Package 'SenSrivastava'

June 26, 2015

Version 2015.6.25

Date 2015-06-25

Title Datasets from Sen & Srivastava

Author Kjetil B Halvorsen <kjetil1001@gmail.com>

Maintainer Kjetil B Halvorsen <kjetil1001@gmail.com>

Depends R (>= 2.5.0), stats

Suggests leaps, car

LazyData TRUE

Description Collection of datasets from Sen & Srivastava: ``Regression Analysis, Theory, Methods and Applications'', Springer. Sources for individual data files are more fully documented in the book.

License GPL (>= 2)

Encoding latin1

Repository CRAN

NeedsCompilation no

Date/Publication 2015-06-26 10:49:58

R topics documented:

E1.1 *Data on density of vehicles and average speed*

Description

The E1.1 data frame has 24 rows and 2 columns.

Usage

```
data(E1.1)
```

Format

This data frame contains the following columns:

DENSITY a numeric vector, vehicles per mile.

SPEED a numeric vector, miles per hour.

Details

Example 1.1 page 2 in Sen and Srivastava.

Source

Huber, M.J (1957) Effect of temporary bridge on Parkway performance. *Highway Research Board Bulletin* **167** 63–74.

Examples

```
data(E1.1)
attach(E1.1)
plot(DENSITY, sqrt(SPEED))
E1.1.m1 <- lm(sqrt(SPEED) ~ DENSITY + I(DENSITY^2), data=E1.1)
summary(E1.1.m1)
```

E1.11 *Data on violent and property crimes in 22 metropolitan areas of the U.S.*

Description

The E1.11 data frame has 23 rows and 4 columns.

Usage

```
data(E1.11)
```

Format

This data frame contains the following columns:

Metro.Area a character vector, names and state of each metropolitan area.

Violent.Crimes a numeric vector, units of measurement not given.

Property.Crimes a numeric vector, units of measurement not given.

Population a numeric vector, in thousands.

Source

Dacey, M.F.(1983) *Social science Theories and Methods I: Models of data*, Evanston: Northwestern University.

Examples

```
data(E1.11)
attach(E1.11)
plot(Population, Violent.Crimes)
detach()
```

E1.15   *Stevens Experiment to compare notes against a standard (80 Db)*

Description

The E1.15 data frame has 10 rows and 3 columns. Stevens (1956) asked a number of persons to compare notes of various decibel levels against a standard (80 decibels) and to assign them a loudness rating with the standard note being a 10. logy is the response variable and x the stimulus.

Usage

```
data(E1.15)
```

Format

This data frame contains the following columns:

x a numeric vector, the stimulus.

y a numeric vector, the median response at x

logy a numeric vector, the log of y.

Source

Dacey,M.F. (1983) *Social science Theories and Methods I: Models of Data* Evanston: Northwestern University, fromStevens (1956).

Examples

```
data(E1.15)
attach(E1.15)
plot(x, logy)
abline(lm( logy ~ x, data=E1.15))
detach()
```

E1.16   *Earnings and Prices of Selected Paper Company Stocks*

Description

The E1.16 data frame has 10 rows and 3 columns.

Usage

```
data(E1.16)
```

Format

This data frame contains the following columns:

Company a character vector, name of the company

Earn.Share a numeric vector, 1972 earnings per share, in dollars.

Price.Share a numeric vector, prive per share, in dollars, in may, 1973.

Source

Dacey (1983, ch 1) from Moodys's Stock Survey, June 4, 1973, p 610.

Examples

```
with(E1.16, plot(Price.Share, Earn.Share))
```

E1.17   *Data on Population Density and Vehicle Thefts*

Description

The E1.17 data frame has 18 rows and 3 columns.

Usage

```
data(E1.17)
```

Format

This data frame contains the following columns:

D a numeric vector, district of Chicago. 1 is downtown Chicago.

pd a numeric vector, population density of each district.

vtt a numeric vector, vehicle thefts per thousand residents.

Source

Mark Buslik, Chicago Police Department.

Examples

```
data(E1.17)
attach(E1.17)
plot(pd, vtt)
cat("Use the mouse to identify the outlier in the plot (click on the outlier)\n")
## Not run: identify(pd, vtt)
```

E1.18   *Data on Simsbury Marriages*

Description

The E1.18 data frame has 8 rows and 3 columns with data on the number of marriages (ma) that occurred between residents of each of 8 annular zones and residents of Simsbury, Connecticut, for the period 1930–39. The number of residents of each zone is given as pop and the midpoint of distance between Simsbury and the band is given as d.

Usage

```
data(E1.18)
```

Format

This data frame contains the following columns:

d a numeric vector, distance between Simsbury and midpoint of annular zone.

pop a numeric vector, population of annular zone.

ma a numeric vector, number of marriages.

Source

Dacey (1983, ch 4) from Ellsworth (1948).

Examples

```
data(E1.18)
summary(E1.18)
```

E1.19   *Data on Book Prices, Pages and Type of Binding*

Description

The E1.19 data frame has 20 rows and 3 columns. Compiled from the catalog of one publisher of American Government books.

Usage

```
data(E1.19)
```

Format

This data frame contains the following columns:

Price a numeric vector, price of book.

P a numeric vector, number of pages of book.

B a factor with levels c p, c is cloth and p is paperback.

Source

Compiled by one of the authors.

Examples

```
data(E1.19)
summary(E1.19)
```

E1.20   *Data on Physical Quality of Life Index (PQLI) Scores and Infant Mortality Rates (IMR) for Selected Indian States*

Description

The E1.20 data frame has 13 rows and 7 columns.

Usage

```
data(E1.20)
```

Format

This data frame contains the following columns:

State a character vector, name of state.

PQLI a numeric vector, Physical Quality of Life Index, a measure of average wealth.

Comb.IMR a numeric vector, combined infant mortality rate.

Rur.M.IMR a numeric vector, rural male infant mortality rate.

Rur.F.IMR a numeric vector, rural female infant mortality rate.

Urb.M.IMR a numeric vector, urban male infant mortality rate.

Urb.F.IMR a numeric vector, urban female infant mortality rate.

Source

Dr. T.N.K.Raju, Department of Neonatology, University of Illinois at Chicago.

Examples

```
data(E1.20)
## Some data reorganization before analysis:
## Maybe reshape could have been used here?
 e1.20 <- data.frame(rbind(as.matrix(E1.20[,c(2,4)]),
                           as.matrix(E1.20[,c(2,5)]),
                           as.matrix(E1.20[,c(2,6)]),
                           as.matrix(E1.20[,c(2,7)])),row.names=1:52)
  attr(e1.20,"names")[[2]] <- "IMR"
 e1.20$Female <- c(rep(0,13), rep(1,13),rep(0,13),rep(1,13))
 e1.20$Urban  <- c(rep(0,26),rep(1,26))
## Now the analysis can start.
summary(e1.20)
```

E1.21   *Data on Loads and Deformation of a Bar*

Description

The E1.21 data frame has 24 rows and 2 columns. Data are on loads, in pounds, and corresponding deformation, in inches, of a mild steel bar, of length 8 inches and average diameter .564 inches.

Usage

```
data(E1.21)
```

Format

This data frame contains the following columns:

L a numeric vector, load, in pounds.

D a numeric vector, corresponding deformation, in inches.

Source

M.R. Khavanin, Department of Mechanical Engineering, University of Illinois at Chicago.

Examples

```
data(E1.21)
attach(E1.21)
plot(L, D)
detach()
```

E1.7   *Data on Population and Number of Telephones*

Description

The `E1.7` data frame has 6 rows and 2 columns. The relation between population and number of telephones have been used to estimate the population in non-census years.

Usage

```
data(E1.7)
```

Format

This data frame contains the following columns:

RES a numeric vector, number of residents.

MAINS a numeric vector, number of telephones.

Source

Prof. Edwin Thomas, Department of Geography, University of Illinois at Chicago.

Examples

```
data(E1.7)
attach(E1.7)
plot(RES, MAINS)
plot(sqrt(RES), sqrt(MAINS))
```

Description

The `E10.1` data frame has 10 rows and 5 columns. The responses were obtained by adding a N(0, 0.01) pseudorandom variate to `x.1+0.5x.2`. The data were made up by the authors.

Usage

```
data(E10.1)
```

Format

This data frame contains the following columns:

x.1 a numeric vector, predictor 1.

x.2 a numeric vector, predictor 2.

y.1 a numeric vector, response 1.

y.2 a numeric vector, response 2.

y.3 a numeric vector, response 3.

Source

The data were made up by the authors.

Examples

```
data(E10.1)
attach(E10.1)
plot(x.1, x.2)
names(E10.1)
hascar <- require(car)
if (hascar) {
   mod <- lm(y.1 ~ x.1+x.2, data=E10.1)
   vif(mod)
}
```

E10.11 *Longley's Data*

Description

The E10.11 data frame has 16 rows and 7 columns. This is a selection of Longley's multicollinear data (1967).

Usage

```
data(E10.11)
```

Format

This data frame contains the following columns:

Def a numeric vector, a price index.

GNP a numeric vector, gross national product.

Unemp a numeric vector, unemployment rate.

AF a numeric vector, employment in the armed forces.

Pop. a numeric vector, noninstitutional population.

Year a numeric vector, the year.

Total a numeric vector, the response, total employment.

Source

Reproduced from the *Journal of the American Statistical Association*, **62**.

Examples

```
data(E10.11)
summary(E10.11)
plot(E10.11)
```

E10.3 *Supervisor Rating Data*

Description

The E10.3 data frame has 30 rows and 6 columns. This is part of a larger data set gathered for other purposes. The six variables are each composites obtained from responses to a questionare. The dependent variable y is a composite of responses towards thr respondent's supervisor and on job satisfaction. The highest possible score is 20. The predictor variables are defined below.

Usage

```
data(E10.3)
```

Format

This data frame contains the following columns:

x.1 a numeric vector, measures the level of social contact each respondent felt he or she had with the supervisor, was based on questions like "Do you see your supervisor outside of your work place?"

x.2 a numeric vector, measures the perceived level of interest from the supervisor in the employees personal life. Based on questions like "Would you discuss a personal problem with your supervisor?"

x.3 a numeric vector, measures the level of support the employee feels from the supervisor. Based on questions like "Is your supervisor supportive of your work?"

x.4 a numeric vector, together with x.5 measure the drive of the supervisor. Based on the emplotees perception of this drive.

x.5 a numeric vector, based on questions like "Does your supervisor encourage you to learn new skills?"

y a numeric vector, the response.

Source

Sen and Srivastava (1990) *Regression Analysis, Theory, Methods and Applications*. Springer-verlag.

Examples

```
data(E10.3)
summary(E10.3)
plot(E10.3)
```

E11.1   *Artificially Created Data for an Example on Variable Search*

Description

The E11.1 data frame has 20 rows and 5 columns.

Usage

```
data(E11.1)
```

Format

This data frame contains the following columns:

x.1 a numeric vector, predictor 1.

x.2 a numeric vector, predictor 2.

x.3 a numeric vector, predictor 3.

x.4 a numeric vector, predictor 4.

y a numeric vector, response.

Source

Data made up by the authors.

Examples

```
data(E11.1)
exleaps <- require("leaps", quietly=TRUE)
if (exleaps) {
   E11.1.m1 <- regsubsets(y ~x.1+x.2+x.3+x.4, data=E11.1)
   summary(E11.1.m1)
   plot(E11.1.m1)
}
```

E2.1   *Data on Grade Point Average and SAT Scores*

Description

The E2.1 data frame has 9 rows and 3 columns.

Usage

```
data(E2.1)
```

Format

This data frame contains the following columns:

GPA a numeric vector, grade point average (maximum=4)

SATV a numeric vector, SAT verbal score.

SATM a numeric vector, SAT mathematical score.

Source

Dacey (1983).

Examples

```
data(E2.1)
summary(E2.1)
```

E2.11   *Demographic Data for the 50 States of the U.S.*

Description

The E2.11 data frame has 50 rows and 27 columns, this combines exhibits E2.10 and E2.11 in the book. The data are for 1980 except as noted.

Usage

```
data(E2.11)
```

Format

This data frame contains the following columns:

State a character vector, two-letter state code.

POP a numeric vector, total population (1000's).

UR a numeric vector, per mil of population living in urban areas.

MV a numeric vector, per mil who moved between 1965 and 1970.

BL a numeric vector, number of blacks (1000's).

SP a numeric vector, number of spanish speaking (1000's)

AI a numeric vector, number of native americans (100's).

IN a numeric vector, number of inmates of all institutions (correctional, mental, TB, etc) in 1970, (1000's).

PR a numeric vector, number of inmates of correctional institutions in 1970 (100's)

MH a numeric vector, Homes and schools for the mentally handicapped (100's)

B a numeric vector, births per thousand.

HT a numeric vector, death rate from hearth disease per 100000 residents.

S a numeric vector, suicide rate, 1978, per 100000.

DI a numeric vector, death rate from diabetes, 1978, per 100000.

MA a numeric vector, marriage rate, per 10000.

D a numeric vector, divorce rate, per 10000.

DR a numeric vector, physicians per 100000.

DN a numeric vector, dentists per 100000.

HS a numeric vector, per mil high school grads.

CR a numeric vector, crime rate per 100000 population.

M a numeric vector, murder rate oer 100000 population.

PI a numeric vector, prison rate (federal and state) per 100000 residents.

RP a numeric vector,

VT a numeric vector,

PH a numeric vector, telephjones per 100 (1979).

INC a numeric vector, per capita income in 1972 dollars.

PL a numeric vector, per mil of population below poverty label.

Source

Compiled by Prof. Siim Soot, Department of Geography, University of Illinois at Chicago, from *Statistical Abstract of the United States*, 1981, *U.S. Bureau of the Census*, Washington, D.C.

Examples

```
data(E2.11)
summary(E2.11)
```

E2.2   *Data on House Prices*

Description

The `E2.2` data frame has 26 rows and 14 columns, data on house prices in different zones of Chicago.

Usage

```
data(E2.2)
```

Format

This data frame contains the following columns:

Price a numeric vector, selling price of house in thousands of dollars.

BDR a numeric vector, number of bedrooms.

FLR a numeric vector, floor space in sq. feet.

FP a numeric vector, number of fireplaces.

RMS a numeric vector, number of rooms.

ST a numeric vector, storm windows (1 present, 0 absent).

LOT a numeric vector, front footage of lot in feet.

TAX a numeric vector, annual taxes.

BTH a numeric vector, number of bathrooms.

CON a numeric vector, construction (0 if frame, 1 if brick).

GAR a numeric vector, garage size (0=no garage, 10 1 auto garage, etc.).

CDN a numeric vector, condition (1=needs work, 0 otherwise).

L1 a numeric vector, indicator for zone A.

L2 a numeric vector, indicator for zone B.

Source

Ms. Terry Tasch of Long-Kogan Realty, Chicago.

Examples

```
data(E2.2)
summary(E2.2)
```

E2.4   *International Car Ownership Data*

Description

The E2.4 data frame has 24 rows and 8 columns, all data are for 1978.

Usage

```
data(E2.4)
```

Format

This data frame contains the following columns:

Country a character vector, name of each country.

AO a numeric vector, cars per person.

POP a numeric vector, population of country in millions.

DEN a numeric vector, population density.

GDP a numeric vector, per capita income in U.S. dollars.

PR a numeric vector, gasoline price in U.S. cents per liter.

CON a numeric vector, Tonnes of gasoline consumed per car per year.

TR a numeric vector, thousands of passenger-kilometers per person of bus and rail use.

Details

Develop a model with A0 as the response variable.

Source

OECD (1982)

Examples

```
data(E2.4)
summary(E2.4)
```

E2.6   *Voltage Data*

Description

The `E2.6` data frame has 10 rows and 2 columns.

Usage

```
data(E2.6)
```

Format

This data frame contains the following columns:

V.a a numeric vector, actual voltage.

V.c a numeric vector, voltage computed from the measured power outout (using light output from electronic flash).

Details

A definition of efficiency is the ratio `V.c`/`V.a`. Obtain a model for efficiency E as a regresion in `V.a`. Use a quadratic polynomial. Examine the fit.

Source

Armin Lehning, Speedotron Corporation.

Examples

```
data(E2.6)
E2.6.m1 <- lm(V.c/V.a ~ V.a + I(V.a^2), data=E2.6)
plot(E2.6.m1)
```

E2.7   *Korean Auto Ownership Data*

Description

The E2.7 data frame has 10 rows and 5 columns.

Usage

```
data(E2.7)
```

Format

This data frame contains the following columns:

Year a numeric vector, year.

AO a numeric vector, number of cars per person.

GNP a numeric vector, per capita GNP in 1000 korean Wons.

CP a numeric vector, average car price in 1000 korean Wons.

OP a numeric vector, gasoline price after taxes, in wons per liter.

Source

KRIHS, (1985) *Study of Road User Charges*. Seoul: Korea Research Institute for Human Settlements.

Examples

```
data(E2.7)
summary(E2.7)
```

E2.8   *Data on per Capita Output of Workers in Shanghai*

Description

The E2.8 data frame has 17 rows and 4 columns, data for 17 factories in Shanghai.

Usage

```
data(E2.8)
```

Format

This data frame contains the following columns:

Output a numeric vector, per capita output in Chinese Yuan.

SI a numeric vector, number of workers in the factory.

SP a numeric vector, land area of the factory in sq. meters per worker.

I a numeric vector, investments in Yuan per worker.

Source

Prof. Zhang Tingwei of Tongji University, Shanghai.

Examples

```
data(E2.8)
summary(E2.8)
```

E2.9   *Data on Capital, Labour and Value Added for Three Sectors*

Description

The E2.9 data frame has 15 rows and 10 columns. The three sectors are "20": Food and kindred products, "36": Equipment and supplies and "37": Transportation equipment.

Usage

```
data(E2.9)
```

Format

This data frame contains the following columns:

YEAR a numeric vector, year without first two digits "19".

Cap.20 a numeric vector, capital of sector 20.

Cap.36 a numeric vector, capital of sector 36.

Cap.37 a numeric vector, capital of sector 37.

Lab.20 a numeric vector, labour of sector 20.

Lab.36 a numeric vector, labour of sector 36.

Lab.37 a numeric vector, labour of sector 37.

Val.20 a numeric vector, real value added of sector 20.

Val.36 a numeric vector, real value added of sector 36.

Val.37 a numeric vector, real value added of sector 37.

Source

Dr. Phillip Israelovich of the Federal Reserve Bank.

Examples

```
data(E2.9)
summary(E2.9)
```

E3.4   *Men's Worlds Record Times for Running and Corresponding Distances*

Description

The E3.4 data frame has 13 rows and 2 columns. World record times as of 1974.

Usage

```
data(E3.4)
```

Format

This data frame contains the following columns:

Dist. a numeric vector, distance in meters.

Time a numeric vector, time in seconds.

Source

Encyclopædia Britannica, 15th Edition, 1974, Micropædia, IX, page 485.

See Also

E3.5, the records for women.

Examples

```
data(E3.4)
summary(E3.4)
```

E3.5   *Women's World Record Times for Running and Corresponding Distances*

Description

The E3.5 data frame has 6 rows and 2 columns. Records are for 1974.

Usage

```
data(E3.5)
```

Format

This data frame contains the following columns:

Dist. a numeric vector, distance run, in meters.

Time a numeric vector, time used, in seconds.

Source

Encyclopædia Britannica, 15th Edition, 1974, Micropædia, IX, page 487.

See Also

E3.4, for the men's records.

Examples

```
data(E3.5)
data(E3.4)
summary(E3.5)
summary(E3.4)
records <- rbind(E3.5,E3.4)
sex <- factor(c(rep("F", 6), rep("M", 13)))
records$sex <- sex
summary(records)
```

E3.6   *Data on Corporations and Corporation Chairmen*

Description

The E3.6 data frame has 50 rows and 6 columns.

Usage

    data(E3.6)

Format

This data frame contains the following columns:

Y84 a numeric vector, salary 1984, in dollars.

Y83 a numeric vector, salary 1983, in dollars.

SHARES a numeric vector, number of shares the chairman holds.

REV a numeric vector, total revenue of the company.

INC a numeric vector, total income of the company.

AGE a numeric vector, age of chairman, in years.

Source

Reprinted with permission from the May 13, 1985, issue of Crain's Chicago Business. Copyright 1985 by Crain's Communications, Inc. The data given are a portion of the original table.

Examples

    data(E3.6)
    summary(E3.6)

E3.7   *Data on Oxygen Demand in Dairy Wastes*

Description

The E3.7 data frame has 20 rows and 7 columns.

Usage

    data(E3.7)

Format

This data frame contains the following columns:

Day a numeric vector, day of measurement, all measurements are on the same sample.

x.1 a numeric vector, biological oxygen demand, mg/liter.

x.2 a numeric vector, total Kjeldahl nitrogen, mg/liter.

x.3 a numeric vector, total solids, mg/liter.

x.4 a numeric vector, total volatile solids, a component of `x.3`, in mg/liter.

x.5 a numeric vector, chemical oxygen demand, mg/liter.

y a numeric vector, the response, log of oxygen demand, mg oxygen per minute.

Details

This is data from an experiment to construct a model for total oxygen demand in dairy wastes as a dunction of five laboratory measurements. Data were collected on samples kept in suspension in water in a laboratory for 220 days. All observations given here were taken on the same sample over time, so are probably dependent.

Source

Moore (1975) *Total Biochemical Oxygen Demand of Animal Manures*. Ph. D. thesis, University of Minnesota, Dept. of Agricultural Engineering.

Examples

```
data(E3.7)
summary(E3.7)
```

E3.8 *Map reading Test scores and Route Finding Scores*

Description

The `E3.8` data frame has 20 rows and 3 columns. 20 student volunteers where given a map reading test and a test of route finding on transit maps.

Usage

```
data(E3.8)
```

Format

This data frame contains the following columns:

y a numeric vector, ability to find routes to a given destination on a transit route map where scored y.

sc a numeric vector, scores on a map reading ability test.

Use a factor with levels `Non.users Users`, users and non-users of transit.

Source

Preof. Siim Soot, Department of Geography, University of Illinois at Chicago.

Examples

```
data(E3.8)
summary(E3.8)
```

E3.9   *Blood Velocity Data*

Description

The `E3.9` data frame has 18 rows and 4 columns. All the observations are for the same person.

Usage

```
data(E3.9)
```

Format

This data frame contains the following columns:

x.1 a numeric vector, cardiac output.

x.2 a numeric vector, carbon dioxide level in the blood.

y a numeric vector, blood flow velocity in the brain.

Aminophylline a factor with levels `no` `with`, Aminophylline used or not. The hypothesis is that aminophylline retards blood flow.

Source

Tonse Raju, M.D., Department of Neonatology, University of Illinois at Chicago.

Examples

```
data(E3.9)
summary(E3.9)
```

E4.1   *Traffic Fatality Data for Illinois*

Description

The `E4.1` data frame has 10 rows and 3 columns. Deaths are in deaths per 100 million vehicle miles.

Usage

```
data(E4.1)
```

Format

This data frame contains the following columns:

Year a numeric vector, the year.

Deaths a numeric vector, number of deaths.

DFR a numeric vector, deaths.t - deaths.(t-1).

Details

The interest are in possible changes after new safety regulations where in effect after 1966.

Source

Illinois Department of Transportation (1972).

Examples

```
data(E4.1)
summary(E4.1)
```

E4.10   *Votes from Chicago's Twenty-second Ward by Precinct*

Description

The `E4.10` data frame has 27 rows and 7 columns.

Usage

```
data(E4.10)
```

Format

This data frame contains the following columns:

Pr. a numeric vector, precinct number.

LATV a numeric vector, number of latin voters.

NONLV a numeric vector, number of non-latin voters.

TURNOUT a numeric vector, total number of votes cast.

GARCIA a numeric vector, number of votes for Garcia.

MARTINEZ a numeric vector, number of votes for Martinez.

YANEZ a numeric vector, number of votes for Yanez.

Details

Note that the votes for the three candidates may not add to the total turnout because of write-in votes, spoilt ballots, etc.

Source

Ray Flores, The Latino Institute, Chicago.

Examples

```
data(E4.10)
summary(E4.10)
```

E4.11   *Data on Cost of Repairing Starters, Ring Gears or Both in Diesel Engines*

Description

The E4.11 data frame has 133 rows and 2 columns.

Usage

```
data(E4.11)
```

Format

This data frame contains the following columns:

Cost a numeric vector, the repair cost in dollars.

Part a factor with levels `Both Ring gear Starter`, the type of part being repaired.

Source

M.R.Khavanin, Department of Mechanical Engineering, University of Illinois at Chicago.

Examples

```
data(E4.11)
E4.11.m1 <- lm(Cost ~ Part - 1, data=E4.11)
summary(E4.11.m1)
```

E4.12 *Time taken by Professional Dieticians and Interns for Four Patient Contact Activities*

Description

The E4.12 data frame has 24 rows and 6 columns. Each row is the activities and time taken by one dietician.

Usage

```
data(E4.12)
```

Format

This data frame contains the following columns:

Time a numeric vector, sum of time taken for all activities.

SC a numeric vector, number of patient contacts for screening.

DC a numeric vector, number of patient contacts for diet class.

MR a numeric vector, number of patient contacts for meal rounds.

TR a numeric vector, number of patient contacts for team rounds.

Dietician a factor with levels `Intern Prof`, dietician is professional or intern.

Source

The data where made available to one of the authors by a student.

Examples

```
m1 <- lm(Time ~ SC+DC+MR+TR-1, data=E4.12, subset=Dietician=="Prof")
summary(m1)
```

E4.13   *Data on Hospital Charges*

Description

The E4.13 data frame has 49 rows and 5 columns. Data on hospital charges for patients with an
identical diagnosis.

Usage

```
data(E4.13)
```

Format

This data frame contains the following columns:

Sex a factor with levels F M , male and female.

MD a factor with levels 499 730 1021, three different medical doctors.

Svty a factor with levels 1 2 3 4, severity of illness.

Chrg a numeric vector, total hospital charge in dollars.

Age a numeric vector, age of patient in years.

Source

Dr. Joseph Feinglass, Northwestern Memorial Hospital, Chicago.

Examples

```
data(E4.13)
summary(E4.13)
```

E4.4   *Measures of Quality for Agencies Delivering Transportation for the
                                   Elderly and the Handicapped*

Description

The E4.4 data frame has 40 rows and 3 columns.

Usage

```
data(E4.4)
```

Format

This data frame contains the following columns:

QUAL a numeric vector, a quality measure made using psychometric methods from results of questionares.

X.1 a numeric vector, an indicator variable for private ownership.

X.2 a numeric vector, an indicator variable for private for profit ownership.

Details

The quality data, QUAL, is constructed from questionares given to users of such services in the state of Illinois. Multiple services in the state of Illinois was scored using this method. The indicator variables was constructed to give first (X.1) a comparison between private and public services, then (X.2) a comparison between private not-for-profit and private for profit services.

Source

Slightly modified version of data supplied by Ms. Claire McKnight of the Department of Civil Engineering, City University of New York.

Examples

```
data(E4.4)
summary(E4.4)
```

E4.7   *Data on Per-Capita Income and Life Expectancy*

Description

The E4.7 data frame has 101 rows and 3 columns.

Usage

```
data(E4.7)
```

Format

This data frame contains the following columns:

Country a character vector, containing names of the countries.

LIFE a numeric vector, life expectancy, years. Early 1970's.

INC a numeric vector, per capita income in 1974 dollars. Early 1970's.

Source

From the *New York Times* (September, 28, 1975, p E-3).

Examples

```
data(E4.7)
attach(E4.7)
plot(INC, LIFE)
plot(log(INC), LIFE)
detach()
```

E6.1   *Data on Automobile Speed and Distance Covered to Come to a Stand-
                                             still after Breaking*

Description

The E6.1 data frame has 62 rows and 2 columns.

Usage

```
data(E6.1)
```

Format

This data frame contains the following columns:

d. a numeric vector, distance covered to come to a standstill after breaking.

sp. a numeric vector, speed before breaking.

Source

From Ezekiel,M. and F. A. Fox, *Methods of Correlation and Regression Analysis.* Copyright 1959
John Wiley and Sons, Inc.

Examples

```
data(E6.1)
attach(E6.1)
plot(sp., d.)
detach()
```

E6.10   *Data on Perceived and Computed Travel Times by Bus*

Description

The E6.10 data frame has 32 rows and 3 columns.

Usage

```
data(E6.10)
```

Format

This data frame contains the following columns:

n a numeric vector, number of respondents, weights for the linear regression.

x a numeric vector, computed travel times between a pair of zones in Chicago.

y a numeric vector, perceived travel times, as reported to the U.S. Census Bureau.

Details

x where computed from bus timetables, adding an average waiting time at the stop, and an average walking time from zone center to bus stop. y is the average reported by n travelers, to the US census bureau. The variable t introduced in the example below is the one for multiple bus transfers, used in example 8.1 page 161.

Source

The data where selected by one of the authors from a larger data set compiled by Cæsar Singh from census tapes, timetables and maps.

Examples

```
data(E6.10)
## Manipulations of the data for example 8.1, page 161:
t <- c(0,1,rep(0,20),1,rep(0,5),1,rep(0,3))
e6.10 <- data.frame(E6.10, t=t)
rm(t)
summary(e6.10)
```

E6.11   *Heights of Fathers and Sons*

Description

The E6.11 data frame has 12 rows and 3 columns.

Usage

```
data(E6.11)
```

Format

This data frame contains the following columns:

Height.Father a numeric vector, height of father to the nearest inch.

Aver.Height.Son a numeric vector, average heights of sons.

No.Fathers a numeric vector, number of fathers in each group.

Source

dacey (1983, Ch. 1) from McNemar (1969, p. 130), *Psycological Statistics*.

Examples

```
data(E6.11)
summary(E6.11)
```

E6.8   *Dial-a-ride Data*

Description

The E6.8 data frame has 54 rows and 7 columns. It has 7 variables describing 54 dial-a-ride services
in U.S. and Canada. It needs weighted regression.

Usage

```
data(E6.8)
```

Format

This data frame contains the following columns:

POP a numeric vector, population of area where service where operating.

AR a numeric vector, area of the place where service where provided.

RDR a numeric vector, number of riders using the system.

HR a numeric vector, hours of operation.

VH a numeric vector, number of vehicles in operation.

F a numeric vector, the fare used.

IND a numeric vector, a composite index, 1 when several ridership enhancing features where present, and 0 elsewhere.

Source

Collected by Louise Stanton-Maston, from 54 services in U.S. and Canada.

Examples

```
data(E6.8)
summary(E6.8)
```

E7.1   *Data on Dental Measurements*

Description

The E7.1 data frame has 4 rows and 12 columns. Dental measurements for girls from 8 to 14 years old. Each measurement is the distance, in mm, from the center of the pituary to the ptery-maxilliary fissure.

Usage

```
data(E7.1)
```

Format

This data frame contains the following columns:

Age a numeric vector, age of girl when measurement was taken.

S.1 a numeric vector, measurements for girl 1.

S.2 a numeric vector, measurements for girl 2.

S.3 a numeric vector, measurements for girl 3.

S.4 a numeric vector, measurements for girl 4.

S.5 a numeric vector, measurements for girl 5.

S.6 a numeric vector, measurements for girl 6.

S.7 a numeric vector, measurements for girl 7.

S.8 a numeric vector, measurements for girl 8.

S.9 a numeric vector, measurements for girl 9.

S.10 a numeric vector, measurements for girl 10.

S.11 a numeric vector, measurements for girl 11.

Source

Pothoff and Roy (1964).

Examples

```
data(E7.1)
summary(E7.1)
```

E7.2   *Prices of Crude Oil, Natural Gas, Bituminous Coal and Lignite, and Anthracite by Year.*

Description

The E7.2 data frame has 32 rows and 5 columns. Prices are in 1972 cents (U.S) by 1000 BTU.

Usage

```
data(E7.2)
```

Format

This data frame contains the following columns:

year a numeric vector, year of observation.

Oil a numeric vector, price of oil.

Gas a numeric vector, price of Gas.

Bit. a numeric vector, price of Bituminous Coal and Lignite.

Anth. a numeric vector, price of Anthracite.

Source

Darrel Sala, Institute of Gas Technology, Chicago.

Examples

```
data(E7.2)
summary(E7.2)
```

E7.3   *Data on Intake/Output Ratio*

Description

The E7.3 data frame has 19 rows and 6 columns. It gives the ratios u of fluid intake to urine output over five consecutive 8-hour periods for 19 babies divided in a control and treatment group.

Usage

```
data(E7.3)
```

Format

This data frame contains the following columns:

G a factor with levels surfactant placebo

u.1 a numeric vector, u for time period 1.

u.2 a numeric vector, u for time period 2.

u.3 a numeric vector, u for time period 3.

u.4 a numeric vector, u for time period 4.

u.5 a numeric vector, u for time period 5.

Source

Rama Bhat, M.D., Department of Pediatrics, University of Illinoi at Chicago. This data is part of a larger data set.

Examples

```
data(E7.3)
summary(E7.3)
```

E7.4   *Data on PCO2 and Cerebral Blood Flow for Five Regions of the Brain of five Chimpanzees*

Description

The E7.4 data frame has 5 rows and 11 columns. Five baby chimpanzees were injected with a heavy dose of HIV infection. After six months, the radio-active microsphere technique was used to measure brain blood flow in ml per 100 grams of brain tissue, from five regions of the brain. The partial pressure of carbon dioxide in millimeters of mercury was also obtained.

Usage

    data(E7.4)

Format

This data frame contains the following columns:

Ch. a numeric vector, id number of chimpanzee.

Fron.x a numeric vector, Frontal partial pressure of carbon dioxide.

Fron.y a numeric vector, Frontal blood flow.

Pari.x a numeric vector, Parietal partial pressure of carbon dioxide.

Pari.y a numeric vector, Parietal blood flow.

Occi.x a numeric vector, Occipital partial pressure of carbon dioxide.

Occi.y a numeric vector, Occipital blood flow.

Temp.x a numeric vector, Temporal partial pressure of carbon dioxide.

Temp.y a numeric vector, Temporal blood flow.

Cere.x a numeric vector, Cerebellum partial pressure of carbon dioxide.

Cere.y a numeric vector, Cerebellum blood flow.

Source

Tonse Raju, M.D. Department of pediatrics, University of Illinois at Chicago.

Examples

    data(E7.4)
    summary(E7.4)

E7.5   *Data on Static Weights and Weight in Motion of Trucks*

Description

The E7.5 data frame has 26 rows and 6 columns.

Usage

    data(E7.5)

Format

This data frame contains the following columns:

sw.1 a numeric vector, static weight of axle 1.

wim.1 a numeric vector, weight in motion of axle 1.

sw.23 a numeric vector, static weight of axles 2–3.

wim.23 a numeric vector, weight in motion of axles 2–3.

sw.45 a numeric vector, static weight of axles 4–5.

wim.45 a numeric vector, weight in motion of axles 4–5.

Details

Trucks can be weighted by two methods. In one, a truck needs to go into a weighting station and each axle is weighted by conventional means. The other is newer and a somewhat experimental method where a thin pad is placed on the highway and axles are weighted as trucks pass over it. Former weight are called static weights (`sw`) while later are called weights in motion (`wim`).

Source

Saleh Mumayiz, Urban Transportation Center, University of Illinois at Chicago, who compiled the data from a data set provided by the Illinois Department of Transportation.

Examples

```
data(E7.5)
summary(E7.5)
plot(E7.5)
```

E7.6 *Community Area Data for the North Part of the City of Chicago*

Description

The `E7.6` data frame has 34 rows and 5 columns.

Usage

```
data(E7.6)
```

Format

This data frame contains the following columns:

Area.Name a character vector, name of area.

PB a numeric vector, percentage of population which are black.

PS a numeric vector, percentage of population which are spanish speaking.

PA a numeric vector, percentage of population over 65.

Income a numeric vector, median family income for each area.

Source

Data set were constructed by Prof. Siim Soot, Dept. of Geography, University of Illinois at Chicago.

See Also

`E7.7`, which is the adjacency matrix for the 34 areas.

Examples

```
data(E7.6)
summary(E7.6)
```

E7.7   *The Contiguity Matrix for the 34 Areas in Northern Chicago*

Description

This is the contiguity matrix for the 34 areas in northern Chicago, given in `E7.6`. Contains only 0's and 1's with the obvious interpretation.

Usage

```
data(E7.7)
```

E8.12   *Data on Lung Cancer Deaths and Cigarette Smoking*

Description

The `E8.12` data frame has 11 rows and 3 columns.

Usage

```
data(E8.12)
```

Format

This data frame contains the following columns:

Country a character vector, the country.

y a numeric vector, male deaths in 1950 for lung cancer, per million.

x a numeric vector, per capita cigarette consumption in 1930.

Source

Tufte, (1974) *Data Analysis for Politics and Policy*. Englewood Cliffs, N.J.: Prentice-Hall. Data are adapted.

Examples

```
data(E8.12)
summary(E8.12)
```

E8.13 *Florida Cumulus Experiment Data*

Description

The E8.13 data frame has 20 rows and 7 columns, giving data on the effects of cloud seeding by silver iodide crystals on precipitation. Each data point is one day.

Usage

```
data(E8.13)
```

Format

This data frame contains the following columns:

A a factor with levels NoSeed Seed

T a numeric vector, number of days after the first day of the experiment.

S a numeric vector, relates to heights of clouds.

C a numeric vector, percentage of clode cover in the experimental area.

P a numeric vector, total rainfall in the study area one hour before seeding (in 10^7 cubic meters).

E a factor with levels Moving Stationary , indicating if the radar echo was mowing or not.

y a numeric vector, the response, natural logarithm of precipitation in the target area in a 6-hour period (in 10^7 cubic meters).

Source

Woodley, et.al (1977) *Rainfall Results 1970–1975: Florida Area Cumulus Experiment.* . Science **95** 735–742. Copyright 1977 by the AAAS.

Examples

```
data(E8.13)
summary(E8.13)
plot(E8.13)
```

E9.11 *Data on Transit Privatization*

Description

The E9.11 data frame has 17 rows and 10 columns.

Usage

```
data(E9.11)
```

Format

This data frame contains the following columns:

V1 a numeric vector, average capacity of buses in service.

V2 a numeric vector, ratio of buses in use during non-peak periods to those in use in peak periods.

V3 a numeric vector, average speed.

V4 a numeric vector, vehicle-miles contracted.

V5 a numeric vector, distance of center from metroploitan area.

V6 a numeric vector, population of metropolitan area.

V7 a numeric vector, percentage of work trips in the metropolitan area that are made by transit.

V8 a numeric vector, Buses owned by sponsor / buses owned by contractor

V9 a numeric vector, per capita income for metropolitan area.

PCS a numeric vector, per cent savings that occurred when some transit lines was given to private companies.

Source

Prof E.K.Morlok, Dept. of Systems Engineering, University of Pennsylvania.

Examples

```
data(E9.11)
summary(E9.11)
plot(E9.11)
```

E9.18	*Data Travel Times and Usage for Automobiles and Public Transportation*

Description

The `E9.18` data frame has 51 rows and 4 columns.

Usage

```
data(E9.18)
```

Format

This data frame contains the following columns:

t.a a numeric vector, travel time by car, in tenth of minutes.

t.r a numeric vector, travel time by public transportation, in tenth of minutes.

m.a a numeric vector, number of those who used a car or van either as driver or passenger.

m.r a numeric vector, number of people using any kind of public transportation.

Details

Travel times modified by one of the authors to reflect the cost of parking. For downtown zones (Chicago) this amounted to about 60 minutes.

Source

Selected by Robert Drozd from Census (US) Urban Transportation Planning Package, for the Chicago area.

Examples

```
data(E9.18)
summary(E9.18)
plot(E9.18)
```

E9.19   *Acceleration data*

Description

The `E9.19` data frame has 50 rows and 4 columns.

Usage

```
data(E9.19)
```

Format

This data frame contains the following columns:

ACC a numeric vector, Acceleration of different vehicles.

WHP a numeric vector, weight-to-horsepower ratio.

SP a numeric vector, speed at which they were travelling.

G a numeric vector, Grade of road, G=0 implies road was horizontal.

Source

Raj Tejwaney, Department of civil Engineering, University of Illinoi at Chicago.

Examples

```
data(E9.19)
summary(E9.19)
plot(E9.19)
```

E9.20   *Stadium Cleanup Data*

Description

The `E9.20` data frame has 16 rows and 3 columns.

Usage

```
data(E9.20)
```

Format

This data frame contains the following columns:

C a numeric vector, cost of cleanup. Units forgotten.

R.HD a numeric vector, sales at hot-dog stands. Units forgotten.

R.B a numeric vector, sales at beer stands. Units forgotten.

Source

The authors of the book.

Examples

```
data(E9.20)
summary(E9.20)
plot(E9.20)
```

E9.21   *Depreciation in Market Value of Large Factories*

Description

The `E9.21` data frame has 11 rows and 2 columns.

Usage

```
data(E9.21)
```

Format

This data frame contains the following columns:

Age a numeric vector, units not given, probably years.

Depr a numeric vector, units not given.

Source

Diamond-stars Motors, Normal, Il. Gary Shultz, general Counsel, made this data available.

Examples

```
data(E9.21)
summary(E9.21)
plot(E9.21)
```

E9.3   *"Areas", lengths and widths of rectangles*

Description

The E9.3 data frame has 50 rows and 3 columns. Made by random sampling numbers.

Usage

```
data(E9.3)
```

Format

This data frame contains the following columns:

y a numeric vector, area of the rectangle.

x1 a numeric vector, length of the rectangle.

x2 a numeric vector, width of the ractangle.

Examples

```
data(E9.3)
E9.3.m1 <- lm(y ~ x1 + x2, data=E9.3)
attach(E9.3)
plot(x1, resid(E9.3.m1))
plot(x2, resid(E9.3.m1))
detach(E9.3)
```

E9.8   *Data on monthly rent, annual income and houshould size*

Description

The E9.8 data frame has 27 rows and 3 columns.

Usage

```
data(E9.8)
```

Format

This data frame contains the following columns:

R a numeric vector, Monthly rent in dollars.

I a numeric vector, annual income in 1000$.

S a numeric vector, household size.

Details

Example 9.8 in Sen and Srivastava, page 201.

Source

Selected by one of the authors from a much larger data set, collected from several sources about 20 years ago.

Examples

```
data(E9.8)
attach(E9.8)
E9.8.m1 <- lm(R ~ I + S, data=E9.8)
summary(E9.8.m1)
plot(I, resid(E9.8.m1, type="partial")[,"I"])
plot(S, resid(E9.8.m1, type="partial")[,"S"])
detach()
```

Ec.8 *Data on asylum requests to the U.S. by country of origen of applicant*

Description

The Ec.8 data frame has 112 rows and 5 columns.

Usage

```
data(Ec.8)
```

Format

This data frame contains the following columns:

Country a character vector, containing country of origen of applicant.

APR a numeric vector, number of successful applications.

DEN a numeric vector, number of denied applications.

H a numeric vector, 1 if country is considered hostile to the U.S., 0 en other case.

E a numeric vector, 1 if country is European or mainly inhabited by people of european descent.

Source

Prof. Barbara Yarnold, Dept. of political science, Saginaw Valley State University, Saginaw, Michigan.

Examples

```
data(Ec.8)
summary(Ec.8)
attach(Ec.8)
Ec.8.m1 <- glm(cbind(APR, DEN) ~ E + H, data=Ec.8, family=binomial)
summary(Ec.8.m1)
detach()
```

Ex.7.7   *U.S. Population in thousands, for exercise 7.7*

Description

The Ex.7.7 data frame has 19 rows and 2 columns.

Usage

```
data(Ex.7.7)
```

Format

This data frame contains the following columns:

y a numeric vector, U.S. population in thousands.

t a numeric vector, year.

Source

Sen and Srivastava.

Examples

```
##---- Should be DIRECTLY executable !! ----
data(Ex.7.7)
with(Ex.7.7, plot(y ~ t))
summary(Ex.7.7)
```

Ex4.4   *Data on Effects of Air Pollution on Interpersonal Attraction*

Description

The Ex4.4 data frame has 24 rows and 3 columns. An experiment was conducted to examine the effects of air pollution on interpersonal attraction. Twenty-four subjects were each placed with a stranger for a 15-minute period in a room which was either odor free or contaminated with ammonium sulfide. The stranger came from a culture which was similar or dissimilar to that of the subject. At the end of the encounter, each subject was asked to assess his degree of attraction towards the stranger on a likert scale of 1–10 with 10 indicating strong attraction.

Usage

```
data(Ex4.4)
```

Format

This data frame contains the following columns:

Likert a numeric vector, attraction on a likert scale.

Odor a factor with levels Free Odor, room was contaminated or not.

Culture a factor with levels Dissimilar Similar, similar or dissimilar culture.

Source

The full data set is given in Srivastava and Carter (1983).

Examples

```
data(Ex4.4)
summary(Ex4.4)
plot(Ex4.4)
```

Index