,
Tianfu Wang, and Kwai Hang Ng. (2003), Population-Based Study of Chlamydial Infection

Examples
## Not run:
### for a description see http://popcenter.uchicago.edu/data/chfls.shtml
library("TH.data")
load(file.path(path.package(package="TH.data"), "rda", "CHFLS.rda"))

tmp <- chfls1[, c("REGION6", "ZJ05", "ZJ06", "A35", "ZJ07", "ZJ16M", "INCRM",
"JK01", "JK02", "JK20", "HY04", "HY07", "A02", "AGEGAPH",

names(tmp) <- c("Region",
"Rgender", ### gender of respondent
"Rage", ### age of respondent
"RagestartA", ### age of respondent at beginning of relationship
"Redu", ### with partner A
"RincomeM", ### education of respondent
"RincomeComp", ### rounded monthly income of respondent
"Rhealth", ### health condition respondent
"Rheight", ### respondent's height
"Rhappy", ### respondent's happiness
"Rmartial", ### respondent's marital status
"RhasA", ### R has current A partner
"Agender", ### gender of partner A
"RAagegap", ### age gap
"RAstartage", ### age at marriage
"Aheight", ### height of partner A
"Aedu", ### education of partner A
"AINcomeM", ### rounded partner A income
"AINcomeEst", ### estimated partner A income
"orgasm", ### orgasm frequency
"AINcomeComp", ### imputed partner A income
"Rsexnow", ### has sex last year
"Rhomosexual") ### R is homosexual

### code missing values
tmp$AincomeM[tmp$AincomeM < 0] <- NA
tmp$RincomeM[tmp$RincomeM < 0] <- NA
tmp$Aheight[tmp$Aheight < 0] <- NA

olevels <- c("never", "rarely", "sometimes", "often", "always")
tmpA <- subset(tmp, Rgender == "female" & Rhomosexual != "yes" & orgasm %in% olevels)

### 1534 subjects
dim(tmpA)

CHFLS <- tmp[, c("Region", "R_age", "R_edu", "RincomeComp", "Rhealth", "Rheight", "Rhappy",
"Aheight", "Aedu", "AincomeComp")]
names(CHFLS) <- c("R_region", "R_age", "R_edu", "R_income", "R_health", "R_height",
"R_happy", "A_height", "A_edu", "A_income")
levels(CHFLS$R_region) <- c("Coastal South", "Coastal East", "Inlands", "North",
"Northeast", "Central West")

CHFLS$R_edu <- ordered(as.character(CHFLS$R_edu), levels = c("no school", "primary",
"low mid", "up mid", "j col", "univ/grad"))
levels(CHFLS$R_edu) <- c("Never attended school", "Elementary school", "Junior high school",
"Senior high school", "Junior college", "University")

CHFLS$A_edu <- ordered(as.character(CHFLS$A_edu), levels = c("no school", "primary",
"low mid", "up mid", "j col", "univ/grad"))
levels(CHFLS$A_edu) <- c("Never attended school", "Elementary school", "Junior high school",
"Senior high school", "Junior college", "University")

CHFLS$R_health <- ordered(as.character(CHFLS$R_health), levels = c("poor", "not good",
"fair", "good", "excellent"))
levels(CHFLS$R_health) <- c("Poor", "Not good", "Fair", "Good", "Excellent")

CHFLS$R_happy <- ordered(as.character(CHFLS$R_happy), levels = c("v unhappy", "not too",
"relatively", "very"))
levels(CHFLS$R_happy) <- c("Very unhappy", "Not too happy", "Relatively happy", "Very happy")

## End(Not run)
Description

Data from an experiment investigating the use of massive amounts of silver iodide (100 to 1000 grams per cloud) in cloud seeding to increase rainfall.

Usage

data("clouds")

Format

A data frame with 24 observations on the following 7 variables.

- **seeding**: a factor indicating whether seeding action occurred (no or yes).
- **time**: number of days after the first day of the experiment.
- **sne**: suitability criterion.
- **cloudcover**: the percentage cloud cover in the experimental area, measured using radar.
- **prewetness**: the total rainfall in the target area one hour before seeding (in cubic metres times $1e+8$).
- **echomotion**: a factor showing whether the radar echo was moving or stationary.
- **rainfall**: the amount of rain in cubic metres times $1e+8$.

Details

Weather modification, or cloud seeding, is the treatment of individual clouds or storm systems with various inorganic and organic materials in the hope of achieving an increase in rainfall. Introduction of such material into a cloud that contains supercooled water, that is, liquid water colder than zero Celsius, has the aim of inducing freezing, with the consequent ice particles growing at the expense of liquid droplets and becoming heavy enough to fall as rain from clouds that otherwise would produce none.

The data available in cloud were collected in the summer of 1975 from an experiment to investigate the use of massive amounts of silver iodide 100 to 1000 grams per cloud) in cloud seeding to increase rainfall. In the experiment, which was conducted in an area of Florida, 24 days were judged suitable for seeding on the basis that a measured suitability criterion (SNE).

Source


Examples

data("clouds", package = "HSAUR3")
layout(matrix(1:2, nrow = 2))
boxplot(rainfall ~ seeding, data = clouds, ylab = "Rainfall")
boxplot(rainfall ~ echomotion, data = clouds, ylab = "Rainfall")
**Description**

Energy output and surface temperature for Star Cluster CYG OB1.

**Usage**

```r
data("CYGOB1")
```

**Format**

A data frame with 47 observations on the following 2 variables.

- `logst` log surface temperature of the star.
- `logli` log light intensity of the star.

**Details**

The Hertzsprung-Russell (H-R) diagram forms the basis of the theory of stellar evolution. The diagram is essentially a plot of the energy output of stars plotted against their surface temperature. Data from the H-R diagram of Star Cluster CYG OB1, calibrated according to VanismaGreve1972 are given here.

**Source**


**Examples**

```r
data("CYGOB1", package = "HSAUR3")
plot(logst ~ logli, data = CYGOB1)
```
EFT  

*Embedded Figures Test Data*

**Description**

Embedded figures test for 24 school children.

**Usage**

data(EFT)

**Format**

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

- **group**: a factor with levels row corner
- **time**: time to complete the pattern
- **EFT**: Embedded Figures Test

**Source**


---

epilepsy  

*Epilepsy Data*

**Description**

A randomised clinical trial investigating the effect of an anti-epileptic drug.

**Usage**

data("epilepsy")

**Format**

A data frame with 236 observations on the following 6 variables.

- **treatment**: the treatment group, a factor with levels placebo and Progabide.
- **base**: the number of seizures before the trial.
- **age**: the age of the patient.
- **seizure.rate**: the number of seizures (response variable).
- **period**: treatment period, an ordered factor with levels 1 to 4.
- **subject**: the patient ID, a factor with levels 1 to 59.
Details
In this clinical trial, 59 patients suffering from epilepsy were randomized to groups receiving either the anti-epileptic drug Progabide or a placebo in addition to standard chemotherapy. The numbers of seizures suffered in each of four, two-week periods were recorded for each patient along with a baseline seizure count for the 8 weeks prior to being randomized to treatment and age. The main question of interest is whether taking progabide reduced the number of epileptic seizures compared with placebo.

Source

Examples

```r
data("epilepsy", package = "HSAUR3")
library(lattice)
dotplot(I(seizure.rate / base) ~ period | subject, data = epilepsy, subset = treatment == "Progabide")
dotplot(I(seizure.rate / base) ~ period | subject, data = epilepsy, subset = treatment == "Progabide")
```

Forbes2000


Description
The Forbes 2000 list is a ranking of the world’s biggest companies, measured by sales, profits, assets and market value.

Usage

```r
data("Forbes2000")
```

Format
A data frame with 2000 observations on the following 8 variables.

- **rank** the ranking of the company.
- **name** the name of the company.
- **country** a factor giving the country the company is situated in.
- **category** a factor describing the products the company produces.
- **sales** the amount of sales of the company in billion USD.
- **profits** the profit of the company in billion USD.
- **assets** the assets of the company in billion USD.
- **marketvalue** the market value of the company in billion USD.
Source


Examples

data("Forbes2000", package = "HSAUR3")
summary(Forbes2000)
### number of countries
length(levels(Forbes2000$country))
### number of industries
length(levels(Forbes2000$category))

foster Foster Feeding Experiment

Description

The data are from a foster feeding experiment with rat mothers and litters of four different genotypes. The measurement is the litter weight after a trial feeding period.

Usage

data("foster")

Format

A data frame with 61 observations on the following 3 variables.

- `litgen` genotype of the litter, a factor with levels A, B, I, and J.
- `motgen` genotype of the mother, a factor with levels A, B, I, and J.
- `weight` the weight of the litter after a feeding period.

Details

Here the interest lies in uncovering the effect of genotype of mother and litter on litter weight.

Source


Examples

data("foster", package = "HSAUR3")
plot.design(foster)
**gardenflowers**

---

**gardenflowers**  
*Garden Flowers Data*

---

**Description**

The dissimilarity matrix of 18 species of garden flowers.

**Usage**

```r
data("gardenflowers")
```

**Format**

An object of class `dist`.

**Details**

The dissimilarity was computed based on certain characteristics of the flowers.

**Source**


**Examples**

```r
data("gardenflowers", package = "HSAUR3")
gardenflowers
```

---

**GHQ**  
*General Health Questionnaire*

---

**Description**

Data from a psychiatric screening questionnaire.

**Usage**

```r
data("GHQ")
```
Format

A data frame with 22 observations on the following 4 variables.

GHQ  the General Health Questionnaire score.
gender  a factor with levels female and male
cases  the number of diseased subjects.
non.cases  the number of healthy subjects.

Details

The data arise from a study of a psychiatric screening questionnaire called the GHQ (General Health Questionnaire, see Goldberg, 1972). Here the main question of interest is to see how caseness is related to gender and GHQ score.

Source


Examples

```r
data("GHQ", package = "HSAUR3")
male <- subset(GHQ, gender == "male")
female <- subset(GHQ, gender == "female")
layout(matrix(1:2, ncol = 2))
barplot(t(as.matrix(male[,c("cases", "non.cases")])), main = "Male", xlab = "GHC score")
barplot(t(as.matrix(male[,c("cases", "non.cases")])), main = "Female", xlab = "GHC score")
```

heptathlon

*Olympic Heptathlon Seoul 1988*

Description


Usage

```r
data("heptathlon")
```
heptathlon

Format

A data frame with 25 observations on the following 8 variables.

- hurdles  results 100m hurdles.
- highjump  results high jump.
- shot  results shot.
- run200m  results 200m race.
- longjump  results long jump.
- javelin  results javelin.
- run800m  results 800m race.
- score  total score.

Details

The first combined Olympic event for women was the pentathlon, first held in Germany in 1928. Initially this consisted of the shot putt, long jump, 100m, high jump and javelin events held over two days. The pentathlon was first introduced into the Olympic Games in 1964, when it consisted of the 80m hurdles, shot, high jump, long jump and 200m. In 1977 the 200m was replaced by the 800m and from 1981 the IAAF brought in the seven-event heptathlon in place of the pentathlon, with day one containing the events-100m hurdles, shot, high jump, 200m and day two, the long jump, javelin and 800m. A scoring system is used to assign points to the results from each event and the winner is the woman who accumulates the most points over the two days. The event made its first Olympic appearance in 1984.

In the 1988 Olympics held in Seoul, the heptathlon was won by one of the stars of women’s athletics in the USA, Jackie Joyner-Kersee. The results for all 25 competitors are given here.

Source


Examples

```r
data("heptathlon", package = "HSAUR3")
plot(heptathlon)
```
**Description**

Survey data on household expenditure on four commodity groups.

**Usage**

```r
data("household")
```

**Format**

A data frame with 40 observations on the following 5 variables.

- **housing**: expenditure on housing, including fuel and light.
- **food**: expenditure on foodstuffs, including alcohol and tobacco.
- **goods**: expenditure on other goods, including clothing, footwear and durable goods.
- **service**: expenditure on services, including transport and vehicles.
- **gender**: a factor with levels female and male

**Details**

The data are part of a data set collected from a survey of household expenditure and give the expenditure of 20 single men and 20 single women on four commodity groups. The units of expenditure are Hong Kong dollars.

**Source**

FIXME

**Examples**

```r
data("household", package = "HSAUR3")
```
HSAURtable

Produce LaTeX Tables

Description

Generate longtable LaTeX environments.

Usage

HSAURtable(object, ...)  
## S3 method for class 'table'  
HSAURtable(object, xname = deparse(substitute(object)), pkg = NULL, ...)
## S3 method for class 'data.frame'  
HSAURtable(object, xname = deparse(substitute(object)), pkg = NULL, nrows = NULL, ...)
## S3 method for class 'tabtab'  
toListe(object, caption = NULL, label = NULL, topcaption = TRUE, index = TRUE, ...)
## S3 method for class 'dftab'  
toListe(object, pcol = 1, caption = NULL, label = NULL, rownames = FALSE, topcaption = TRUE, index = TRUE, ...)

Arguments

object an object of table or data.frame.
xname the name of the object.
pkg the package object comes from, optionally.
nrows the number of rows actually printed for a data.frame.
caption the (optional) caption of the table without label.
label the (optional) label to be defined for this table.
pcol the number of parallel columns.
rownames logical, should the rownames be printed in the first row without column name?
topcaption logical, should the captions be placed on top (default) of the table?
index logical, should an index entry be generated?
... additional arguments, currently ignored.

Details

Based on the data in object, an object from which a Latex table (in a longtable environment) may be constructed (via toLatex) is generated.
Value

An object of class `tabtab` or `dftab` for which `toLatex` methods are available. `toLatex` produces objects of class `Latex`, a character vector, essentially.

Examples

```r
data("rearrests", package = "HSAUR3")
toLatex(HSAURtable(rearrests),
    caption = "Rearrests of juvenile felons.",
    label = "rearrests_tab")
```

---

### Lanza Prevention of Gastrointestinal Damages

Description

Data from four randomised clinical trials on the prevention of gastrointestinal damages by Misoprostol reported by Lanza et al. (1987, 1988a,b, 1989).

Usage

```r
data("Lanza")
```

Format

A data frame with 198 observations on the following 3 variables.

- `study` a factor with levels I, II, III, and IV describing the study number.
- `treatment` a factor with levels Misoprostol Placebo
- `classification` an ordered factor with levels 1 < 2 < 3 < 4 < 5 describing an ordered response variable.

Details

The response variable is defined by the number of haemorrhages or erosions.

Source


Examples

```r
data("Lanza", package = "HSAUR3")
layout(matrix(1:4, nrow = 2))
pl <- tapply(1:nrow(Lanza), Lanza$study, function(indx)
  mosaicplot(table(Lanza[indx,"treatment"],
              Lanza[indx,"classification"]),
            main = "", shade = TRUE))
```

### Examples

```r
data("Lanza", package = "HSAUR3")
pl <- tapply(1:nrow(Lanza), Lanza$study, function(indx)
  mosaicplot(table(Lanza[indx,"treatment"],
              Lanza[indx,"classification"]),
            main = "", shade = TRUE))
```

**Survival Times after Mastectomy of Breast Cancer Patients**

**Description**

Survival times in months after mastectomy of women with breast cancer. The cancers are classified as having metastasized or not based on a histochemical marker.

**Usage**

```r
data("mastectomy")
```

**Format**

- A data frame with 42 observations on the following 3 variables.
  - **time** survival times in months.
  - **event** a logical indicating if the event was observed (TRUE) or if the survival time was censored (FALSE).
  - **metastasized** a factor at levels yes and no.

**Source**


**Examples**

```r
data("mastectomy", package = "HSAUR3")
table(mastectomy$metastasized)
```
**men1500m**

*Winners of the Olympic Men's 1500m*

**Description**

The data gives the winners of the men's 1500m race for the Olympic Games 1896 to 2004.

**Usage**

```r
data("men1500m")
```

**Format**

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

- **year**: the olympic year.
- **venue**: city where the games took place.
- **winner**: winner of men's 1500m race.
- **country**: country the winner came from.
- **time**: time (in seconds) of the winner.

**Examples**

```r
data("men1500m", package = "HSAUR3")
op <- par(las = 2)
plot(time ~ year, data = men1500m, axes = FALSE)
yrs <- seq(from = 1896, to = 2004, by = 4)
axis(1, at = yrs, labels = yrs)
axis(2)
box()
par(op)
```

---

**meteo**

*Meteorological Measurements for 11 Years*

**Description**

Several meteorological measurements for a period between 1920 and 1931.

**Usage**

```r
data("meteo")
```
Format

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

year  the years.
rainNovDec  rainfall in November and December (mm).
temp  average July temperature.
rainJuly  rainfall in July (mm).
radiation  radiation in July (millilitres of alcohol).
yield  average harvest yield (quintals per hectare).

Details

Carry out a principal components analysis of both the covariance matrix and the correlation matrix of the data and compare the results. Which set of components leads to the most meaningful interpretation?

Source


Examples

```r
data("meteo", package = "HSAUR3")
meto
```

orallesions  Oral Lesions in Rural India

Description

The distribution of the oral lesion site found in house-to-house surveys in three geographic regions of rural India.

Usage

```r
data("orallesions")
```

Format

A two-way classification, see table.

Source

Examples

```r
data("orallesions", package = "HSAUR3")
mosaicplot(orallesions)
```

---

**phosphate**  
*Phosphate Level Data*

Description

Plasma inorganic phosphate levels from 33 subjects.

Usage

```r
data("phosphate")
```

Format

A data frame with 33 observations on the following 9 variables.

- **group**: a factor with levels `control` and `obese`.
- **t0**: baseline phosphate level.
- **t0.5**: phosphate level after 1/2 an hour.
- **t1**: phosphate level after one an hour.
- **t1.5**: phosphate level after 1 1/2 hours.
- **t2**: phosphate level after two hours.
- **t3**: phosphate level after three hours.
- **t4**: phosphate level after four hours.
- **t5**: phosphate level after five hours.

Source


Examples

```r
data("phosphate", package = "HSAUR3")
plot(t0 ~ group, data = phosphate)
```
pistonrings | Piston Rings Failures
---|---

**Description**
Number of failures of piston rings in three legs of four steam-driven compressors.

**Usage**
data("pistonrings")

**Format**
A two-way classification, see table.

**Details**
The data are given in form of a table. The table gives the number of piston-ring failures in each of three legs of four steam-driven compressors located in the same building. The compressors have identical design and are oriented in the same way. The question of interest is whether the two classification variables (compressor and leg) are independent.

**Source**

**Examples**

data("pistonrings", package = "HSAUR3")
mosaicplot(pistonrings)

---

planets | Exoplanets Data
---|---

**Description**
Data on planets outside the Solar System.

**Usage**
data("planets")
Format

A data frame with 101 observations from 101 exoplanets on the following 3 variables.

- **mass**: Jupiter mass of the planet.
- **period**: period in earth days.
- **eccen**: the radial eccentricity of the planet.

Details

From the properties of the exoplanets found up to now it appears that the theory of planetary development constructed for the planets of the Solar System may need to be reformulated. The exoplanets are not at all like the nine local planets that we know so well. A first step in the process of understanding the exoplanets might be to try to classify them with respect to their known properties.

Source


Examples

```r
data("planets", package = "HSAUR3")
require("scatterplot3d")
scatterplot3d(log(planets$mass), log(planets$period), log(planets$eccen),
              type = "h", highlight.3d = TRUE, angle = 55, scale.y = 0.7, pch = 16)
```

---

plasma  

*Blood Screening Data*

Description

The erythrocyte sedimentation rate and measurements of two plasma proteins (fibrinogen and globulin).

Usage

```r
data("plasma")
```

Format

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

- **fibrinogen**: the fibrinogen level in the blood.
- **globulin**: the globulin level in the blood.
- **ESR**: the erythrocyte sedimentation rate, either less or greater 20 mm/hour.
Details

The erythrocyte sedimentation rate (ESR) is the rate at which red blood cells (erythrocytes) settle out of suspension in blood plasma, when measured under standard conditions. If the ESR increases when the level of certain proteins in the blood plasma rise in association with conditions such as rheumatic diseases, chronic infections and malignant diseases, its determination might be useful in screening blood samples taken form people suspected to being suffering from one of the conditions mentioned. The absolute value of the ESR is not of great importance rather it is whether it is less than 20mm/hr since lower values indicate a healthy individual.

The question of interest is whether there is any association between the probability of an ESR reading greater than 20mm/hr and the levels of the two plasma proteins. If there is not then the determination of ESR would not be useful for diagnostic purposes.

Source


Examples

data("plasma", package = "HSAUR3")
layout(matrix(1:2, ncol = 2))
boxplot(fibrinogen ~ ESR, data = plasma, varwidth = TRUE)
boxplot(globulin ~ ESR, data = plasma, varwidth = TRUE)

---

polyps

Familial Andenomatous Polyposis

Description

Data from a placebo-controlled trial of a non-steroidal anti-inflammatory drug in the treatment of familial adenomatous polyposis (FAP).

Usage

data("polyps")

Format

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

number number of colonic polyps at 12 months.
treat treatment arms of the trail, a factor with levels placebo and drug.
age the age of the patient.
Details

Giardiello et al. (1993) and Piantadosi (1997) describe the results of a placebo-controlled trial of a non-steroidal anti-inflammatory drug in the treatment of familial adenomatous polyposis (FAP). The trial was halted after a planned interim analysis had suggested compelling evidence in favour of the treatment. Here we are interested in assessing whether the number of colonic polyps at 12 months is related to treatment and age of patient.

Source


Examples

data("polyps", package = "HSAUR3")
plot(number ~ age, data = polyps, pch = as.numeric(polyps$treat))
legend(40, 40, legend = levels(polyps$treat), pch = 1:2, bty = "n")

---

polyps3  *Familial Adenomatous Polyposis*

Description

Data from a placebo-controlled trial of a non-steroidal anti-inflammatory drug in the treatment of familial adenomatous polyposis (FAP).

Usage

data("polyps3")

Format

A data frame with 22 observations on the following 5 variables.

gender  a factor with levels female and male.
treatment a factor with levels placebo and active.
baseline  the baseline number of polyps.
age  the age of the patient.
number3m  the number of polyps after three month.
Details

The data arise from the same study as the \textit{polyps} data. Here, the number of polyps after three months are given.

Source


Examples

```r
data("polyps3", package = "HSAUR3")
plot(number3m ~ age, data = polyps3, pch = as.numeric(polyps3$treatment))
legend("topright", legend = levels(polyps3$treatment), pch = 1:2, bty = "n")
```

---

**pottery**

\textit{Romano-British Pottery Data}

Description

Chemical composition of Romano-British pottery.

Usage

```r
data("pottery")
```

Format

A data frame with 45 observations on the following 9 chemicals.

- **Al2O3** aluminium trioxide.
- **Fe2O3** iron trioxide.
- **MgO** magnesium oxide.
- **CaO** calcium oxide.
- **Na2O** natrium oxide.
- **K2O** calium oxide.
- **TiO2** titanium oxide.
- **MnO** mangan oxide.
- **BaO** barium oxide.
- **kiln** site at which the pottery was found.
Details
The data gives the chemical composition of specimens of Romano-British pottery, determined by atomic absorption spectrophotometry, for nine oxides.

Source

Examples
```r
data("pottery", package = "HSAUR3")
plot(pottery)
```

---

rearrests

Rearrests of Juvenile Felons

Description
Rearrests of juvenile felons by type of court in which they were tried.

Usage
data("rearrests")

Format
A two-way classification, see table.

Details
The data (taken from Agresti, 1996) arise from a sample of juveniles convicted of felony in Florida in 1987. Matched pairs were formed using criteria such as age and the number of previous offences. For each pair, one subject was handled in the juvenile court and the other was transferred to the adult court. Whether or not the juvenile was rearrested by the end of 1988 was then noted. Here the question of interest is whether the true proportions rearrested were identical for the adult and juvenile court assignments?

Source

Examples
```r
data("rearrests", package = "HSAUR3")
rearrests
```
Description

The respiratory status of patients recruited for a randomised clinical multicenter trial.

Usage

data("respiratory")

Format

A data frame with 555 observations on the following 7 variables.

centre the study center, a factor with levels 1 and 2.
treatment the treatment arm, a factor with levels placebo and treatment.
gender a factor with levels female and male.
age the age of the patient.
status the respiratory status (response variable), a factor with levels poor and good.
month the month, each patient was examined at months 0, 1, 2, 3 and 4.
subject the patient ID, a factor with levels 1 to 111.

Details

In each of two centres, eligible patients were randomly assigned to active treatment or placebo. During the treatment, the respiratory status (categorised poor or good) was determined at each of four, monthly visits. The trial recruited 111 participants (54 in the active group, 57 in the placebo group) and there were no missing data for either the responses or the covariates. The question of interest is to assess whether the treatment is effective and to estimate its effect.

Note that the data are in long form, i.e, repeated measurements are stored as additional rows in the data frame.

Source


Examples

data("respiratory", package = "HSAUR3")
mosaicplot(xtabs(~ treatment + month + status, data = respiratory))
roomwidth

Students Estimates of Lecture Room Width

Description

Lecture room width estimated by students in two different units.

Usage

data("roomwidth")

Format

A data frame with 113 observations on the following 2 variables.

unit  a factor with levels feet and metres.
width  the estimated width of the lecture room.

Details

Shortly after metric units of length were officially introduced in Australia, each of a group of 44 students was asked to guess, to the nearest metre, the width of the lecture hall in which they were sitting. Another group of 69 students in the same room was asked to guess the width in feet, to the nearest foot. The data were collected by Professor T. Lewis and are taken from Hand et al (1994). The main question is whether estimation in feet and in metres gives different results.

Source


Examples

data("roomwidth", package = "HSAUR3")
convert <- ifelse(roomwidth$unit == "feet", 1, 3.28)
boxplot(I(width * convert) ~ unit, data = roomwidth)
Schizophrenia

Age of Onset of Schizophrenia Data

Description

Data on sex differences in the age of onset of schizophrenia.

Usage

data("schizophrenia")

Format

A data frame with 251 observations on the following 2 variables.

age  age at the time of diagnosis.

gender  a factor with levels female and male

Details

A sex difference in the age of onset of schizophrenia was noted by Kraepelin (1919). Subsequently epidemiological studies of the disorder have consistently shown an earlier onset in men than in women. One model that has been suggested to explain this observed difference is know as the subtype model which postulates two type of schizophrenia, one characterised by early onset, typical symptoms and poor premorbid competence, and the other by late onset, atypical symptoms, and good premorbid competence. The early onset type is assumed to be largely a disorder of men and the late onset largely a disorder of women.

Source


Examples

data("schizophrenia", package = "HSAUR3")
boxplot(age ~ gender, data = schizophrenia)


schizophrenia2  Schizophrenia Data

Description

Though disorder and early onset of schizophrenia.

Usage

data("schizophrenia2")

Format

A data frame with 220 observations on the following 4 variables.

subject  the patient ID, a factor with levels 1 to 44.
onset  the time of onset of the disease, a factor with levels < 20 yrs and > 20 yrs.
disorder  whether thought disorder was absent or present, the response variable.
month  month after hospitalisation.

Details

The data were collected in a follow-up study of women patients with schizophrenia. The binary response recorded at 0, 2, 6, 8 and 10 months after hospitalisation was thought disorder (absent or present). The single covariate is the factor indicating whether a patient had suffered early or late onset of her condition (age of onset less than 20 years or age of onset 20 years or above). The question of interest is whether the course of the illness differs between patients with early and late onset?

Source


Examples

data("schizophrenia2", package = "HSAUR3")
mosaicplot(xtabs(~onset+month+disorder, data = schizophrenia2))
schooldays

Days not Spent at School

Description

Data from a sociological study, the number of days absent from school is the response variable.

Usage

data("schooldays")

Format

A data frame with 154 observations on the following 5 variables.

- `race`  race of the child, a factor with levels aboriginal and non-aboriginal.
- `gender`  the gender of the child, a factor with levels female and male.
- `school`  the school type, a factor with levels F0 (primary), F1 (first), F2 (second) and F3 (third form).
- `learner`  how good is the child in learning things, a factor with levels average and slow.
- `absent`  number of days absent from school.

Details

The data arise from a sociological study of Australian Aboriginal and white children reported by Quine (1975).

In this study, children of both sexes from four age groups (final grade in primary schools and first, second and third form in secondary school) and from two cultural groups were used. The children in age group were classified as slow or average learners. The response variable was the number of days absent from school during the school year. (Children who had suffered a serious illness during the years were excluded.)

Source

S. Quine (1975), Achievement Orientation of Aboriginal and White Adolescents. Doctoral Dissertation, Australian National University, Canberra.

Examples

data("schooldays", package = "HSAUR3")
plot.design(schooldays)
skulls  

Egyptian Skulls

Description

Measurements made on Egyptian skulls from five epochs.

Usage

data("skulls")

Format

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

- **epoch**: the epoch the skull as assigned to, a factor with levels c4000BC, c3300BC, c1850BC, c200BC, and cAD150, where the years are only given approximately, of course.
- **mb**: maximum breaths of the skull.
- **bh**: basibregmatic heights of the skull.
- **bl**: basialveolar length of the skull.
- **nh**: nasal heights of the skull.

Details

The question is whether the measurements change over time. Non-constant measurements of the skulls over time would indicate interbreeding with immigrant populations.

Source


Examples

data("skulls", package = "HSAUR3")
means <- tapply(1:nrow(skulls), skulls$epoch, function(i)
               apply(skulls[i, colnames(skulls)[-1]], 2, mean))
means <- matrix(unlist(means), nrow = length(means), byrow = TRUE)
colnames(means) <- colnames(skulls)[-1]
rownames(means) <- levels(skulls$epoch)
pairs(means,
       panel = function(x, y) {
           text(x, y, levels(skulls$epoch))
       })
Description

Data from a meta-analysis on nicotine gum and smoking cessation

Usage

data("smoking")

Format

A data frame with 26 observations (studies) on the following 4 variables.

qt  the number of treated subjects who stopped smoking.

nt  the total number of treated subjects.

qc  the number of subjects who stopped smoking without being treated.

tc  the total number of subjects not being treated.

Details

Cigarette smoking is the leading cause of preventable death in the United States and kills more Americans than AIDS, alcohol, illegal drug use, car accidents, fires, murders and suicides combined. It has been estimated that 430,000 Americans die from smoking every year. Fighting tobacco use is, consequently, one of the major public health goals of our time and there are now many programs available designed to help smokers quit. One of the major aids used in these programs is nicotine chewing gum, which acts as a substitute oral activity and provides a source of nicotine that reduces the withdrawal symptoms experienced when smoking is stopped. But separate randomized clinical trials of nicotine gum have been largely inconclusive, leading Silagy (2003) to consider combining the results studies found from an extensive literature search. The results of these trials in terms of numbers of people in the treatment arm and the control arm who stopped smoking for at least 6 months after treatment are given here.

Source


Examples

data("smoking", package = "HSAUR3")
boxplot(smoking$qt/smoking$tt, smoking$qc/smoking$tc,
       names = c("Treated", "Control"), ylab = "Percent Quitters")
Description

Number of smokers in a case-control study.

Usage

data(Smoking_DollHill1950)

Format

The format is: table [1:6, 1:2, 1:2] 2 33 250 196 136 32 27 55 293 190 ... - attr(*, "dimnames")=List of 3 ..$ Smok- ing : chr [1:6] "Nonsmoker" "1-" "5-" "15-" ... ..$ Diagnosis: chr [1:2] "Lung cancer" "Other" ..$ Sex : chr [1:2] "Male" "Female"

Details

This is Table V from Doll and Hill (1950).

Source

Richard Doll and A. Bradford Hill (1950), Smoking and Carcinoma of the Lung, British Medical Journal, 2, 739-748

References

Richard Doll (1998), Uncovering the effects of smoking: historical perspective. Statistical Methods in Medical Research, 7(87), 87-117

Description

Number of smokers in a case-control study.

Usage

data(Smoking_Mueller1940)

Format

The format is: table [1:5, 1:2] 25 18 13 27 3 4 5 22 41 14 - attr(*, "dimnames")=List of 2 ..$ Smoking_type: chr [1:5] "Extreme smoker" "Very heavy smoker" "Heavy smoker" "Moderate smoker" ... ..$ Group : chr [1:2] "Lung cancer" "Healthy control"
Details

Extreme smoker: 10-15 cigars, >35 cigarettes, or >50 g pipe tobacco/day. Very heavy smoker: 7-9 cigars, 26-35 cigarettes, or 36-50 g pipe tobacco/day. Heavy smoker: 4-6 cigars, 16-25 cigarettes, or 21-35 g pipe tobacco/day. Moderate smoker: 1-3 cigars, 1-15 cigarettes, or 1-20 g pipe tobacco/day.

Source


References

Richard Doll (1998), Uncovering the effects of smoking: historical perspective. Statistical Methods in Medical Research, 7(87), 87-117

Examples

data(Smoking_Mueller1940)
## maybe str(Smoking_Mueller1940) ; plot(Smoking_Mueller1940) ...

---

Smoking_SchraireSchoeniger1944

Smoking and Lung Cancer (II)

Description

Number of smokers in a case-control study.

Usage

data(Smoking_SchraireSchoeniger1944)

Format

The format is: table [1:5, 1:7] 3 11 31 19 29 2 0 4 6 3 ... - attr(*, "dimnames")=List of 2 ..$ Smoking : chr [1:5] "Nonsmoker" "Moderate smoker" "Medium smoker" "Heavy smoker" "... ..$ Diagnosis: chr [1:7] "Lung cancer" "Lip cancer" "Throat cancer" "Stomach cancer" ...

Source

E. Schairer and E. Schoninger (1944), Lungenkrebs und Tabakverbrauch, Zeitschrift fuer Krebsforschung, 54(4), 261-269

References

Richard Doll (1998), Uncovering the effects of smoking: historical perspective. Statistical Methods in Medical Research, 7(87), 87-117
**Smoking_Wassink1945  Smoking and Lung Cancer (III)**

**Description**

Number of smokers in a case-control study.

**Usage**

```r
data(Smoking_Wassink1945)
```

**Format**


**Source**


**References**

Richard Doll (1998), Uncovering the effects of smoking: historical perspective. *Statistical Methods in Medical Research*, 7(87), 87-117

---

**students  Student Risk Taking**

**Description**

Students were administered two parallel forms of a test after a random assignment to three different treatments.

**Usage**

```r
data("students")
```

**Format**

A data frame with 35 observations on the following 3 variables.

- **treatment** a factor with levels AA, C, and NC.
- **low** the result of the first test.
- **high** the result of the second test.
suicides

Details
The data arise from a large study of risk taking (Timm, 2002). Students were randomly assigned to three different treatments labelled AA, C and NC. Students were administered two parallel forms of a test called low and high. The aim is to carry out a test of the equality of the bivariate means of each treatment population.

Source

Examples

```r
data("students", package = "HSAUR3")
layout(matrix(1:2, ncol = 2))
boxplot(low ~ treatment, data = students, ylab = "low")
boxplot(high ~ treatment, data = students, ylab = "high")
```

suicides
Crowd Baiting Behaviour and Suicides

Description
Data from a study carried out to investigate the causes of jeering or baiting behaviour by a crowd when a person is threatening to commit suicide by jumping from a high building.

Usage
data("suicides")

Format
A two-way classification, see table.

Source

Examples

```r
data("suicides", package = "HSAUR3")
mosaicplot(suicides)
```
suicides2  

**Male Suicides Data**

**Description**

Number of suicides in different age groups and countries.

**Usage**

data("suicides2")

**Format**

A data frame with 15 observations on the following 5 variables.

- A25.34 number of suicides (per 100000 males) between 25 and 34 years old.
- A35.44 number of suicides (per 100000 males) between 35 and 44 years old.
- A45.54 number of suicides (per 100000 males) between 45 and 54 years old.
- A55.64 number of suicides (per 100000 males) between 55 and 64 years old.
- A65.74 number of suicides (per 100000 males) between 65 and 74 years old.

**Details**

Each of the numbers gives the number of suicides per 100000 male inhabitants of the countries given by the row names.

toenail  

**Toenail Infection Data**

**Description**

Results of a clinical trial to compare two competing oral antifungal treatments for toenail infection.

**Usage**

data("toenail")

**Format**

A data frame with 1908 observations on the following 5 variables.

- patientID a unique identifier for each patient in the trial.
- outcome degree of separation of the nail plate from the nail bed (onycholysis).
- treatment a factor with levels itraconazole and terbinafine.
- time the time in month when the visit actually took place.
- visit number of visit attended.
De Backer et al. (1998) describe a clinical trial to compare two competing oral antifungal treatments for toenail infection (dermatophyte onychomycosis). A total of 378 patients were randomly allocated into two treatment groups, one group receiving 250mg per day of terbinafine and the other group 200mg per day of itraconazole. Patients were evaluated at seven visits, intended to be at weeks 0, 4, 8, 12, 24, 36, and 48 for the degree of separation of the nail plate from the nail bed (onycholysis) dichotomized into moderate or severe and none or mild. But patients did not always arrive exactly at the scheduled time and the exact time in months that they did attend was recorded. The data is not balanced since not all patients attended for all seven planned visits.


**Examples**

```r
data("toenail", package = "HSAUR3")
```

**Description**

Meta-analysis of studies comparing two different toothpastes.

**Usage**

```r
data("toothpaste")
```

**Format**

A data frame with 9 observations on the following 7 variables.

- **Study** the identifier of the study.
- **nA** number of subjects using toothpaste A.
- **meanA** mean DMFS index of subjects using toothpaste A.
- **sdA** standard deviation of DMFS index of subjects using toothpaste A.
- **nB** number of subjects using toothpaste B.
- **meanB** mean DMFS index of subjects using toothpaste B.
- **sdB** standard deviation of DMFS index of subjects using toothpaste B.
Details

The data are the results of nine randomised trials comparing two different toothpastes for the prevention of caries development. The outcomes in each trial was the change, from baseline, in the decayed, missing (due to caries) and filled surface dental index (DMFS).

Source


Examples

```r
data("toothpaste", package = "HSAUR3")
toothpaste
```

---

### USairpollution

**Air Pollution in US Cities**

**Description**

Air pollution data of 41 US cities.

**Usage**

```r
data("USairpollution")
```

**Format**

A data frame with 41 observations on the following 7 variables.

- **SO2**: SO2 content of air in micrograms per cubic metre.
- **temp**: average annual temperature in Fahrenheit.
- **manu**: number of manufacturing enterprises employing 20 or more workers.
- **popul**: population size (1970 census); in thousands.
- **wind**: average annual wind speed in miles per hour.
- **precip**: average annual precipitation in inches.
- **predays**: average number of days with precipitation per year.

**Details**

The annual mean concentration of sulphur dioxide, in micrograms per cubic metre, is a measure of the air pollution of the city. The question of interest here is what aspects of climate and human ecology as measured by the other six variables in the data determine pollution?
USmelanoma

Source


Examples

data("USairpollution", package = "HSAUR3")

<table>
<thead>
<tr>
<th>USmelanoma</th>
<th>USA Malignant Melanoma Data</th>
</tr>
</thead>
</table>

Description

USA mortality rates for white males due to malignant melanoma 1950-1969.

Usage

data("USmelanoma")

Format

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

- mortality: number of white males died due to malignant melanoma 1950-1969 per one million inhabitants.
- latitude: latitude of the geographic centre of the state.
- longitude: longitude of the geographic centre of each state.
- ocean: a binary variable indicating contiguity to an ocean at levels no or yes.

Details

Fisher and van Belle (1993) report mortality rates due to malignant melanoma of the skin for white males during the period 1950-1969, for each state on the US mainland. Questions of interest about these data include how do the mortality rates compare for ocean and non-ocean states?

Source

Fisher and van Belle (1993)

Examples

data("USmelanoma", package = "HSAUR3")
USstates  

Description
Socio-demographic variables for ten US states.

Usage
data(USstates)

Format
A data frame with 10 observations on the following 7 variables.

Population population size divided by 1000
Income average per capita income
Illiteracy illiteracy rate (per cent population)
Life.Expectancy life expectancy (years)
Homicide homicide rate (per 1000)
Graduates percentage of high school graduates
Freezing average number of days per below freezing

Details
The data set contains values of seven socio-demographic variables for ten states in the USA.

UStemp  

Description
Lowest temperatures in Fahrenheit in 22 US cities in four months.

Usage
data(UStemp)

Format
A data frame with 22 observations on the following 4 variables.

January lowest temperature in Fahrenheit
April lowest temperature in Fahrenheit
July lowest temperature in Fahrenheit
October lowest temperature in Fahrenheit
House of Representatives Voting Data

Description

Voting results for 15 congressmen from New Jersey.

Usage

data("voting")

Format

A 15 times 15 matrix.

Details

Romesburg (1984) gives a set of data that shows the number of times 15 congressmen from New Jersey voted differently in the House of Representatives on 19 environmental bills. Abstentions are not recorded.

Source


Examples

data("voting", package = "HSAUR3")
require("MASS")
voting_mds <- isoMDS(voting)
plot(voting_mds$points[,1], voting_mds$points[,2],
     type = "n", xlab = "Coordinate 1", ylab = "Coordinate 2",
     xlim = range(voting_mds$points[,1])*1.2)
text(voting_mds$points[,1], voting_mds$points[,2],
     labels = colnames(voting))
voting_sh <- Shepard(voting[lower.tri(voting)], voting_mds$points)
Description
The mortality and drinking water hardness for 61 cities in England and Wales.

Usage
data("water")

Format
A data frame with 61 observations on the following 4 variables.

location a factor with levels North and South indicating whether the town is as north as Derby.
town the name of the town.
mortality averaged annual mortality per 100,000 male inhabitants.
hardness calcium concentration (in parts per million).

Details
The data were collected in an investigation of environmental causes of disease. They show the annual mortality per 100,000 for males, averaged over the years 1958-1964, and the calcium concentration (in parts per million) in the drinking water for 61 large towns in England and Wales. The higher the calcium concentration, the harder the water. Towns at least as far north as Derby are identified in the table. Here there are several questions that might be of interest including, are mortality and water hardness related, and do either or both variables differ between northern and southern towns?

Source

Examples
data("water", package = "HSAUR3")
plot(mortality ~ hardness, data = water,
    col = as.numeric(water$location))
watervoles  Water Voles Data

Description

Percentage incidence of the 13 characteristics of water voles in 14 areas.

Usage

data("watervoles")

Format

A dissimilarity matrix for the following 14 variables, i.e., areas: Surrey, Shropshire, Yorkshire, Perthshire, Aberdeen, Elean Gamhna, Alps, Yugoslavia, Germany, Norway, Pyrenees I, Pyrenees II, North Spain, and South Spain.

Details

Corbet et al. (1970) report a study of water voles (genus Arvicola) in which the aim was to compare British populations of these animals with those in Europe, to investigate whether more than one species might be present in Britain. The original data consisted of observations of the presence or absence of 13 characteristics in about 300 water vole skulls arising from six British populations and eight populations from the rest of Europe. The data are the percentage incidence of the 13 characteristics in each of the 14 samples of water vole skulls.

Source


Examples

data("watervoles", package = "HSAUR3")
watervoles
waves

Electricity from Wave Power at Sea

Description

Measurements of root mean square bending moment by two different mooring methods.

Usage

data("waves")

Format

A data frame with 18 observations on the following 2 variables.

- `method1` Root mean square bending moment in Newton metres, mooring method 1
- `method2` Root mean square bending moment in Newton metres, mooring method 2

Details

In a design study for a device to generate electricity from wave power at sea, experiments were carried out on scale models in a wave tank to establish how the choice of mooring method for the system affected the bending stress produced in part of the device. The wave tank could simulate a wide range of sea states and the model system was subjected to the same sample of sea states with each of two mooring methods, one of which was considerably cheaper than the other. The question of interest is whether bending stress differs for the two mooring methods.

Source


Examples

data("waves", package = "HSAUR3")
plot(method1 ~ method2, data = waves)
weightgain

Gain in Weight of Rats

Description

The data arise from an experiment to study the gain in weight of rats fed on four different diets, distinguished by amount of protein (low and high) and by source of protein (beef and cereal).

Usage

data("weightgain")

Format

A data frame with 40 observations on the following 3 variables.

- source  source of protein given, a factor with levels Beef and Cereal.
- type    amount of protein given, a factor with levels High and Low.
- weightgain  weight gain in grams.

Details

Ten rats are randomized to each of the four treatments. The question of interest is how diet affects weight gain.

Source


Examples

data("weightgain", package = "HSAUR3")
interaction.plot(weightgain$type, weightgain$source, weightgain$weightgain)
Womens Role in Society

Description

Data from a survey from 1974 / 1975 asking both female and male responders about their opinion on the statement: Women should take care of running their homes and leave running the country up to men.

Usage

data("womensrole")

Format

A data frame with 42 observations on the following 4 variables.

- education  years of education.
- gender  a factor with levels Male and Female.
- agree  number of subjects in agreement with the statement.
- disagree  number of subjects in disagreement with the statement.

Details

The data are from Haberman (1973) and also given in Collett (2003). The questions here are whether the response of men and women differ.

Source


Examples

```
data("womensrole", package = "HSAUR3")
summary(subset(womensrole, gender == "Female"))
summary(subset(womensrole, gender == "Male"))
```
Index

* datasets
  agefat, 3
  aspirin, 4
  backpain, 5
  BCG, 6
  birds, 7
  birthdeathrates, 8
  bladdercancer, 8
  bp, 9
  BtheB, 10
  CHFLS, 11
  clouds, 14
  CYGOB1, 15
  EFT, 16
  epilepsy, 16
  Forbes2000, 17
  foster, 18
  gardenflowers, 19
  GHQ, 19
  heptathlon, 20
  household, 22
  Lanza, 24
  mastectomy, 25
  men1500m, 26
  meteo, 26
  orallesions, 27
  phosphate, 28
  pistonrings, 29
  planets, 29
  plasma, 30
  polyps, 31
  polyps3, 32
  pottery, 33
  rearrests, 34
  respiratory, 35
  roomwidth, 36
  schizophrenia, 37
  schizophrenia2, 38
  schooldays, 39
  skulls, 40
  smoking, 41
  Smoking_DollHill1950, 42
  Smoking_Mueller1940, 42
  Smoking_SchairerSchoeniger1944, 43
  Smoking_Wassink1945, 44
  students, 44
  suicides, 45
  suicides2, 46
  toenail, 46
  toothpaste, 47
  USairpollution, 48
  USmelanoma, 49
  USstates, 50
  UStemp, 50
  voting, 51
  water, 52
  watervoles, 53
  waves, 54
  weightgain, 55
  womensrole, 56
* misc
  HSAURtable, 23

agefat, 3
aspirin, 4
backpain, 5
BCG, 6
birds, 7
birthdeathrates, 8
bladdercancer, 8
bp, 9
BtheB, 10
CHFLS, 11
clouds, 13
CYGOB1, 15
dist, 19
EFT, 16
epilepsy, 16
Forbes2000, 17
foster, 18
gardenflowers, 19
GHQ, 19
heptathlon, 20
household, 22
HSAURtable, 23
Lanza, 24
mastectomy, 25
men1500m, 26
meteo, 26
orallesions, 27
phosphate, 28
pistonrings, 29
planets, 29
plasma, 30
polyps, 31, 33
polyp3, 32
pottery, 33
rearrests, 34
respiratory, 35
roomwidth, 36
schizophrenia, 37
schizophrenia2, 38
schooldays, 39
skulls, 40
smoking, 41
Smoking_DollHill1950, 42
Smoking_Mueller1940, 42
Smoking_SchairerSchoeniger1944, 43
Smoking_Wassink1945, 44
students, 44
suicides, 45
suicides2, 46
table, 27, 29, 34, 45
toenail, 46
toLatex, 23, 24
toLatex.tabtab (HSAURtable), 23
toothpaste, 47
USairpollution, 48
USmelanoma, 49
USstates, 50
USTemp, 50
voting, 51
water, 52
watervoles, 53
waves, 54
weightgain, 55
womensrole, 56