Package 'ISwR'

January 20, 2020

Version 2.0-8

Date 2019-12-30

Title Introductory Statistics with R

Author Peter Dalgaard

Maintainer Peter Dalgaard <pd.mes@cbs.dk>

Description Data sets and scripts for text examples and exercises in
P. Dalgaard (2008), `Introductory Statistics with R', 2nd ed., Springer Verlag, ISBN 978-0387790534.

Depends R (>= 2.6.0)

Suggests survival,MASS

License GPL (>= 2)

ZipData no

LazyData yes

NeedsCompilation no

Repository CRAN

Date/Publication 2020-01-20 09:50:02 UTC

R topics documented:

`alkfos` *Alkaline phosphatase data*

Description

Repeated measurements of alkaline phosphatase in a randomized trial of Tamoxifen treatment of breast cancer patients.

Usage

```
alkfos
```

Format

A data frame with 43 observations on the following 8 variables.

`grp` a numeric vector, group code (1=placebo, 2=Tamoxifen).

`c0` a numeric vector, concentration at baseline.

`c3` a numeric vector, concentration after 3 months.

`c6` a numeric vector, concentration after 6 months.

`c9` a numeric vector, concentration after 9 months.

`c12` a numeric vector, concentration after 12 months.

`c18` a numeric vector, concentration after 18 months.

`c24` a numeric vector, concentration after 24 months.

Source

Original data.

References

B. Kristensen et al. (1994), Tamoxifen and bone metabolism in postmenopausal low-risk breast cancer patients: a randomized study. *Journal of Clinical Oncology*, 12(2):992–997.

`ashina` *Ashina's crossover trial*

Description

The `ashina` data frame has 16 rows and 3 columns. It contains data from a crossover trial for the effect of an NO synthase inhibitor on headaches. Visual analog scale recordings of pain levels were made at baseline and at five time points after infusion of the drug or placebo. A score was calculated as the sum of the differences from baseline. Data were recorded during two sessions for each patient. Six patients were given treatment on the first occasion and the placebo on the second. Ten patients had placebo first and then treatment. The order of treatment and the placebo was randomized.

Usage

```
ashina
```

Format

This data frame contains the following columns:

`vas.active` a numeric vector, summary score when given active substance.

`vas.plac` a numeric vector, summary score when given placebo treatment.

`grp` a numeric vector code, 1: placebo first, 2: active first.

Source

Original data.

References

M.Ashina et al. (1999), Effect of inhibition of nitric oxide synthase on chronic tension-type headache: a randomised crossover trial. *Lancet* 353, 287–289

Examples

```
plot(vas.active~vas.plac,pch=grp,data=ashina)
abline(0,1)
```

---

bcmort *Breast cancer mortality*

Description

Danish study on the effect of screening for breast cancer.

Usage

```
bcmort
```

Format

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

`age` a factor with levels `50-54`, `55-59`, `60-64`, `65-69`, `70-74`, and `75-79`.

`cohort` a factor with levels `Study gr.`, `Nat.ctr.`, `Hist.ctr.`, and `Hist.nat.ctr.`.

`bc.deaths` a numeric vector, number of breast cancer deaths.

`p.yr` a numeric vector, person-years under study.

Details

Four cohorts were collected. The "study group" consists of the population of women in the appropriate age range in Copenhagen and Frederiksberg after the introduction of routine mammography screening. The "national control group" consisted of the population in the parts of Denmark in which routine mammography screening was not available. These two groups were both collected in the years 1991–2001. The "historical control group" and the "historical national control group" are similar cohorts from 10 years earlier (1981–1991), before the introduction of screening in Copenhagen and Frederiksberg. The study group comprises the entire population, not just those accepting the invitation to be screened.

Source

A.H. Olsen et al. (2005), Breast cancer mortality in Copenhagen after introduction of mammography screening. *British Medical Journal*, 330: 220–222.

bp.obese *Obesity and blood pressure*

Description

The bp.obese data frame has 102 rows and 3 columns. It contains data from a random sample of Mexican-American adults in a small California town.

Usage

```
bp.obese
```

Format

This data frame contains the following columns:

sex a numeric vector code, 0: male, 1: female.

obese a numeric vector, ratio of actual weight to ideal weight from New York Metropolitan Life Tables.

bp a numeric vector,systolic blood pressure (mm Hg).

Source

B.W. Brown and M. Hollander (1977), *Statistics: A Biomedical Introduction,* Wiley.

Examples

```
plot(bp~obese,pch = ifelse(sex==1, "F", "M"), data = bp.obese)
```

caesarean *Caesarean section and maternal shoe size*

Description

The table caesar.shoe contains the relation between caesarean section and maternal shoe size (UK sizes!).

Usage

```
caesar.shoe
```

Format

A matrix with two rows and six columns.

Source

D.G. Altman (1991), *Practical Statistics for Medical Research*, Table 10.1, Chapman & Hall.

Examples

```
prop.trend.test(caesar.shoe["Yes",],margin.table(caesar.shoe,2))
```

`coking` *Coking data*

Description

The `coking` data frame has 18 rows and 3 columns. It contains the time to coking in an experiment where the oven width and temperature were varied.

Usage

```
coking
```

Format

This data frame contains the following columns:

`width` a factor with levels `4`, `8`, and `12`, giving the oven width in inches.

`temp` a factor with levels `1600` and `1900`, giving the temperature in Fahrenheit.

`time` a numeric vector, time to coking.

Source

R.A. Johnson (1994), *Miller and Freund's Probability and Statistics for Engineers*, 5th ed., Prentice-Hall.

Examples

```
attach(coking)
matplot(tapply(time,list(width,temp),mean))
detach(coking)
```

cystfibr *Cystic fibrosis lung function data*

Description

The cystfibr data frame has 25 rows and 10 columns. It contains lung function data for cystic fibrosis patients (7–23 years old).

Usage

```
cystfibr
```

Format

This data frame contains the following columns:

age a numeric vector, age in years.

sex a numeric vector code, 0: male, 1:female.

height a numeric vector, height (cm).

weight a numeric vector, weight (kg).

bmp a numeric vector, body mass (% of normal).

fev1 a numeric vector, forced expiratory volume.

rv a numeric vector, residual volume.

frc a numeric vector, functional residual capacity.

tlc a numeric vector, total lung capacity.

pemax a numeric vector, maximum expiratory pressure.

Source

D.G. Altman (1991), *Practical Statistics for Medical Research*, Table 12.11, Chapman & Hall.

References

O'Neill et al. (1983), The effects of chronic hyperinflation, nutritional status, and posture on respiratory muscle strength in cystic fibrosis, *Am. Rev. Respir. Dis.,* 128:1051–1054.

eba1977 *Lung cancer incidence in four Danish cities 1968–1971*

Description

This data set contains counts of incident lung cancer cases and population size in four neighbouring Danish cities by age group.

Usage

```
eba1977
```

Format

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

`city` a factor with levels `Fredericia`, `Horsens`, `Kolding`, and `Vejle`.

`age` a factor with levels `40-54`, `55-59`, `60-64`, `65-69`, `70-74`, and `75+`.

`pop` a numeric vector, number of inhabitants.

`cases` a numeric vector, number of lung cancer cases.

Details

These data were "at the center of public interest in Denmark in 1974", according to Erling Andersen's paper. The city of Fredericia has a substantial petrochemical industry in the harbour area.

Source

E.B. Andersen (1977), Multiplicative Poisson models with unequal cell rates, *Scandinavian Journal of Statistics*, 4:153–158.

References

J. Clemmensen et al. (1974), *Ugeskrift for Læger*, pp. 2260–2268.

energy *Energy expenditure*

Description

The `energy` data frame has 22 rows and 2 columns. It contains data on the energy expenditure in groups of lean and obese women.

Usage

```
energy
```

Format

This data frame contains the following columns:

`expend` a numeric vector, 24 hour energy expenditure (MJ).

`stature` a factor with levels `lean` and `obese`.

Source

D.G. Altman (1991), *Practical Statistics for Medical Research*, Table 9.4, Chapman & Hall.

Examples

```
plot(expend~stature,data=energy)
```

`ewrates` *Rates of lung and nasal cancer mortality, and total mortality.*

Description

England and Wales mortality rates from lung cancer, nasal cancer, and all causes, 1936–1980. The 1936 rates are repeated as 1931 rates in order to accommodate follow-up for the `nickel` study.

Usage

```
ewrates
```

Format

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

`year` calendar period, 1931: 1931–35, 1936: 1936–40, ...

`age` age class, 10: 10–14, 15:15–19, ...

`lung` lung cancer mortality rate per 1 million person-years

`nasal` nasal cancer mortality rate per 1 million person-years

`other` all cause mortality rate per 1 million person-years

Details

Taken from the "Epi" package by Bendix Carstensen et al.

Source

N.E. Breslow, and N. Day (1987). *Statistical Methods in Cancer Research. Volume II: The Design and Analysis of Cohort Studies*, Appendix IX. IARC Scientific Publications, Lyon.

fake.trypsin *Trypsin by age groups*

Description

The trypsin data frame has 271 rows and 3 columns. Serum levels of immunoreactive trypsin in healthy volunteers (faked!).

Usage

```
fake.trypsin
```

Format

This data frame contains the following columns:

trypsin a numeric vector, serum-trypsin in ng/ml.

grp a numeric vector, age coding. See below.

grpf a factor with levels 1: age 10–19, 2: age 20–29, 3: age 30–39, 4: age 40–49, 5: age 50–59, and 6: age 60–69.

Details

Data have been simulated to match given group means and SD.

Source

D.G. Altman (1991), *Practical Statistics for Medical Research*, Table 9.12, Chapman & Hall.

Examples

```
plot(trypsin~grp, data=fake.trypsin)
```

graft.vs.host *Graft versus host disease*

Description

The gvhd data frame has 37 rows and 7 columns. It contains data from patients receiving a nondepleted allogenic bone marrow transplant with the purpose of finding variables associated with the development of acute graft-versus-host disease.

Usage

```
graft.vs.host
```

Format

This data frame contains the following columns:

`pnr` a numeric vector patient number.

`rcpage` a numeric vector, age of recipient (years).

`donage` a numeric vector, age of donor (years).

`type` a numeric vector, type of leukaemia coded 1: AML, 2: ALL, 3: CML for acute myeloid, acute lymphatic, and chronic myeloid leukaemia.

`preg` a numeric vector code indicating whether donor has been pregnant. 0: no, 1: yes.

`index` a numeric vector giving an index of mixed epidermal cell-lymphocyte reactions.

`gvhd` a numeric vector code, graft-versus-host disease, 0: no, 1: yes.

`time` a numeric vector, follow-up time

`dead` a numeric vector code, 0: no (censored), 1: yes

Source

D.G. Altman (1991), *Practical Statistics for Medical Research*, Exercise 12.3, Chapman & Hall.

Examples

```
plot(jitter(gvhd,0.2)~index,data=graft.vs.host)
```

---

heart.rate *Heart rates after enalaprilat*

Description

The `heart.rate` data frame has 36 rows and 3 columns. It contains data for nine patients with congestive heart failure before and shortly after administration of enalaprilat, in a balanced two-way layout.

Usage

```
heart.rate
```

Format

This data frame contains the following columns:

`hr` a numeric vector, heart rate in beats per minute.

`subj` a factor with levels `1` to `9`.

`time` a factor with levels `0` (before), `30`, `60`, and `120` (minutes after administration).

Source

D.G. Altman (1991), *Practical Statistics for Medical Research*, Table 12.2, Chapman & Hall.

Examples

```
evalq(interaction.plot(time,subj,hr), heart.rate)
```

`hellung` *Growth of Tetrahymena cells*

Description

The `hellung` data frame has 51 rows and 3 columns. diameter and concentration of *Tetrahymena* cells with and without glucose added to growth medium.

Usage

```
hellung
```

Format

This data frame contains the following columns:

`glucose` a numeric vector code, 1: yes, 2: no.

`conc` a numeric vector, cell concentration (counts/ml).

`diameter` a numeric vector, cell diameter (μm).

Source

D. Kronborg and L.T. Skovgaard (1990), *Regressionsanalyse*, Table 1.1, FADLs Forlag (in Danish).

Examples

```
plot(diameter~conc,pch=glucose,log="xy",data=hellung)
```

`IgM` *Immunoglobulin G*

Description

Serum IgM in 298 children aged 6 months to 6 years.

Usage

```
IgM
```

Format

A single numeric vector (g/l).

Source

D.G. Altman (1991), *Practical Statistics for Medical Research*, Table 3.2, Chapman & Hall.

Examples

```
stripchart(IgM,method="stack")
```

`intake` *Energy intake*

Description

The `intake` data frame has 11 rows and 2 columns. It contains paired values of energy intake for 11 women.

Usage

```
intake
```

Format

This data frame contains the following columns:

`pre` a numeric vector, premenstrual intake (kJ).

`post` a numeric vector, postmenstrual intake (kJ).

Source

D.G. Altman (1991), *Practical Statistics for Medical Research*, Table 9.3, Chapman & Hall.

Examples

```
plot(intake$pre, intake$post)
```

juul *Juul's IGF data*

Description

The juul data frame has 1339 rows and 6 columns. It contains a reference sample of the distribution of insulin-like growth factor (IGF-I), one observation per subject in various ages, with the bulk of the data collected in connection with school physical examinations.

Usage

```
juul
```

Format

This data frame contains the following columns:

age a numeric vector (years).

menarche a numeric vector. Has menarche occurred (code 1: no, 2: yes)?

sex a numeric vector (1: boy, 2: girl).

igf1 a numeric vector, insulin-like growth factor (μg/l).

tanner a numeric vector, codes 1–5: Stages of puberty ad modum Tanner.

testvol a numeric vector, testicular volume (ml).

Source

Original data.

Examples

```
plot(igf1~age, data=juul)
```

juul2 *Juul's IGF data, extended version*

Description

The juul2 data frame has 1339 rows and 8 columns; extended version of ljuull.

Usage

```
juul2
```

Format

This data frame contains the following columns:

`age` a numeric vector (years).

`height` a numeric vector (cm).

`menarche` a numeric vector. Has menarche occurred (code 1: no, 2: yes)?

`sex` a numeric vector (1: boy, 2: girl).

`igf1` a numeric vector, insulin-like growth factor (μg/l).

`tanner` a numeric vector, codes 1–5: Stages of puberty ad modum Tanner.

`testvol` a numeric vector, testicular volume (ml).

`weight` a numeric vector, weight (kg).

Source

Original data.

Examples

```
plot(igf1~age, data=juul2)
```

---

kfm *Breast-feeding data*

Description

The `kfm` data frame has 50 rows and 7 columns. It was collected by Kim Fleischer Michaelsen and contains data for 50 infants of age approximately 2 months. They were weighed immediately before and after each breast feeding. and the measured intake of breast milk was registered along with various other data.

Usage

```
kfm
```

Format

This data frame contains the following columns:

`no` a numeric vector, identification number.

`dl.milk` a numeric vector, breast-milk intake (dl/24h).

`sex` a factor with levels `boy` and `girl`.

`weight` a numeric vector, weight of child (kg).

`ml.suppl` a numeric vector, supplementary milk substitute (ml/24h).

`mat.weight` a numeric vector, weight of mother (kg).

`mat.height` a numeric vector, height of mother (cm).

Note

The amount of supplementary milk substitute refers to a period before the data collection.

Source

Original data.

Examples

```
plot(dl.milk~mat.height,pch=c(1,2)[sex],data=kfm)
```

`lung` *Methods for determining lung volume*

Description

The `lung` data frame has 18 rows and 3 columns. It contains data on three different methods of determining human lung volume.

Usage

```
lung
```

Format

This data frame contains the following columns:

`volume` a numeric vector, measured lung volume.

`method` a factor with levels `A`, `B`, and `C`.

`subject` a factor with levels `1`–`6`.

Source

Anon. (1977), Exercises in Applied Statistics, Exercise 4.15, Dept.\ of Theoretical Statistics, Aarhus University.

malaria *Malaria antibody data*

Description

The malaria data frame has 100 rows and 4 columns.

Usage

```
malaria
```

Format

This data frame contains the following columns:

subject subject code.

age age in years.

ab antibody level.

mal a numeric vector code, Malaria: 0: no, 1: yes.

Details

A random sample of 100 children aged 3–15 years from a village in Ghana. The children were followed for a period of 8 months. At the beginning of the study, values of a particular antibody were assessed. Based on observations during the study period, the children were categorized into two groups: individuals with and without symptoms of malaria.

Source

Unpublished data.

Examples

```
summary(malaria)
```

melanom *Survival after malignant melanoma*

Description

The melanom data frame has 205 rows and 7 columns. It contains data relating to the survival of patients after an operation for malignant melanoma, collected at Odense University Hospital by K.T. Drzewiecki.

Usage

```
melanom
```

Format

This data frame contains the following columns:

`no` a numeric vector, patient code.

`status` a numeric vector code, survival status; 1: dead from melanoma, 2: alive, 3: dead from other cause.

`days` a numeric vector, observation time.

`ulc` a numeric vector code, ulceration; 1: present, 2: absent.

`thick` a numeric vector, tumor thickness (1/100 mm).

`sex` a numeric vector code; 1: female, 2: male.

Source

P.K. Andersen, Ø. Borgan, R.D. Gill, and N. Keiding (1991), *Statistical Models Based on Counting Processes*, Appendix 1, Springer-Verlag.

Examples

```
require(survival)
plot(survfit(Surv(days,status==1)~1,data=melanom))
```

---

nickel *Nickel smelters in South Wales*

Description

The data concern a cohort of nickel smelting workers in South Wales, with information on exposure, follow-up period, and cause of death.

Usage

```
nickel
```

Format

A data frame containing 679 observations of the following 7 variables:

`id` subject identifier (numeric).

`icd` ICD cause of death if dead, 0 otherwise (numeric).

`exposure` exposure index for workplace (numeric)

`dob` date of birth (numeric).

`age1st` age at first exposure (numeric).

`agein` age at start of follow-up (numeric).

`ageout` age at end of follow-up (numeric).

Details

Taken from the "Epi" package by Bendix Carstensen et al. For comparison purposes, England and Wales mortality rates (per 1,000,000 per annum) from lung cancer (ICDs 162 and 163), nasal cancer (ICD 160), and all causes, by age group and calendar period, are supplied in the data set ewrates.

Source

N.E. Breslow and N. Day (1987). *Statistical Methods in Cancer Research. Volume II: The Design and Analysis of Cohort Studies*, IARC Scientific Publications, Lyon.

`nickel.expand` *Nickel smelters in South Wales, expanded*

Description

The data concern a cohort of nickel smelting workers in South Wales, with information on exposure, follow-up period, and cause of death, as in the nickel data. This version has follow-up times split according to age groups and is merged with the mortality rates in ewrates.

Usage

```
nickel.expand
```

Format

A data frame with 3724 observations on the following 12 variables:

`agr` age class: 10: 10–14, 15: 15–19, ...

`ygr` calendar period, 1931: 1931–35, 1936: 1936–40, ...

`id` subject identifier (numeric).

`icd` ICD cause of death if dead, 0 otherwise (numeric).

`exposure` exposure index for workplace (numeric).

`dob` date of birth (numeric).

`age1st` age at first exposure (numeric).

`agein` age at start of follow-up (numeric).

`ageout` age at end of follow-up (numeric).

`lung` lung cancer mortality rate per 1 million person-years.

`nasal` nasal cancer mortality rate per 1 million person-years.

`other` all cause mortality rate per 1 million person-years.

Source

Computed from nickel and ewrates data sets.

philion *Dose response data*

Description

Four small experiments with the purpose of estimating the EC50 of a biological dose-response relation.

Usage

```
philion
```

Format

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

`experiment` a numeric vector; codes 1 through 4 denote the experiment number.

`dose` a numeric vector, the dose.

`response` a numeric vector, the response (counts).

Details

These data were discussed on the R mailing lists, initially suggesting a log-linear Poisson regression, but actually a relation like $y = y_{\max}/(1 + (x/\beta)^{\alpha})$ is more suitable.

Source

Original data from Vincent Philion, IRDA, Qu\'ebec.

References

http://tolstoy.newcastle.edu.au/R/help/03b/1121.html

react *Tuberculin reactions*

Description

The numeric vector `react` contains differences between two nurses' determinations of 334 tuberculin reaction sizes.

Usage

```
react
```

Format

A single vector, differences between reaction sizes in mm.

Source

Anon. (1977), *Exercises in Applied Statistics*, Exercise 2.9, Dept.\ of Theoretical Statistics, Aarhus University.

Examples

```
hist(react) # not good because of discretization effects...
plot(density(react))
```

---

red.cell.folate *Red cell folate data*

Description

The `folate` data frame has 22 rows and 2 columns. It contains data on red cell folate levels in patients receiving three different methods of ventilation during anesthesia.

Usage

```
red.cell.folate
```

Format

This data frame contains the following columns:

`folate` a numeric vector, folate concentration (μg/l).

`ventilation` a factor with levels `N2O+O2,24h`: 50% nitrous oxide and 50% oxygen, continuously for 24 hours; `N2O+O2,op`: 50% nitrous oxide and 50% oxygen, only during operation; `O2,24h`: no nitrous oxide but 35%–50% oxygen for 24 hours.

Source

D.G. Altman (1991), *Practical Statistics for Medical Research*, Table 9.10, Chapman & Hall.

Examples

```
plot(folate~ventilation,data=red.cell.folate)
```

rmr *Resting metabolic rate*

Description

The rmr data frame has 44 rows and 2 columns. It contains the resting metabolic rate and body weight data for 44 women.

Usage

```
rmr
```

Format

This data frame contains the following columns:

body.weight a numeric vector, body weight (kg).

metabolic.rate a numeric vector, metabolic rate (kcal/24hr).

Source

D.G. Altman (1991), *Practical Statistics for Medical Research*, Exercise 11.2, Chapman & Hall.

Examples

```
plot(metabolic.rate~body.weight,data=rmr)
```

secher *Birth weight and ultrasonography*

Description

The secher data frame has 107 rows and 4 columns. It contains ultrasonographic measurements of fetuses immediately before birth and their subsequent birth weight.

Usage

```
secher
```

Format

This data frame contains the following columns:

bwt a numeric vector, birth weight (g).

bpd a numeric vector, biparietal diameter (mm).

ad a numeric vector, abdominal diameter (mm).

no a numeric vector, observation number.

Source

D. Kronborg and L.T. Skovgaard (1990), *Regressionsanalyse*, Table 3.1, FADLs Forlag (in Danish).

Secher et al. (1987), European Journal of Obstetrics, Gynecology, and Reproductive Biology, 24: 1–11.

Examples

```
plot(bwt~ad, data=secher, log="xy")
```

`secretin` *Secretin-induced blood glucose changes*

Description

The `secretin` data frame has 50 rows and 6 columns. It contains data from a glucose response experiment.

Usage

```
secretin
```

Format

This data frame contains the following columns:

`gluc` a numeric vector, blood glucose level.

`person` a factor with levels `A`–`E`.

`time` a factor with levels `20`, `30`, `60`, `90` (minutes since injection), and `pre` (before injection).

`repl` a factor with levels `a`: 1st sample; `b`: 2nd sample.

`time20plus` a factor with levels `20+`: 20 minutes or longer since injection; `pre`: before injection.

`time.comb` a factor with levels `20`: 20 minutes since injection; `30+`: 30 minutes or longer since injection; `pre`: before injection.

Details

Secretin is a hormone of the duodenal mucous membrane. An extract was administered to five patients with arterial hypertension. Primary registrations (double determination) of blood glucose were on graph paper and later quantified with the smallest of the two measurements recorded first.

Source

Anon. (1977), *Exercises in Applied Statistics*, Exercise 5.8, Dept.\ of Theoretical Statistics, Aarhus University.

`stroke` *Estonian stroke data*

Description

All cases of stroke in Tartu, Estonia, during the period 1991–1993, with follow-up until January 1, 1996.

Usage

```
stroke
```

Format

A data frame with 829 observations on the following 10 variables.

`sex` a factor with levels `Female` and `Male`.

`died` a Date, date of death.

`dstr` a Date, date of stroke.

`age` a numeric vector, age at stroke.

`dgn` a factor, diagnosis, with levels `ICH` (intracranial haemorrhage), `ID` (unidentified). `INF` (infarction, ischaemic), `SAH` (subarchnoid haemorrhage).

`coma` a factor with levels `No` and `Yes`, indicating whether patient was in coma after the stroke.

`diab` a factor with levels `No` and `Yes`, history of diabetes.

`minf` a factor with levels `No` and `Yes`, history of myocardial infarction.

`han` a factor with levels `No` and `Yes`, history of hypertension.

`obsmonths` a numeric vector, observation times in months (set to 0.1 for patients dying on the same day as the stroke).

`dead` a logical vector, whether patient died during the study.

Source

Original data.

References

J. Korv, M. Roose, and A.E. Kaasik (1997). Stroke Registry of Tartu, Estonia, from 1991 through 1993. Cerebrovascular Disorders 7:154–162.

tb.dilute *Tuberculin dilution assay*

Description

The `tb.dilute` data frame has 18 rows and 3 columns. It contains data from a drug test involving dilutions of tuberculin.

Usage

```
tb.dilute
```

Format

This data frame contains the following columns:

`reaction` a numeric vector, reaction sizes (average of diameters) for tuberculin skin pricks.

`animal` a factor with levels `1`–`6`.

`logdose` a factor with levels `0.5`, `0`, and `-0.5`.

Details

The actual dilutions were 1:100, $1 : 100\sqrt{10}$, 1:1000. Setting the middle one to 1 and using base-10 logarithms gives the `logdose` values.

Source

Anon. (1977), *Exercises in Applied Statistics*, part of Exercise 4.15, Dept.\ of Theoretical Statistics, Aarhus University.

thuesen *Ventricular shortening velocity*

Description

The `thuesen` data frame has 24 rows and 2 columns. It contains ventricular shortening velocity and blood glucose for type 1 diabetic patients.

Usage

```
thuesen
```

Format

This data frame contains the following columns:

`blood.glucose` a numeric vector, fasting blood glucose (mmol/l).

`short.velocity` a numeric vector, mean circumferential shortening velocity (%/s).

Source

D.G. Altman (1991), *Practical Statistics for Medical Research*, Table 11.6, Chapman & Hall.

Examples

```
plot(short.velocity~blood.glucose, data=thuesen)
```

---

tlc *Total lung capacity*

Description

The `tlc` data frame has 32 rows and 4 columns. It contains data on pretransplant total lung capacity (TLC) for recipients of heart-lung transplants by whole-body plethysmography.

Usage

```
tlc
```

Format

This data frame contains the following columns:

`age` a numeric vector, age of recipient (years).

`sex` a numeric vector code, female: 1, male: 2.

`height` a numeric vector, height of recipient (cm).

`tlc` a numeric vector, total lung capacity (l).

Source

D.G. Altman (1991), *Practical Statistics for Medical Research*, Exercise 12.5, 10.1, Chapman & Hall.

Examples

```
plot(tlc~height,data=tlc)
```

vitcap *Vital capacity*

Description

The vitcap data frame has 24 rows and 3 columns. It contains data on vital capacity for workers in the cadmium industry. It is a subset of the vitcap2 data set.

Usage

```
vitcap
```

Format

This data frame contains the following columns:

group a numeric vector; group codes are 1: exposed > 10 years, 3: not exposed.

age a numeric vector, age in years.

vital.capacity a numeric vector, vital capacity (a measure of lung volume) in liters.

Source

P. Armitage and G. Berry (1987), *Statistical Methods in Medical Research*, 2nd ed., Blackwell, p.286.

Examples

```
plot(vital.capacity~age, pch=group, data=vitcap)
```

vitcap2 *Vital capacity, full data set*

Description

The vitcap2 data frame has 84 rows and 3 columns. Age and vital capacity for workers in the cadmium industry.

Usage

```
vitcap2
```

Format

This data frame contains the following columns:

`group` a numeric vector; group codes are 1: exposed > 10 years, 2: exposed < 10 years, 3: not exposed.

`age` a numeric vector, age in years.

`vital.capacity` a numeric vector, vital capacity (a measure of lung volume) (l).

Source

P. Armitage and G. Berry (1987), *Statistical Methods in Medical Research*, 2nd ed., Blackwell, p.286.

Examples

```
plot(vital.capacity~age, pch=group, data=vitcap2)
```

`wright` *Comparison of Wright peak-flow meters*

Description

The `wright` data frame has 17 rows and 2 columns. It contains data on peak expiratory flow rate with two different flow meters on each of 17 subjects.

Usage

```
wright
```

Format

This data frame contains the following columns:

`std.wright` a numeric vector, data from large flow meter (l/min).

`mini.wright` a numeric vector, data from mini flow meter (l/min).

Source

J.M. Bland and D.G. Altman (1986), Statistical methods for assessing agreement between two methods of clinical measurement, *Lancet*, 1:307–310.

Examples

```
plot(wright)
abline(0,1)
```

`zelazo` *Age at walking*

Description

The `zelazo` object is a list with four components.

Usage

```
zelazo
```

Format

This is a list containing data on age at walking (in months) for four groups of infants:

`active` test group receiving active training; these children had their walking and placing reflexes trained during four three-minute sessions that took place every day from their second to their eighth week of life.

`passive` passive training group; these children received the same types of social and gross motor stimulation, but did not have their specific walking and placing reflexes trained.

`none` no training; these children had no special training, but were tested along with the children who underwent active or passive training.

`ctr.8w` eighth-week controls; these children had no training and were only tested at the age of 8 weeks.

Note

When asked to enter these data from a text source, many students will use one vector per group and will need to reformat data into a data frame for some uses. The rather unusual format of this data set mimics that situation.

Source

P.R. Zelazo, N.A. Zelazo, and S. Kolb (1972), "Walking" in the newborn, *Science*, 176: 314–315.

Index