Package 'Sleuth2'

October 11, 2010

Title Data sets from Ramsey and Schafer's "Statistical Sleuth (2nd ed)"

Version 1.0-2

Date 2010-10-10

Author Original by F.L. Ramsey and D.W. Schafer, modifications by Daniel W Schafer, Jeannie Sifneos and Berwin A Turlach

Description Data sets from Ramsey, F.L. and Schafer, D.W. (2002), "The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)" Duxbury.

Maintainer Berwin A Turlach <berwin@maths.uwa.edu.au>

LazyData yes

Depends R (>= 2.6.0)

Suggests lattice

License GPL (>= 2)

URL http://r-forge.r-project.org/projects/sleuth2/

R topics documented:

`Sleuth2-package` *The R Sleuth2 package*

Description

Data sets from Ramsey and Schafer's "Statistical Sleuth (2nd ed)"

Details

This package contains a variety of datasets. For a complete list, use `library(help="Sleuth2")` or `Sleuth2Manual()`.

Author(s)

Original by F.L. Ramsey and D.W. Schafer

Modifications by Daniel W Schafer, Jeannie Sifneos and Berwin A Turlach

Maintainer: Berwin A Turlach `<berwin@maths.uwa.edu.au>`

case0101 *Motivation and Creativity*

Description

Data from an experiment concerning the effects of intrinsic and extrinsic motivation on creativity. Subjects with considerable experience in creative writing were randomly assigned to on of two treatment groups.

Usage

```
case0101
```

Format

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

`Score` creativity score

`Treatment` factor denoting the treatment group

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Amabile, T. (1985). Motivation and Creativity: Effects of Motivational Orientation on Creative Writers, *Journal of Personality and Social Psychology* **48**(2): 393–399.

Examples

```
str(case0101)
boxplot(Score~Treatment, case0101)
```

case0102 *Sex Discrimination in Employment*

Description

The data are the beginning salaries for all 32 male and all 61 female skilled, entry–level clerical employees hired by a bank between 1969 and 1977.

Usage

```
case0102
```

Format

A data frame with 93 observations on the following 2 variables.

`Salary` starting salaries (in US$)

`Sex` sex of the clerical employee

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Roberts, H.V. (1979). Harris Trust and Savings Bank: An Analysis of Employee Compensation, *Report 7946*, Center for Mathematical Studies in Business and Economics, University of Chicago Graduate School of Business.

See Also

`case1202`

Examples

```
str(case0102)
boxplot(Salary~Sex, case0102)
```

`case0201` *Bumpus's Data on Natural Selection (Humerus)*

Description

As evidence in support of natural selection, Bumpus presented measurements on house sparrows brought to the Anatomical Laboratory of Brown University after an uncommonly severe winter storm. Some of these birds had survived and some had perished. Bumpus asked whether those that perished did so because they lacked physical characteristics enabling them to withstand the intensity of that particular instance of selective elimination. The data are on the humerus (arm bone) lengths for the 24 adult male sparrows that perished and for the 35 adult males that survived.

Usage

```
case0201
```

Format

A data frame with 59 observations on the following 2 variables.

`Humerus` Humerus length of adult male sparrows (in inches)

`Status` factor variable indicating whether the sparrow perished or survived in a winter storm

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

See Also

ex0221, ex2016

Examples

```
str(case0201)
with(subset(case0201, Status=="Perished"), stem(Humerus, scale=10))
with(subset(case0201, Status=="Survived"), stem(Humerus))
```

---

case0202 *Anatomical Abnormalities Associated with Schizophrenia*

Description

Are any physiological indicators associated with schizophrenia? In a 1990 article, researchers reported the results of a study that controlled for genetic and socioeconomic differences by examining 15 pairs of monozygotic twins, where one of the twins was schizophrenic and the other was not. The researchers used magnetic resonance imaging to measure the volumes (in cm^3) of several regions and subregions of the twins' brains.

Usage

```
case0202
```

Format

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

`Unaffect` volume of left hippocampus of unaffected twin (in cm^3)

`Affected` volume of left hippocampus of affected twin (in cm^3)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Suddath, R.L., Christison, G.W., Torrey, E.F., Casanova, M.F. and Weinberger, D.R. (1990). Anatomical Abnormalities in the Brains of Monozygotic Twins Discordant for Schizophrenia, *New England Journal of Medicine* **322**(12): 789–794.

Examples

```
str(case0202)
with(case0202, stem(Unaffect-Affected, scale=2))
```

`case0301` *Cloud Seeding*

Description

Does dropping silver iodide onto clouds increase the amount of rainfall they produce? In a randomized experiment, researchers measured the volume of rainfall in a target area (in acre-feet) on 26 suitable days in which the clouds were seeded and on 26 suitble days in which the clouds were not seeded.

Usage

```
case0301
```

Format

A data frame with 52 observations on the following 2 variables.

`Rainfall` the volume of rainfall in the target area (in acre-feet)

`Treatment` a factor with levels `"Unseeded"` and `"Seeded"` indicating whether the clouds were unseeded or seeded.

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Simpson, J., Olsen, A., and Eden, J. (1975). A Bayesian Analysis of a Multiplicative Treatment Effect in Weather Modification. *Technometrics* **17**: 161–166.

Examples

```
str(case0301)
boxplot(Rainfall ~ Treatment, case0301)
boxplot(log(Rainfall) ~ Treatment, case0301)
library(lattice)
bwplot(Treatment ~ log(Rainfall), case0301)
bwplot(log(Rainfall) ~ Treatment, case0301)
```

`case0302` *Agent Orange*

Description

In 1987, researchers measured the TCDD concentration in blood samples from 646 U.S. veterans of the Vietnam War and from 97 U.S. veterans who did not serve in Vietnam. TCDD is a carcinogenic dioxin in the herbicide called Agent Orange, which was used to clear jungle hiding areas by the U.S. military in the Vietnam War between 1962 and 1970.

Usage

```
data(case0302)
```

Format

A data frame with 743 observations on the following 2 variables.

`Dioxin` the concentration of TCDD, in parts per trillion

`Veteran` factor variable with two levels, `"Vietnam"` and `"Other"`, to indicate the type of veteran

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Centers for Disease Control Veterans Health Studies: Serum 2,3,7,8-Tetraclorodibenzo-p-dioxin Levels in U.S. Army Vietnam-era Veterans. *Journal of the American Medical Association* **260**: 1249–1254.

Examples

```
str(case0302)
boxplot(Dioxin ~ Veteran, case0302)
t.test(Dioxin ~ Veteran, case0302)
## To examine results with largest dioxin omitted
t.test(Dioxin ~ Veteran, case0302, subset=(Dioxin < 40))
```

case0401 *Space Shuttle*

Description

The number of space shuttle O-ring incidents for 4 space shuttle launches when the air temperatures was below 65 degrees F and for 20 space shuttle launches when the air temperature was above 65 degrees F.

Usage

```
case0401
```

Format

A data frame with 24 observations on the following 2 variables.

`Incidents` the number of O-ring incidents

`Launch` factor variable with two levels—`"Cool"` and `"Warm"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Feynman, R.P. (1988). *What do You Care What Other People Think?* W. W. Norton.

See Also

`ex2011`, `ex2223`

Examples

```
str(case0401)
stem(subset(case0401, Launch=="Cool", Incidents, drop=TRUE))
stem(subset(case0401, Launch=="Warm", Incidents, drop=TRUE))
```

case0402 *Cognitive Load*

Description

Educational researchers randomly assigned 28 ninth-year students in Australia to receive coordinate geometry training in one of two ways: a conventional way and a modified way. After the training, the students were asked to solve a coordinate geometry problem. The time to complete the problem was recorded, but five students in the "conventional" group did not complete the solution in the five minute alloted time.

Usage

```
case0402
```

Format

A data frame with 28 observations on the following 3 variables.

`Time` the time (in seconds) that the student worked on the problem

`Treatmt` factor variable with two levels—`"Modified"` and `"Conventional"`

`Censor` 1 if the individual did not complete the problem in 5 minutes, 0 if they did

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Sweller, J., Chandler, P., Tierney, P. and Cooper, M. (1990). Cognitive Load as a Factor in the Structuring of Technical Material, *Journal of Experimental Psychology General* **119**(2): 176–192.

Examples

```
str(case0402)
stem(subset(case0402, Treatmt=="Conventional", Time, drop=TRUE))
stem(subset(case0402, Treatmt=="Modified", Time, drop=TRUE))
wilcox.test(Time ~ Treatmt, case0402)
```

---

case0501 *Diet Restriction and Longevity*

Description

Female mice were randomly assigned to six treatment groups to investigate whether restricting dietary intake increases life expectancy. Diet treatments were:

1. `"NP"`—mice ate unlimited amount of nonpurified, standard diet
2. `"N/N85"`—mice fed normally before and after weaning. After weaning, ration was controlled at 85 kcal/wk
3. `"N/R50"`—normal diet before weaning and reduced calorie diet (50 kcal/wk) after weaning
4. `"R/R50"`—reduced calorie diet of 50 kcal/wk both before and after weaning
5. `"N/R50 lopro"`—normal diet before weaning, restricted diet (50 kcal/wk) after weaning and dietary protein content decreased with advancing age
6. `"N/R40"`—normal diet before weaning and reduced diet (40 Kcal/wk) after weaning.

Usage

```
case0501
```

Format

A data frame with 349 observations on the following 2 variables.

`Lifetime` the lifetime of the mice (in months)

`Diet` factor variable with six levels—`"NP"`, `"N/N85"`, `"lopro"`, `"N/R50"`, `"R/R50"` and `"N/R40"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Weindruch, R., Walford, R.L., Fligiel, S. and Guthrie D. (1986). The Retardation of Aging in Mice by Dietary Restriction: Longevity, Cancer, Immunity and Lifetime Energy Intake, *Journal of Nutrition* **116**(4):641–54.

Examples

```
str(case0501)
boxplot(Lifetime~Diet, width=c(rep(.8,6)), data=case0501,
        xlab="Diet", ylab="Lifetime in months")
summary(subset(case0501, Diet=="NP", Lifetime))
```

---

case0502 *The Spock Conspiracy Trial*

Description

In 1968, Dr. Benjamin Spock was tried in Boston on charges of conspiring to violate the Selective Service Act by encouraging young men to resist being drafted into military service for Vietnam. The defence in the case challenged the method of jury selection claiming that women were underrepresented. Boston juries are selected in three stages. First 300 names are selected at random from the City Directory, then a venire of 30 or more jurors is selected from the initial list of 300 and finally, an actual jury is selected from the venire in a nonrandom process allowing each side to exclude certain jurors. There was one woman on the venire and no women on the final list. The defence argued that the judge in the trial had a history of venires in which women were systematically underrepresented and compared the judge's recent venires with the venires of six other Boston area district judges.

Usage

```
case0502
```

Format

A data frame with 46 observations on the following 2 variables.

`Percent` is the percent of women on the venire's of the Spock trial judge and 6 other Boston area judges

`Judge` a factor with levels `"Spock's"`, `"A"`, `"B"`, `"C"`, `"D"`, `"E"` and `"F"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Zeisel, H. and Kalven, H. Jr. (1972). Parking Tickets and Missing Women: Statistics and the Law *in* Tanur, J.M. et al. (eds.) *Statistics: A Guide to the Unknown*, Holden-Day.

Examples

```
str(case0502)
boxplot(Percent~Judge, data=case0502,
        xlab="Judge",ylab="Percentage of Women")
percent.spocks <- subset(case0502, Judge == "Spock's", Percent)
percent.others <- subset(case0502, Judge != "Spock's", Percent)
t.test( percent.spocks,percent.others)
summary(aov(Percent~Judge, case0502, subset = Judge != "Spock's"))

#as in Display 5.10
summary(aov(Percent~Judge, case0502))
```

---

case0601 *Discrimination Against the Handicapped*

Description

Study explores how physical handicaps affect people's perception of employment qualifications. Researchers prepared 5 videotaped job interviews using actors with a script designed to reflect an interview with an applicant of average qualifications. The 5 tapes differed only in that the applicant appeared with a different handicap in each one. Seventy undergraduate students were randomly assigned to view the tapes and rate the qualification of the applicant on a 0-10 point scale.

Usage

```
case0601
```

Format

A data frame with 70 observations on the following 2 variables.

`Score` is the score each student gave to the applicant

`Handicap` is a factor variable with 5 levels—`"None"`, `"Amputee"`, `"Crutches"`, `"Hearing"` and `"Wheelchair"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Cesare, S.J., Tannenbaum, R.J. and Dalessio, A. (1990). Interviewers' Decisions Related to Applicant Handicap Type and Rater Empathy, *Human Performance* **3**(3): 157–171.

Examples

```
str(case0601)
boxplot(Score~Handicap, data=case0601, ylab="Score")
aov.handicap <- aov(Score ~ Handicap, case0601)
summary(aov.handicap)
TukeyHSD(aov.handicap)

#Calculate confidence interval for linear combination
#(wheelchair+crutches)/2 - (amputee+hearing)/2 as in Display 6.4
mean.handicaps <- with(case0601, tapply(Score, Handicap, mean))
var.handicaps <- with(case0601, tapply(Score, Handicap, var))

n <- 14
s.pooled <- sqrt(sum((n-1)*var.handicaps)/sum((n-1)*5))

## either
cr.wh <- mean.handicaps["Wheelchair"] + mean.handicaps["Crutches"]
am.he <- mean.handicaps["Amputee"] + mean.handicaps["Hearing"]
g <- cr.wh/2 - am.he/2
## or
contr <- c(0, -1, 1, -1, 1)/2
g <- sum(contr * mean.handicaps)

se.g <- s.pooled * sqrt(sum(contr^2)/n)
t.65 <- qt(.975, 65)
## ci
g + c(-1,1) * t.65 * se.g
```

---

case0602 *Mate Preference of Platyfish*

Description

Do female Platyfish prefer male Platyfish with yellow swordtails? A.L. Basolo proposed and tested a selection model in which females have a pre-existing bias for a male trait even before the males possess it. Six pairs of males were surgically given artificial, plastic swordtails—one pair received a bright yellow sword, the other a transparent sword. Females were given the opportunity to engage in courtship activity with either of the males. Of the total time spent by each female engaged in courtship during a 20 minute observation period, the percentages of time spent with the yellow-sword male were recorded.

Usage

```
case0602
```

Format

A data frame with 84 observations on the following 3 variables.

`Proportion` The proportion of courtship time spent by 84 females with the yellow-sword males

`Pair` Factor variable with 6 levels—`"Pair 1"`, `"Pair 2"`, `"Pair 3"`, `"Pair 4"`, `"Pair 5"` and `"Pair 6"`

`Length` Body size of the males

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Basolo, A.L. (1990). Female Preference Predates the Evolution of the Sword in Swordtail Fish, *Science* **250**: 808–810.

Examples

```
str(case0602)
boxplot(Proportion~Pair, case0602, ylab="Proportion")

#as in Display 6.5
summary(aov(Proportion~Pair, case0602))

n.fish  <- with(case0602, tapply(Proportion, Pair, length))
av.fish <- with(case0602, tapply(Proportion, Pair, mean))
sd.fish <- with(case0602, tapply(Proportion, Pair, sd))
male.body.size <- with(case0602, tapply(Length, Pair, unique))
mean.body <- mean(male.body.size)
table.fish <- data.frame(n.fish, round(av.fish*100,2),
                         round(sd.fish*100,2), male.body.size,
                         2*(male.body.size-mean.body))
names(table.fish) <- c("n", "average", "sd", "male.body.size", "coefficient")
s.pooled <- with(table.fish, round(sqrt(sum(sd^2*(n-1))/sum(n-1)),2))
g <- with(table.fish, sum(average*coefficient))
se.g <- with(table.fish, round(s.pooled*sqrt(sum(coefficient^2/n)),2))
g/se.g
```

`case0701` *The Big Bang*

Description

Hubble's initial data on 24 nebulae outside the Milky Way.

Usage

```
case0701
```

Format

A data frame with 24 observations on the following 2 variables.

`Velocity` recession velocity (in kilometres per second)

`Distance` distance from earth (in magaparsec)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Hubble, E. (1929). A Relation Between Distance and Radial Velocity Among Extragalactic Nebulae, *Proceedings of the National Academy of Science* **15**: 168–173.

See Also

`ex0727`

Examples

```
str(case0701)
plot(case0701)
```

---

case0702 *Meat Processing and pH*

Description

A certain kind of meat processing may begin once the pH in postmortem muscle of a steer carcass has decreased sufficiently. To estimate the timepoint at which pH has dropped sufficiently, 10 steer carcasses were assigned to be measured for pH at one of five times after slaughter.

Usage

```
case0702
```

Format

A data frame with 10 observations on the following 2 variables.

`Time` time after slaughter (hours)

`pH` pH level in postmortem muscle

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Schwenke, J.R. and Milliken, G.A. (1991). On the Calibration Problem Extended to Nonlinear Models, *Biometrics* **47**(2): 563–574.

See Also

ex0816

Examples

```
str(case0702)
plot(case0702)
```

---

case0801 *Island Area and Number of Species*

Description

The data are the numbers of reptile and amphibian species and the island areas for seven islands in the West Indies.

Usage

```
case0801
```

Format

A data frame with 7 observations on the following 2 variables.

`Area` area of island (in square miles)

`Species` number of reptile and amphibian species on island

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
str(case0801)
plot(case0801)
```

case0802 *Breakdown Times for Insulating Fluid under different Voltage*

Description

In an industrial laboratory, under uniform conditions, batches of electrical insulating fluid were subjected to constant voltages until the insulating property of the fluids broke down. Seven different voltage levels were studied and the measured reponses were the times until breakdown.

Usage

```
case0802
```

Format

A data frame with 76 observations on the following 3 variables.

`Time` times until breakdown (in minutes)

`Voltage` voltage applied (in kV)

`Group` factor variable (group number)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
str(case0802)
plot(log(Time)~Voltage, case0802)
```

case0901 *Effects of Light on Meadowfoam Flowering*

Description

Meadowfoam is a small plant found growing in moist meadows of the US Pacific Northwest. Researchers reported the results from one study in a series designed to find out how to elevate meadowfoam production to a profitable crop. In a controlled growth chamber, they focused on the effects of two light–related factors: light intensity and the timeing of the onset of the ligth treatment.

Usage

```
case0901
```

Format

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

`Flowers` average number of flowers per meadowfoam plant

`Time` time light intensity regiments started

`Intens` light intensity (in μmol/m^2/sec)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
str(case0901)
plot(Flowers~Intens, case0901, pch= ifelse(Time=="Early", 19, 21))
```

---

case0902 *Why Do Some Mammals Have Large Brains for Their Size?*

Description

The data are the average values of brain weight, body weight, gestation lengths (length of pregnancy) and litter size for 96 species of mammals.

Usage

```
case0902
```

Format

A data frame with 96 observations on the following 5 variables.

`Species` species

`Brain` average brain weight (in grams)

`Body` average body weight (in kilograms)

`Gestation` gestation period (in days)

`Litter` average litter size

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

See Also

`case0902`

Examples

```
str(case0902)
pairs(log(Brain)~log(Body)+log(Litter)+Gestation, case0902)
```

case1001 *Galileo's Data on the Motion of Falling Bodies*

Description

In 1609 Galileo proved mathematically that the trajectory of a body falling with a horizontal velocity component is a parabola. His search for an experimental setting in which horizontal motion was not affected appreciably (to study inertia) let him to construct a certain apparatus. The data comes from one of his experiments.

Usage

```
case1001
```

Format

A data frame with 7 observations on the following 2 variables.

`Distance` horizontal distances (in punti)

`Height` initial height (in punti)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
str(case1001)
plot(Distance ~ Height, case1001)
```

case1002 *The Energy Costs of Echolocation by Bats*

Description

The data are on in–flight energy expenditure and body mass from 20 energy studies on three types of flying vertebrates: echolocating bats, non–echolocating bats and non–echolocating birds.

Usage

```
case1002
```

Format

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

`Species` species

`Mass` mass (in grams)

`Type` a factor with 3 levels indicating the type of flying vertebrate

`Energy` in–flight energy expenditure (in W)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Speakman, J.R. and Racey, P.A. (1991). No cost of Echolocation for Bats in Flight, *Nature* **350**: 421–423.

Examples

```
str(case1002)

plot(log(Energy)~log(Mass), case1002,
     pch = ifelse(Type=="echolocating bats", 19,
                  ifelse(Type=="non-echolocating birds", 21, 24)))

plot(Energy~Mass, case1002, log="xy",
     xlab = "Body Mass (g) (log scale)",
     ylab = "Energy Expenditure (W) (log scale)",
     pch = ifelse(Type=="echolocating bats", 19,
                  ifelse(Type=="non-echolocating birds", 21, 24)))
legend(7, 50, pch=c(24, 21, 19),
     c("Non-echolocating bats", "Non-echolocating birds","Echolocating bats"))

library(lattice)
yticks <- c(1,2,5,10,20,50)
xticks <- c(10,20,50,100,200,500)
xyplot(Energy ~ Mass, case1002, groups=Type,
       scales = list(log=TRUE, y=list(at=yticks), x=list(at=xticks)),
       ylab = "Energy Expenditure (W) (log scale)",
       xlab = "Body Mass (g) (log scale)",
       auto.key = list(x = 0.2, y = 0.9, corner = c(0, 1), border = TRUE))
```

---

case1101 *Alcohol Metabolism in Men and Women*

Description

These data were collected on 18 women and 14 men to investigate a certain theory on why women exhibit a lower tolerance for alcohol and develop alcohol–related liver disease more readily than men.

Usage

```
case1101
```

Format

A data frame with 32 observations on the following 5 variables.

`Subject` subject number in the study

`Metabol` first–pass metabolism of alcohol in the stomach (in mmol/liter-hour)

`Gastric` gastric alcohol dehydrogenase activity in the stomach (in μmol/min/g of tissue)

`Sex` sex of the subject

`Alcohol` whether the subject is alcoholic or not

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
str(case1101)

plot(Metabol~Gastric, case1101,
     pch=ifelse(Sex=="Female", 19, 21),
     col=ifelse(Alcohol=="Alcoholic", "red", "green"))
legend(1,12, pch=c(19,21,19,21), col=c("green","green", "red", "red"),
             c("Non-alcoholic Females", "Non-alcoholic Males",
               "Alcoholic Females", "Alcoholic Males"))

library(lattice)
xyplot(Metabol~Gastric|Sex*Alcohol, case1101)
xyplot(Metabol~Gastric, case1101, groups=Sex:Alcohol,
        auto.key=list(x=0.2, y=0.8, corner=c(0,0), border=TRUE))
```

---

case1102 *The Blood–Brain Barrier*

Description

The human brain is protected from bacteria and toxins, which course through the blood–stream, by a single layer of cells called the blood–brain barrier. These data come from an experiment (on rats, which process a similar barrier) to study a method of disrupting the barrier by infusing a solution of concentrated sugars.

Usage

```
case1102
```

Format

A data frame with 34 observations on the following 9 variables.

`Brain` Brain tumor count (per gm)

`Liver` Liver count (per gm)

`Time` Sacrifice time (in hours)

`Treat` Treatment received

`Days` Days post inoculation

`Sex` Sex of the rat

`Weight` Initial weight (in grams)

`Loss` Weight loss (in grams)

`Tumor` Tumor weight (in 10^{-4} grams)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
str(case1102)

plot(Brain/Liver ~ Time, case1102, log="xy", pch=ifelse(Treat=="BD", 19,21))
legend(10,0.1, pch=c(19,21), c("Saline control", "Barrier disruption"))
```

---

case1201 *State Average SAT Scores*

Description

Data on the average SAT scores for US states in 1982 and possible associated factors.

Usage

```
case1201
```

Format

A data frame with 50 observations on the following 8 variables.

`State` US state

`SAT` state averages of the total SAT (verbal + quantitative) scores

`Takers` the percentage of the total eligible students (high school seniors) in the state who took the exam

`Income` the median income of families of test–takers (in hundreds of dollars)

`Years` the average number of years that the test–takers had formal studies in social sciences, natural sciences and humanities

`Public` the percentage of the test–takers who attended public secondary schools

`Expend` the total state expenditure on secondary schools (in hundreds of dollars per student)

`Rank` the median percentile ranking of the test–takers within their secondary school classes

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
str(case1201)
pairs(SAT~Rank+Years+Income+Public+Expend, case1201)
```

case1202 *Sex discrimination in Employment*

Description

Data on employees from one job category (skilled, entry–level clerical) of a bank that was sued for sex discrimination. The data are on 32 male and 61 female employees, hired between 1965 and 1975.

Usage

```
case1202
```

Format

A data frame with 93 observations on the following 7 variables.

`Bsal` Annual salary at time of hire

`Sal77` Salary as of March 1975

`Sex` Sex of employee

`Senior` Seniority (months since first hired)

`Age` Age of employee (in months)

`Educ` Education (in years)

`Exper` Work experience prior to employment with the bank (months)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Roberts, H.V. (1979). Harris Trust and Savings Bank: An Analysis of Employee Compensation, *Report 7946*, Center for Mathematical Studies in Business and Economics, University of Chicago Graduate School of Business.

See Also

case0102

Examples

```
str(case1202)
pairs(Sal77~Bsal+Senior+Age+Exper, case1202)
```

`case1301` *Seaweed Grazers*

Description

To study the influence of ocean grazers on regeneration rates of seaweed in the intertidal zone, a researcher scraped rock plots free of seaweed and observed the degree of regeneration when certain types of seaweed-grazing animals were denied access. The grazers were limpets (L), small fishes (f) and large fishes (F). Each plot received one of six treatments named by which grazers were allowed access. In addition, the researcher applied the treatments in eight blocks of 12 plots each. Within each block she randomly assigned treatments to plots. The blocks covered a wide range of tidal conditions.

Usage

```
case1301
```

Format

A data frame with 96 observations on the following 3 variables.

`Cover` percent of regenerated seaweed cover

`Block` a factor with levels `"B1"`, `"B2"`, `"B3"`, `"B4"`, `"B5"`, `"B6"`, `"B7"` and `"B8"`

`Treat` a factor indicating treatment, with levels `"C"`, `"f"`, `"fF"`, `"L"`, `"Lf"` and `"LfF"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Olson, A. (1993). Evolutionary and Ecological Interactions Affecting Seaweeds, Ph.D. Thesis. Oregon State University.

Examples

```
str(case1301)

# full two-way model with interactions
fitfull <- aov(Cover ~ Treat*Block, case1301)
# Residual plot indicates a transformation might help
plot(fitfull)

# Log of seaweed "regeneration ratio"
y <- with(case1301, log(Cover/(100-Cover)))
# Full two-way model with interactions
fitfull <- aov(y~Treat*Block, case1301)
# No problems indicated by residual plot
plot(fitfull)
# Note that interactions are not statistically significant
anova(fitfull)
# Additive model (no interactions)
```

```
fitadditive <- aov(y ~ Treat + Block, case1301)

# Make indicator variables for presence of limpets, small fish, and large fish
lmp <- with(case1301, ifelse(Treat %in% c("L", "Lf", "LfF"), 1, 0))
sml <- with(case1301, ifelse(Treat %in% c("f", "fF", "Lf", "LfF"), 1, 0))
big <- with(case1301, ifelse(Treat %in% c("fF", "LfF"), 1, 0))

fitsimple <- lm(y ~ Block + lmp + sml + big, case1301)
# Model with main effects of 3 "presence" factors seems ok.
anova(fitsimple, fitadditive)
summary(fitsimple, cor=FALSE)
```

`case1302` *Pygmalion Effect*

Description

One company of soldiers in each of 10 platoons was assigned to a Pygmalion treatment group, with remaining companies in the platoon assigned to a control group. Leaders of the Pygmalion platoons were told their soldiers had done particularly well on a battery of tests which were, in fact, non-existent. In this randomised block experiment, platoons are experimental units, companies are blocks, and average Practical Specialty test score for soldiers in a platoon is the response. The researchers wished to see if the platoon response was affected by the artificially-induced expectations of the platoon leader.

Usage

```
case1302
```

Format

A data frame with 29 observations on the following 3 variables.

`Company` a factor indicating company identification, with levels `"C1"`, `"C2"`, ..., `"C10"`

`Treat` a factor indicating treatment with two levels, `"Pygmalion"` and `"Control"`

`Score` average score on practical specialty test of all soldiers in the platoon

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Eden, D. (1990). Pygmalion Without Interpersonal Contrast Effects: Whole Groups Gain from Raising Manager Expectations, *Journal of Applied Psychology* **75**(4): 395–398.

Examples

```
str(case1302)

# two-way model with interactions
fitfull <- aov(Score ~ Company*Treat, case1302)
# No problems are indicated by residual plot
plot(fitfull)
# Interaction terms are not statistically significant
anova(fitfull)
# Additive model, with "treatment contrast" for treatment:
fitadditive <- aov(Score ~ Company + Treat, case1302)
# Interpret treatment effect as coefficient of Treat
anova(fitadditive)
```

---

case1401 *Chimp Learning Times*

Description

Researchers taught each of 4 chimps to learn 10 words in American sign language and recorded the learning time for each word for each chimp. They wished to describe chimp differences and word differences.

Usage

```
case1401
```

Format

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

`Minutes` learning time in minutes

`Chimp` a factor indicating chimp, with four levels `"Booee"`, `"Cindy"`, `"Bruno"` and `"Thelma"`

`Sign` a factor indicating word taught, with 10 levels

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Fouts, R.S. (1973). Acquisition and Testing of Gestural Signs in Four Young Chimpanzees, *Science* **180**: 978-980.

Examples

```
str(case1401)

fitadditive <- aov(Minutes ~ Chimp + Sign, case1401)
# Residual plot indicates a transformation may help
plot(fitadditive)

fitadditive <- aov(log(Minutes) ~ Chimp + Sign, case1401)
# No problems are indicated by residual plot
plot(fitadditive)
anova(fitadditive)

# Tukey multiple comparisons of sign differences
mcSign <- TukeyHSD(fitadditive,"Sign")
mcSign
plot(mcSign)
mcChimp <- TukeyHSD(fitadditive,"Chimp")
mcChimp
par(cex=.7)
plot(mcChimp)
```

---

case1402 *Effect of Ozone, SO2 and Drought on Soybean Yield*

Description

In a completely randomized design with a 2x3x5 factorial treatment structure, researchers randomly assigned one of 30 treatment combinations to open-topped growing chambers, in which two soybean cultivars were planted. The responses for each chamber were the yields of the two types of soybean.

Usage

```
case1402
```

Format

A data frame with 30 observations on the following 5 variables.

`Stress` a factor indicating treatment, with two levels `"Well-watered"` and `"Stressed"`

`SO2` a quantitative treatment with three levels 0, 0.02 and 0.06

`O3` a quantitative treatment with five levels 0.02, 0.05, 0.07, 0.08 and 0.10

`Forrest` the yield of the Forrest cultivar of soybean (in kg/ha)

`William` the yield of the Williams cultivar of soybean (in kg/ha)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Heggestad, H.E. and Lesser, V.M. (1990). Effects of Chronic Doses of Sulfur Dioxide, Ozone, and Drought on Yields and Growth of Soybeans Under Field Conditions, *Journal of Environmental Quality* **19**: 488–495.

Examples

```
str(case1402)

plot(Forrest ~ O3,  case1402, log="y", pch=ifelse(Stress=="Stressed",19,21))
plot(Forrest ~ SO2, case1402, log="y", pch=ifelse(Stress=="Stressed",19,21))

fitbig <- lm(log(Forrest) ~  O3*SO2*Stress, case1402)
# Residual plot does not indicate any problem.
plot(fitbig)
# The 3-factor interaction is not statistically significant.
anova(fitbig)
# Drop the three-factor interaction
fit2 <- update(fitbig, ~ . - O3:SO2:Stress)
anova(fit2)

fitadditive <- lm(log(Forrest) ~ O3 + SO2 + Stress, case1402)
summary(fitadditive)
```

---

case1501 *Logging and Water Quality*

Description

Data from an observational study of nitrate levels measured at three week intervals for five years in two watersheds. One of the watersheds was undisturbed and the other had been logged with a patchwork pattern.

Usage

```
case1501
```

Format

A data frame with 88 observations on the following 3 variables.

`Week` week after the start of the study

`Patch` residual nitrate level in the logged watershed (ppm) (see Display 15.3 of Statistical Sleuth)

`Nocut` residual nitrate level in the undisturbed watershed (ppm)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Harr, R.D., Friderksen, R.L., and Rothacher, J. (1979). Changes in Streamflow Following Timber Harvests in Southwestern Oregon, USDA/USFS Research Paper PNW-249, Pacific NW Forest and Range Experiment Station, Portland, Oregon.

Examples

```
str(case1501)

par(mfrow=c(2,1)) # Make 2 plots on one page
plot(Nocut ~ Week, case1501)
plot(Patch ~ Week, case1501)
par(mfrow=c(1,1))
lag.plot(case1501$Nocut,do.lines=FALSE)
lag.plot(case1501$Patch,do.lines=FALSE)

# Compute pooled estimate of first autocorrelation coefficient
# First auto covariance, Nocut
ac1nocut <- acf(case1501$Nocut,lag.max=1,type="covariance",plot=FALSE)$acf[2]
n <- length(case1501$Nocut)
# Zeroth autocovariance for Nocut
ac0nocut <- var(case1501$Nocut[2:n])*(n-2)/(n-1)
# First auto covariance, Patch
ac1patch <- acf(case1501$Patch,lag.max=1,type="covariance",plot=FALSE)$acf[2]
# Zeroth autocovariance for PATCH
ac0patch <- var(case1501$Patch [2:n])*(n-2)/(n-1)

ac1pool <- (ac1nocut + ac1patch)/2
ac0pool <- (ac0nocut + ac0patch)/2

acorr1 <- ac1pool/ac0pool
acorr1  # Pooled estimate of first lag serial coefficient
```

---

case1502 *Global Warming*

Description

The data are the temperatures (in degrees Celsius) averaged for the northern hemisphere over a full year, for years 1880 to 1987. The 108-year average temperature has been subtracted, so each observation is the temperature difference from the series average.

Usage

```
case1502
```

Format

A data frame with 108 observations on the following 2 variables.

`Year` year in which yearly average temperature was computed, from 1880 to 1987

`Temp` northern hemisphere temperature minus the 108-year average (degrees Celsius)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Jones, P.D. (1988). Hemispheric Surface Air Temperature Variations—Recent Trends Plus an Update to 1987, *Journal of Climatology* **1**: 654–660.

Examples

```
str(case1502)

# Residuals from regression fit, ignoring autocorrelation
resids <- lm(Temp ~ Year, case1502)$res
# PACF plot shows evidence of 1st order auto correlation
acf(resids,type="partial")
# 1st autocorrelation coef.
acorr1 <- acf(resids,type="correlation",plot=FALSE)$acf[2]

# Fit regression with filtered response and explanatory variables:
n <- length(case1502$Temp)
y <- with(case1502, Temp [2:n] - acorr1* Temp [1:(n-1)])
x <- with(case1502, Year [2:n] - acorr1* Year [1:(n-1)])
fit <- lm(y ~ x)
summary(fit)  # Interpret coefficient of x as coefficient of Year
```

---

case1601 *Sites of Short- and Long-Term Memory*

Description

Researchers taught 18 monkeys to distinguish each of 100 pairs of objects, 20 pairs each at 16, 12, 8, 4, and 2 weeks prior to a treatment. After this training, they blocked access to the hippocampal formation in 11 of the monkeys. All monkeys were then tested on their ability to distinguish the objects. The five-dimensional response for each monkey is the number of correct objects distinguished among those taught at 16, 12, 8, 4, and 2 weeks prior to treatment.

Usage

```
case1601
```

Format

A data frame with 18 observations on the following 7 variables.

`Monkey` Monkey name

`Treatment` a treatment factor with levels `"Control"` and `"Treated"`

`Week2` percentage of 20 objects taught 2 weeks prior to treatment that were correctly distinguished in the test

`Week4` percentage of 20 objects taught 4 weeks prior to treatment that were correctly distinguished in the test

`Week8` percentage of 20 objects taught 8 weeks prior to treatment that were correctly distinguished in the test

`Week12` percentage of 20 objects taught 12 weeks prior to treatment that were correctly distinguished in the test

`Week16` percentage of 20 objects taught 16 weeks prior to treatment that were correctly distinguished in the test

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Sola-Morgan, S. M. and Squire, L. R. (1990). The Primate Hippocampal Formation: Evidence for a Time-limited Role in Memory Storage, *Science* **250**: 288–290.

Examples

```
str(case1601)

# short-term response
short <- with(case1601, (Week2 + Week4)/2)
# long-term response
long <- with(case1601, (Week8 + Week12 + Week16)/3)
# Multivariate analysis of variance
mfit <- manova(cbind(short,long) ~ Treatment, case1601)
summary(mfit)
```

---

case1602 *Oat Bran and Cholesterol*

Description

In a randomized, double-blind, crossover experiment, researchers randomly assigned 20 volunteer hospital employees to either a low-fiber or low-fiber treatment group. The subjects followed the diets for six weeks. After two weeks on their normal diet, all patients crossed over to the other treatment group for another six weeks. The total serum cholesterol (in mg/dl) was measured on each patient before the first treatment, at the end of the first six week treatment, and at the end of the second six week treatment.

Usage

```
case1602
```

Format

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

`Baseline` total serum cholesterol before treatment

`Hifiber` total serum cholesterol after the high fiber diet

`Lofiber` total serum cholesterol after the low fiber diet

`Order` factor to identify order of treatment, with two levels `"HL"` and `"LH"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Swain, J.F., Rouse, I.L., Curley, C.B., and Sacks, F.M. (1990). Comparison of the Effects of Oat Bran and Low-fiber Wheat on Serum Lipoprotein Levels and Blood Pressure, *New England Journal of Medicine* **320**: 1746–1747.

Examples

```
str(case1602)

subjects <- 1:20
ordersubjects <- order(case1602$Baseline)
plot(1:20, case1602$Baseline[ordersubjects], pch=24,
     xlab="Subjects (Ordered According to Baseline Cholesterol)",
     ylab="Total Serum Cholesterol (mg/dl)")
points(1:20, case1602$Lofiber[ordersubjects], pch=19, col=5)
points(1:20, case1602$Hifiber[ordersubjects], pch=21, col=3)
legend(1,245,legend=c("Baseline","After Low Fiber Diet","After High Fiber Diet"),pch=c(24

diff <- with(case1602, Hifiber-Lofiber)
plot(subjects, diff, pch=ifelse(case1602$Order=="HL",19,21))
abline(h=0)
t.test(diff ~ Order, case1602) # Test for order of treatment effect
t.test(diff) # Test for treatment effect
```

---

case1701 *Magnetic Force on Printer Rods*

Description

Engineers manipulated three factors (with 3, 2, and 4 levels each) in the construction and operation of printer rods, to see if they influenced the magnetic force around the rod.

Usage

```
case1701
```

Format

A data frame with 44 observations on the following 14 variables.

Name	Description
L1, L2,..., L11	the magnetic force at each of the equally-spaces positions 1, 2, ..., 11 on the printer rod
Current	electric current passing through the rod, with three levels "0", "250" and "500" (milliamperes)
Configur	a factor identifying the configuration, with two levels "0" and "1"
Material	a factor identifying the type of metal from which the rod was made, with four levels "1", "2", "3" a

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
str(case1701)

pca <- princomp(case1701[,1:11])
summary(pca)
# The first 3 principal components account for 99.7% of the variation
screeplot(pca)
# The loadings suggest the following meaningful summaries...
loadings(pca)

overallaverage <- with(case1701, (L1 + L2 + L3 + L4 + L5 + L6 + L7 + L8 + L9 + L10 + L11)
rightleftdiff <- with(case1701, (L9 + L10 + L11)/3 - (L1 + L2 + L3)/3)
middleleftdiff <- with(case1701, L6 - (L1 + L2)/2)

# Note 4 clusters and 1 outlier
pairs(cbind(overallaverage, rightleftdiff, middleleftdiff))

fit1 <- lm(overallaverage ~ Current*Configur*Material, case1701)
anova(fit1)
```

---

case1702 *Love and Marriage*

Description

Thirty couples participated in a study of love and marriage. Wives and husbands responded separately to four questions:

1. What is the level of passionate love you feel for your spouse?
2. What is the level of passionate love your spouse feels for you?
3. What is the level of compassionate love you feel for your spouse?
4. What is the level of compassionate love your spouse feels for you?

Each response was recorded on a five-point scale: 1=None, 2=Very Little, 3=Some, 4=A Great Deal and 5=A Tremendous Amount.

Usage

```
case1702
```

Format

A data frame with 30 observations on the following 9 variables.

`Couple` couple identification number

`Hps` level of passionate love husband feels for spouse

`Wps` level of passionate love wife feels for spouse

`Hcs` level of compassionate love husband feels for spouse

`Wcs` level of compassionate love wife feels for spouse

`Hpy` level of passionate love husband perceives spouse to have for him

`Wpy` level of passionate love wife perceives spouse to have for her

`Hcy` level of compassionate love husband perceives spouse to have for him

`Wcy` level of compassionate love husband perceives spouse to have for her

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Johnson, R.A. and Wichern, D.W. (1988). *Applied Multivariate Statistical Analysis (2nd ed)*, Prentice-Hall.

Examples

```
str(case1702)

# feelings about spouse
tospouse <- with(case1702, cbind(Hps, Wps, Hcs, Wcs))
# perceived feelings from spouse
fromspouse <- with(case1702, cbind(Hpy, Wpy, Hcy, Wcy))
cca <- cancor(tospouse,fromspouse)
# Examine loadings of first canonical variables:
par(mfrow=c(2,1))
barplot(cca$xcoef[,1], ylab="first 'to spouse' loadings",
     names=c("Hps","Wps","Hcs","Wcs"))
barplot(cca$ycoef[,1], ylab="first 'from spouse' loadings",
     names=c("Hpy","Wpy","Hcy","Wcy"))

# The first canonical variable for 'to spouse" is mostly Hcs
# The first canonical variable for 'fom spouse' is mostly Hcy

can.to <- tospouse
can.from <- fromspouse
can.to.1 <- can.to[,1] # first canonical variable
can.from.1 <- can.from[,1] # first canonical variable
pairs(cbind(can.to.1, case1702$Hcs, can.from.1, case1702$Hcy),
labels=c("1st cv 'to'","husband's compassionate","1st cv
        'from'","husband's perceived compassionate"))
```

---

case1902 *Death Penalty and Race*

Description

Lawyers collected data on convicted murderers in the state of Georgia to see whether convicted murderers who were black were more likely to receive the death penalty than convicted murderers who were white, after accounting for aggravation level of the murder. They categorized murders into 6 progressively more serious types. Category 1 comprises barroom brawls, liquor-induced arguments lovers' quarrels, and similar crimes. Category 6 includes the most vicious, cruel, cold=blooded, unprovoked crimes.

Usage

```
case1902
```

Format

A data frame with 12 observations on the following 4 variables.

`Aggravation` the aggravation level of the crime, a factor with levels `"1"`, `"2"`, `"3"`, `"4"`, `"5"` and `"6"`

`Victim` a factor indicating race of murder victim, with levels `"White"` and `"Black"`

`Death` number in the aggravation and victim category who received the death penalty

`Nodeath` number in the aggravation and victim category who did not receive the death penalty

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Woodworth, G.C. (1989). Statistics and the Death Penalty, *Stats* **2**: 9–12.

Examples

```
str(case1902)

# Add smidgeon to denominator because of zeros
empiricalodds <- with(case1902, Death/(Nodeath + .5))
plot(empiricalodds ~ as.numeric(Aggravation), case1902, log="y", pch=ifelse(Victim=="Whit
legend(3.8,.02,legend=c("White Victim Murderers","Black Victim Murderers"),pch=c(21,19))

fitbig <- glm(cbind(Death,Nodeath) ~ Aggravation*Victim, case1902, family=binomial)
# No evidence of overdispersion; no statistically significant evidence
# of interactive effect
anova(fitbig, test="Chisq")
fitlinear <- glm(cbind(Death,Nodeath) ~ Aggravation + Victim, case1902, family=binomial)
summary(fitlinear)

# Mantel Haenszel Test, as an alternative
table1902   <- with(case1902, rbind(Death,Nodeath))
dim(table1902) <- c(2,2,6)
mantelhaen.test(table1902)
```

case2001 *Survival in the Donner Party*

Description

This data frame contains the ages and sexes of the adult (over 15 years) survivors and nonsurvivors of the Donner party.

Usage

```
case2001
```

Format

A data frame with 45 observations on the following 3 variables.

`Age` Age of person

`Sex` Sex of person

`Status` Whether the person survived or died

Details

In 1846 the Donner and Reed families left Springfield, Illinois, for California by covered wagon. In July, the Donner Party, as it became known, reached Fort Bridger, Wyoming. There its leaders decided to attempt a new and untested rote to the Sacramento Valley. Having reached its full size of 87 people and 20 wagons, the party was delayed by a difficult crossing of the Wasatch Range and again in the crossing of the desert west of the Great Salt Lake. The group became stranded in the eastern Sierra Nevada mountains when the region was hit by heavy snows in late October. By the time the last survivor was rescued on April 21, 1847, 40 of the 87 members had died from famine and exposure to extreme cold.

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Grayson, D.K. (1990). Donner Party Deaths: A Demographic Assessment, *Journal of Anthropological Research* **46**: 223–242.

See Also

ex1918

Examples

```
str(case2001)
```

case2002 *Birdkeeping and Lung Cancer*

Description

A 1972–1981 health survey in The Hague, Netherlands, discovered an association between keeping pet birds and increased risk of lung cancer. To investigate birdkeeping as a risk factor, researchers conducted a *case–control* study of patients in 1985 at four hospitals in The Hague (population 450,000). They identified 49 cases of lung cancer among the patients who were registered with a general practice, who were age 65 or younger and who had resided in the city since 1965. They also selected 98 controls from a population of residents having the same general age structure.

Usage

```
case2002
```

Format

A data frame with 147 observations on the following 7 variables.

- `LC` Whether subject has lung cancer
- `FM` Sex of subject
- `SS` Socioeconomic status, determined by occupation of the household's principal wage earner
- `BK` Indicator for birdkeeping (caged birds in the home for more that 6 consecutive months from 5 to 14 years before diagnosis (cases) or examination (control))
- `AG` Age of subject (in years)
- `YR` Years of smoking prior to diagnosis or examination
- `CD` Average rate of smoking (in cigarettes per day)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Holst, P.A., Kromhout, D. and Brand, R. (1988). For Debate: Pet Birds as an Independent Risk Factor for Lung Cancer, *British Medical Journal* **297**: 13–21.

Examples

```
str(case2002)
```

`case2101` *Island Size and Bird Extinctions*

Description

In a study of the Krunnit Islands archipelago, researchers presented results of extensive bird surveys taken over four decades. They visited each island several times, cataloguing species. If a species was found on a specific island in 1949, it was considered to be at risk of extinction for the next survey of the island in 1959. If it was not found in 1959, it was counted as an "extinction", even though it might reappear later. This data frame contains data on island size, number of species at risk to become extinct and number of extinctions.

Usage

```
case2101
```

Format

A data frame with 18 observations on the following 4 variables.

`Island` Name of Island

`Area` Area of Island

`Atrisk` Number of species at risk

`Extinct` Number of extinctions

Details

Scientists agree that preserving certain habitats in their natural states is necessary to slow the accelerating rate of species extinctions. But they are divided on how to construct such reserves. Given a finite amount of available land, is it better to have many small reserves or a few large one? Central to the debate on this question are observational studies of what has happened in island archipelagos, where nearly the same fauna tries to survive on islands of different sizes.

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

V\"ais\"anen, R.A. and J\"arvinen, O. (1977). Dynamics of Protected Bird Communities in a Finnish Archipelago, *Journal of Animal Ecology* **46**: 891–908.

Examples

```
str(case2101)
logit <- function(p) log(p/(1-p))
plot(logit(Extinct/Atrisk) ~ log(Area), case2101)
```

`case2102` *Moth Coloration and Natural Selection*

Description

This data was collected by J.A. Bishop. Bishop selected seven locations progressively farther from Liverpool. At each location, Bishop chose eight trees at random. Equal number of dead (frozen) light (*Typicals*) and dark (*Carbonaria*) moths were glued to the trunks in lifelike positions. After 24 hours, a count was taken of the numbers of each morph that had been removed—presumably by predators.

Usage

```
case2102
```

Format

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

`Morph` Morph, a factor with levels `"light"` and `"dark"`

`Distance` Distance from Liverpool (in km)

`Placed` Number of moths placed

`Removed` Number of moths removed

Details

Population geneticists consider clines particularly favourable situations for investigating evolutionary phenomena. A cline is a region where two colour morphs of one species arrange themselves at opposite ends of an environmental gradient, with increasing mixtures occurring between. Such a cline exists near Liverpool, England, where a dark morph of a local moth has flourished in response to the blackening of tree trunks by air pollution from the mills. The moths are nocturnal, resting during the day on tree trunks, where their coloration acts as camouflage against predatory birds. In Liverpool, where tree trunks are blackened by smoke, a high percentage of the moths are of the dark morph. One encounters a higher percentage of the typical (pepper–and–salt) morph as one travels from the city into the Welsh countryside, where tree trunks are lighter. J.A. Bishop used this cline to study the intensity of natural selection.

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Bishop, J.A. (1972). An Experimental Study of the Cline of Industrial Melanism in *Biston betularia* [Lepidoptera] Between Urban Liverpool and Rural North Wales, *Journal of Animal Ecology* **41**: 209–243.

Examples

```
str(case2102)
```

case2201 *Age and Mating Success of Male Elephants*

Description

Although male elephants are capable of reproducing by 14 to 17 years of age, your adult males are usually unsuccessful in competing with their larger elders for the attention of receptive females. Since male elephants continue to grow throughout their lifetimes, and since larger males tend to be more successful at mating, the males most likely to pass their genes to future generations are those whose characteristics enable them to live long lives. Joyce Poole studied a population of African elephants in Amboseli National Park, Kenya, for 8 years. This data frame contains the number of successful matings and ages (at the study's beginning) of 41 male elephants.

Usage

```
case2201
```

Format

A data frame with 41 observations on the following 2 variables.

`Age` Age of elephant at beginning of study

`Matings` Number of successful matings

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Poole, J.H. (1989). Mate Guarding, Reproductive Success and Female Choice in African Elephants, *Animal Behavior* **37**: 842–849.

Examples

```
str(case2201)
plot(case2201)
```

case2202 *Characteristics Associated with Salamander Habitat*

Description

The Del Norte Salamander (*plethodon elongates*) is a small (5–7 cm) salamander found among rock rubble, rock outcrops and moss-covered talus in a narrow range of northwest California. To study the habitat characteristics of the species and particularly the tendency of these salamanders to reside in dwindling old-growth forests, researchers selected 47 sites from plausible salamander habitat in national forest and parkland. Randomly chosen grid points were searched for the presence of a site with suitable rocky habitat. At each suitable site, a 7 metre by 7 metre search are was examined for the number of salamanders it contained. This data frame contains the counts of salamanders at the sites, along with the percentage of forest canopy and age of the forest in years.

Usage

```
case2202
```

Format

A data frame with 47 observations on the following 4 variables.

`Site` Investigated site

`Salaman` Number of salamanders found in 49 m^2 area

`PctCover` Percentage of canopy cover

`Forestage` Forest age

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Welsh, H.H. and Lind, A.J. (1995). *Journal of Herpetology* **29**(2): 198–210.

Examples

```
str(case2202)
```

---

ex0112 *Fish Oil and Blood Pressure*

Description

Researchers used 7 red and 7 black playing cards to randomly assign 14 volunteer males with high blood pressure to one of two diets for four weeks: a fish oil diet and a standard oil diet. These data are the reductions in diastolic blood pressure.

Usage

```
ex0112
```

Format

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

`BP` reduction in diastolic blood pressure (in mm of mercury)

`Diet` factor variable indicating the diet that the subject followed

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Knapp, H.R. and FitzGerald, G.A. (1989). The Antihypertensive Effects of Fish Oil, *New England Journal of Medicine* **320**: 1037–1043.

Examples

```
str(ex0112)
```

`ex0116` *Planet Distances and Order from Sun*

Description

The data are the distances from the sun (scaled so that earth=10) and the order from the sun for the 9 planets in our solar system plus the asteroid belt (treated here as the fifth body from the sun).

Usage

```
ex0116
```

Format

A data frame with 10 observations on the following 3 variables.

`Planet` name of body (planet or asteroid belt)

`Order` order from sun

`Distance` distance from sun (scaled so that earth's distance is 10)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
str(ex0116)
```

ex0211 *Lifetimes of Guinea Pigs*

Description

The data are survival times (in days) of guinea pigs that were randomly assigned either to a control group or to a treatment group that received a dose of tubercle bacilli.

Usage

```
ex0211
```

Format

A data frame with 122 observations on the following 2 variables.

`Lifetime` survival time of guinea pig (in days)

`Group` a factor with levels `"bacilli"` and `"control"`, indicating the group to which the guinea pig was assigned

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Doksum, K. (1974). Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two–sample Case, *Annals of Statistics* **2**: 267–277.

Examples

```
str(ex0211)
```

ex0221 *Bumpus's Data on Natural Selection (Weight)*

Description

As evidence in support of natural selection, Bumpus presented measurements on house sparrows brought to the Anatomical Laboratory of Brown University after an uncommonly severe winter storm. Some of these birds had survived and some had perished. Bumpus asked whether those that perished did so because they lacked physical characteristics enabling them to withstand the intensity of that particular instance of selective elimination. The data are on the the weights, in grams, for the 24 adult male sparrows that perished and for the 35 adult males that survived.

Usage

```
ex0221
```

Format

A data frame with 59 observations on the following 2 variables.

`Weight` weight of adult male sparrows (in grams)

`Status` factor variable indicating whether the sparrow perished or survived in a winter storm

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

See Also

`case0201`, `ex2016`

Examples

```
str(ex0221)
```

---

ex0222 *Cholesterol in Urban and Rural Guatemalans*

Description

This data comes from an observational study to contrast cholesterol levels in rural and urban Guatemalan Indians

Usage

```
ex0222
```

Format

A data frame with 94 observations on the following 2 variables.

`Cholesterol` Serum total cholesterol of individual (in mg/l)

`Group` a factor with levels `"Rural"` and `"Urban"` indicating to which group the individual belongs

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Tejada, C., Charm, S., Guzman, M., Mendez, J. and Kurland, G. (1964). The Blood Viscosity of Various Socioeconomic Groups in Guatemala, *American Journal of Clinical Nutrition*: 303–308.

Examples

```
str(ex0222)
```

ex0223 *Speed Limits and Traffic Fatalities*

Description

The National Highway System Designation Act was signed into law in the United States on November 28, 1995. Among other things, the act abolished the federal mandate of 55 mile per hour maximum speed limits on roads in the United States and permitted states to establish their own limits. Of the 50 states (plus the District of Columbia), 32 increased their speed limits at the beginning of 1996 or sometime during 1996. These data are the percentage changes in interstate highway traffic fatalities from 1995 to 1996.

Usage

```
ex0223
```

Format

A data frame with 51 observations on the following 3 variables.

`State` US state

`Increase` a factor with levels `"No"` `"Yes"`, indicating whether the state increased its speed limit

`FatalitiesChange` percentage change in interstate traffic fatalities between 1995 and 1996

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Report to Congress: The Effect of Increased Speed Limits in the Post-NMSL Era, National Highway Traffic Safety Administration, February, 1998; available in the reports library at http://www-fars.nhtsa.dot.gov/.

Examples

```
str(ex0223)
```

ex0321 *Umpire Life Lengths*

Description

Researchers collected historical and current data on umpires to investigate their life expectancies following the collapse and death of a U.S. major league baseball umpire. They were investigating speculation that stress associated with the job posed a health risk. Data were found on 227 umpires who had died or had retired and were still living. The data set includes the dates of birth and death.

Usage

```
ex0321
```

Format

A data frame with 227 observations on the following 3 variables.

`Lifelength` observed lifetime for those umpires who had died by the time of the study or current age of those still living

`Censored` 0 for those who had died by the time of the study or 1 for those who were still living

`Expected` length from actuarial life tables for individuals who were alive at the time the person first became an umpire

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Cohen, R.S., Kamps, C.A., Kokoska, S., Segal E.M. and Tucker, J.B.(2000). Life Expectancy of Major League Baseball Umpires, *The Physician and Sportsmedicine* **28**(5): 83–89.

Examples

```
str(ex0321)
```

ex0323 *Solar Radiation and Skin Cancer*

Description

Data contains yearly skin cancer rates (per 100,000 people) in Connecticut from 1938 to 1972 with a code indicating those years that came two years after higher than average sunspot activity and those years that came two years after lower than average sunspot activity.

Usage

```
ex0323
```

Format

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

`Year` year

`Rate` skin cancer rate per 100,000 people

`Sunspot` a factor with levels `"High"` and `"Low"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Andrews, D.F. and Herzberg, A.M. (1985). *Data: A Collection of Problems from many Fields for the Student and Research Worker*, Springer-Verlag.

Examples

```
str(ex0323)
```

---

ex0327 *Life Expectancy and Per Capita Income*

Description

Life expectancy and per capita income for 20 industrialized countries and 9 petroleum exporting countries. Note that there is a missing value for South Africa.

Usage

```
ex0327
```

Format

A data frame with 29 observations on the following 4 variables.

`Country` a character vector indicating the country

`Life` life expectancy (years)

`Income` income in 1974 (U.S. dollars)

`Type` factor variable with levels `"Industrialized"` and `"Petroleum"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Leinhardt, S. and Wasserman, S.S. (1979). Teaching Regression: An Exploratory Approach, *The American Statistician* **33**(4): 196–203.

Examples

```
str(ex0327)
```

`ex0328` *Pollen Removal*

Description

As part of a study to investigate reproductive strategies in plants, biologists recorded the time spent at sources of pollen and the proportions of pollen removed by bumblebee queens and honeybee workers pollinating a species of lily.

Usage

```
ex0328
```

Format

A data frame with 47 observations on the following 3 variables.

`Removed` proportion of pollen removed

`Duration` duration of visit (in seconds)

`Bee` factor variable with levels `"Queen"` and `"Worker"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Harder, L.D. and Thompson, J.D. (1989). Evolutionary Options for Maximizing Pollen Dispersal of Animal-pollinated Plants, *American Naturalist* **133**: 323–344.

Examples

```
str(ex0328)
```

ex0331 *Iron Supplementation*

Description

A randomized experiment was performed on mice to determine whether two forms of iron are retained differently. If one type is retained especially well it may be more useful as a dietary supplement for humans.

Usage

```
ex0331
```

Format

A data frame with 36 observations on the following 2 variables.

`Iron` percentage of iron retained in each mouse

`Supplement` factor variable with levels `"Fe3"` and `"Fe4"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Rice, J. (1987). *Mathematical Statistics and Data Analysis*, Wadsworth.

Examples

```
str(ex0331)
```

ex0332 *College Tuition*

Description

Tuition in dollars of 20 private and 20 public U.S. colleges and universities for 1993–1994.

Usage

```
ex0332
```

Format

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

`Private` tuition in dollars of 20 private schools

`PubIn` tuition in dollars of 20 public schools (in-state tuition)

`PubOut` tuition in dollars of 20 public schools (out-of-state tuition)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

1995 U.S. News and World Report's Guide to America's Best Colleges.

Examples

```
str(ex0332)
```

`ex0333` *Brain Size and Litter Size*

Description

Relative brain weights for 51 species of mammal whose average litter size is less than 2 and for 45 species of mammal whose average litter size is greater than or equal to 2.

Usage

```
ex0333
```

Format

A data frame with 96 observations on the following 2 variables.

`Brainsize` relative brain sizes (1000 * Brain weight/Body weight) for 96 species of mammals

`Littersize` factor variable with levels `"Small"` and `"Large"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Sacher, G.A. and Staffeldt, E.F. (1974). Relation of Gestation Time to Brain Weight for Placental Mammals: Implications for the Theory of Vertebrate Growth, *American Naturalist* **108**: 593–613.

See Also

`case0902`

Examples

```
str(ex0333)
```

ex0428 *Darwin's Data*

Description

Plant heights (inches) for 15 pairs of plants of the same age, one of which was grown from a seed from a cross-fertilized flower and the other of which was grown from a seed from a self-fertilized flower.

Usage

```
ex0428
```

Format

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

`Cross` height (inches) of cross-fertilized plant

`Self` height (inches) of self-fertilized plant

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Andrews, D.F. and Herzberg, A.M. (1985). *Data: A Collection of Problems from many Fields for the Student and Research Worker*, Springer-Verlag.

Examples

```
str(ex0428)
```

ex0429 *Exercise and Walking Time*

Description

Can active exercise shorten the time it takes an infant to walk alone? Twelve, one week old, male infants from white, middle-class families were randomly allocated to one of two treatment groups. Those in the active-exercise group received stimulation of the walking reflexes during four 3 minute sessions each day from the beginning of the second through the end of the eighth week. Those in the other group received no stimulation.

Usage

```
ex0429
```

Format

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

`Age` age (months) at which infants first walked alone

`Exercise` a factor with levels `"Active"` and `"None"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Zelazo, P.R. (1972). Walking in the Newborn, *Science* **176**: 314–315.

Examples

```
str(ex0429)
```

---

ex0430 *Sunlight Protection Factor*

Description

Tolerance to sunlight (in minutes) for 13 patients prior to and after treatment with a sunscreen.

Usage

```
ex0430
```

Format

A data frame with 13 observations on the following 2 variables.

`Control` tolerance to sunlight (minutes) prior to sunscreen application

`Sunscreen` tolerance to sunlight (minutes) after sunscreen application

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Fusaro, R.M. and Johnson, J.A. (1974). Sunlight Protection for Erythropoietic Protoporphyria Patients, *Journal of the American Medical Association* **229**(11): 1420.

Examples

```
str(ex0430)
```

ex0431 *Effect of Group Therapy on Survival of Breast Cancer Patients*

Description

Researchers randomly assigned metastatic breast cancer patients to either a control group or a group that received weekly 90 minute sessions of group therapy and self-hypnosis, to see whether the latter treatment improved the patients' quality of life.

Usage

```
ex0431
```

Format

A data frame with 58 observations on the following 3 variables.

`Survival` months of survival after beginning of study

`Group` a factor with levels `"Control"` and `"Therapy"`

`Censor` 0 if entire lifetime observed, 1 if patient known to have lived at least 122 months

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Spiegel, D., Bloom, J.R., Kraemer, H.C. and Gottheil, E. (1989). Effect of Psychosocial Treatment on Survival of Patients with Metastatic Breast Cancer, *Lancet* **334**(8668): 888–891.

Examples

```
str(ex0431)
```

ex0432 *Therapeutic Marijuana*

Description

To investigate the capacity of marijuana to reduce the side effects of cancer chemotherapy, researchers performed a double-blind, randomized, crossover trial. Fifteen cancer patients on chemotherapy were randomly assigned to receive either a marijuana treatment or a placebo treatment after their first three sessions of chemotherapy. They were then crossed over to the opposite treatment for their next 3 sessions.

Usage

```
ex0432
```

Format

A data frame with 15 observations on the following 3 variables.

`Subject` subject number 1–15

`Marijuana` total number of vomiting and retching episodes under marijuana treatment

`Placebo` total number of vomiting and retching episodes under placebo treatment

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Chang, A.E., Shiling, D.J., Stillman, R.C., Goldberg, N.H., Seipp, C.A., Barofsky, I., Simon, R.M. and Rosenberg, S.A. (1979). Delta-9-Tetrahydrocannabinol as an Antiemetic in Cancer Patients Receiving High Dose Methotrexate, *Annals of Internal Medicine* **91**(6): 819–824.

Examples

```
str(ex0432)
```

---

ex0518 *Fatty Acid*

Description

A randomized experiment was performed to estimate the effect of a certain fatty acid CPFA on the level of a certain protein in rat livers.

Usage

```
ex0518
```

Format

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

`Protein` levels of protein (x 10) found in rat livers

`Treatment` a factor with levels `"Control"`, `"CPFA50"`, `"CPFA150"`, `"CPFA300"`, `"CPFA450"` and `"CPFA600"`

`Day` a factor with levels `"Day1"`, `"Day2"`, `"Day3"`, `"Day4"` and `"Day5"`

`Group` a factor with levels `"Group1"`, `"Group2"`, ..., `"Group10"`; the observed levels of the `Treatment` and `Day` interaction

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
str(ex0518)
```

ex0523 *Was Tyrannosaurus Rex Warm-Blooded?*

Description

Data frame with measurements of oxygen isotopic composition of vertebrate bone phosphate (per mil deviations from SMOW) in 12 bones of a singe Tyrannosaurus rex specimen

Usage

```
ex0523
```

Format

A data frame with 52 observations on the following 2 variables.

`Oxygen` oxygen isotopic composition

`Bonegrp` a factor with levels `"Bone1"`, `"Bone2"`, ..., `"Bone12"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Barrick, R.E. and Showers, W.J. (1994). Thermophysiology of *Tyrannosaurus rex*: Evidence from Oxygen Isotopes, *Science* **265**(5169): 222–224.

See Also

ex1120

Examples

```
str(ex0523)
```

ex0524 *Vegetarians and Zinc: An Observational Study*

Description

Previous studies suggest that vegetarians may not receive enough zinc in their diets and the zinc requirement is especially important during pregnancy. Twenty-three women were monitored: twelve vegetarians who were pregnant, six nonvegetarians who were pregnant, and five vegetarians who were not pregnant. Is there any evidence that pregnant vegetarians tend to have lower zinc levels than pregnant nonvegetarians?

Usage

```
ex0524
```

Format

A data frame with 23 observations on the following 2 variables.

`Zinc` levels of Zinc (μg/g) in the hair of women

`Group` a factor with levels `"PregNonVeg"`, `"PregVeg"` and `"NonPregVeg"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

King, J.C., Stein, T. and Doyle, M. (1981). Effect of Vegetarianism on the Zinc Status of Pregnant Women, *American Journal of Clinical Nutrition* **34**(6): 1049–1055.

Examples

```
str(ex0524)
```

---

ex0621 *Failure Times of Bearings*

Description

Data consist of times to fatigue failure (in units of millions of cycles) for 10 high-speed turbine engine bearings made from five different compounds.

Usage

```
ex0621
```

Format

A data frame with 50 observations on the following 2 variables.

`Time` failure times of bearings (millions of cycles)

`Compound` a factor with levels `"I"`, `"II"`, `"III"`, `"IV"` and `"V"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

McCool, J.I. (1979). Analysis of Single Classification Experiments Based on Censored Samples from the Two-parameter Weibull Distribution, *Journal of Statistical Planning and Inference* **3**(1): 39–68.

Examples

```
str(ex0621)
```

ex0622 *A Biological Basis for Homosexuality*

Description

Is there a physiological basis for sexual preference? Researchers measured the volumes of four cell groups in the interstitial nuclei of the anterior hypothalamus in postmortem tissue from 41 subjects at autopsy from seven metropolitan hospitals in New York and California.

Usage

```
ex0622
```

Format

A data frame with 41 observations on the following 2 variables.

`Volume` volumes of INAH3 ($1000 \times \text{mm}^3$) cell clusters from 41 humans

`Group` a factor with levels

`"Group1"`	heterosexual male with AIDS death
`"Group2"`	heterosexual male with Non-AIDS death
`"Group3"`	homosexual male with AIDS death
`"Group4"`	heterosexual female with AIDS death
`"Group5"`	heterosexual female with Non-AIDS death

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

LeVay, S. (1991). A Difference in Hypothalamic Structure Between Heterosexual and Homosexual Men, *Science* **253**(5023): 1034–1037.

Examples

```
str(ex0622)
```

ex0723 *Old Faithful*

Description

Old Faithful Geyser in Yellowstone National Park, Wyoming, derives its name and its considerable fame from the regularity (and beauty) of its eruptions. As they do with most geysers in the park, rangers post the predicted tiems of eruptions on signs nearby and people gather beforehand to witness the show. R.A. Hutchinson, a park geologist, collected measurements of the eruption durations (X, in minutes) and the subsequent intervals before the next eruption (Y, in minutes) over an 8–day period.

Usage

```
ex0723
```

Format

A data frame with 107 observations on the following 3 variables.

`Date` date of observation (August 1 to August 8, 1978)

`Interval` length of interval before the next eruption (in minutes)

`Duration` duration of eruption (in minutes)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Weisberg, S. (1985). *Applied Linear Regression*, John Wiley \& Sons, New York, p. 231.

Examples

```
str(ex0723)
```

---

ex0724 *Crab Claw Size and Force*

Description

As part of a study of the effects of predatory intertidal crab species on snail populations, researchers measured the mean closing forces and the propdus heights of the claws on several crabs of three species.

Usage

```
ex0724
```

Format

A data frame with 38 observations on the following 3 variables.

`Force` closing strength of claw of the crab

`Height` propodus height of claw of the crab

`Species` species to which the crab belongs

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Yamada, S.B. and Boulding, E.G. (1992). Shell–breaking Efficiency of Predatory Crabs Influences the Distribution of an Intertidal Snail, Technical Report, Zoology Department, Oregon State University.

Examples

```
str(ex0724)
```

---

ex0726 *Decline in Male Births*

Description

The data are on the proportion of male birts in Denmark, The Netherlands, Canada and the United States for a number of yeras. Notice that the proportions for Canada and the United States are only provided for the years 1970 to 1990, while Denmark and The Netherlands have data listed for 1950 to 1994.

Usage

```
ex0726
```

Format

A data frame with 45 observations on the following 5 variables.

`Year` year of observation

`Denmark` male birth rate of Denmark for given year

`Netherlands` male birth rate of The Netherlands for given year

`Canada` male birth rate of Canada for given year

`Usa` male birth rate of the United States for given year

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Davis, D.L., Gottlieb, M.B. and Stampnitzky, J.R. (1998). Reduced ratio of male to female births in several industrial countries, *Journal of the American Medical Association* **279**(13): 1018–1023.

Examples

```
str(ex0726)
```

ex0727 *The Big Bang II*

Description

These data are measured distances and recession velocities for 10 clusters of nebulae, much farther from earth than the nebulae reported in `case0701`.

Usage

```
ex0727
```

Format

A data frame with 10 observations on the following 2 variables.

`Cluster` name of the cluster of nebulae

`Distance` distance from earth (in million parsec)

`Velocity` recession velocity (in kilometres per second)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Hubble, E. and Humason, M. (1931). The Velocity–Distance Relation Among Extra–calactic Nebulae, *Astrophysics Journal* **74**: 43–50.

See Also

`case0701`

Examples

```
str(ex0727)
```

ex0728 *Number of Stories and Building Height*

Description

The *1994 World Almanac* reports heights and number of stories for notable tall buildings in North America. The data in this data frame are a random sample of size 60 of those for which dates of completion were available.

Usage

```
ex0728
```

Format

A data frame with 60 observations on the following 3 variables.

`Year` year of completion

`Height` height of building

`Stories` number of stories of building

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
str(ex0728)
```

---

ex0729 *Male Displays*

Description

Black wheatears are small birds in Spain and Morocco. Males of the species demonstrate an exaggerated sexual display by carrying many heavy stones to nesting cavities. This 35–gram bird transports, on average, 3.1 kg of stones per nesting season! Different males carry somewhat different sized stones, prompting a study on whether larger stones may be a signal of higher health status. Soler et al. calculated the average stone mass (g) carried by each of 21 male black wheatears, along with T-cell response measurements reflecting their immune systems' strengths.

Usage

```
ex0729
```

Format

A data frame with 21 observations on the following 2 variables.

`Mass` average mass of stones carried by bird (in g)

`Tcell` T-cell response measurement (in mm)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Soler, M., Martín-Vivaldi, M., Marín J. and Møller, A. (1999). Weight lifting and health status in the black wheatears, *Behavioral Ecology* **10**(3): 281–286.

Examples

```
str(ex0729)
```

ex0730 *Brain Activity in Violin and String Players*

Description

Studies over the past two decades have shown that activity can effect the reorganisation of the human central nervous system. For example, it is known that the part of the brain associated with activity of a finger or limb is taken over for other purposes in individuals whose limb or finger has been lost. In one study, psychologists used magnetic source imaging (MSI) to measure neuronal activity in the brains of nine string players (six violinists, two cellists and one guitarist) and six controls who had never played a musical instrument, when the thumb and fifth finger of the left hand were exposed to mild stimulation. The researchers felt that stringed instrument players, who use the fingers of their left hand extensively, might show different behaviour—as a result of this extensive physical activity—than individuals who did not play stringed instruments.

Usage

```
ex0730
```

Format

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

`Years` years that the individual has been playing

`Activity` neuronal activity index

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B. and Taub E. (1995). Increased cortical representation of the fingers of the left hand in string players, *Science* **270**(5234): 305–307.

Examples

```
str(ex0730)
```

ex0816 *Meat Processing*

Description

The data in `case0702` are a subset of the complete data on postmortum pH in 12 steer carcasses.

Usage

```
ex0816
```

Format

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

`Time` time after slaughter (hours)

`pH` pH level in postmortem muscle

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Schwenke, J.R. and Milliken, G.A. (1991). On the Calibration Problem Extended to Nonlinear Models, *Biometrics* **47**(2): 563–574.

See Also

`case0702`

Examples

```
str(ex0816)
```

---

ex0817 *Biological Pest Control*

Description

In a study of the effectiveness of biological control of the exotic weed tansy ragwort, researchers manipulated the exposure to the ragwort flea beetle on 15 plots that had been planted with a high density of ragwort. Harvesting the plots the next season, they measured the average dry mass of ragwort remaining (grams/plant) and the flea beetle load (beetles/gram of ragwort dry mass) to see if the ragwort plants in plots with high flea beetle loads were smaller as a result of herbivory by the beetles.

Usage

```
ex0817
```

Format

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

`Load` flee beetle load (in beetles/gram of ragwort dry mass)

`Mass` dry mass of ragwort weed

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

McEvoy, P. and Cox, C. (1991). Successful Biological Control of Ragwort, *Senecio jacobaea*, by introducing insects in Oregon, *Ecological Applications* **1**(4): 430–442.

Examples

```
str(ex0817)
```

`ex0818` *Chernobyl Fallout*

Description

One of the most dangerous contaminants deposited over European countries following the Chernobyl accident in April 1986 was radioactive cesium. To study cesium transfer from contaminated soil to plants, researchers collected soil samples and samples of mushroom mycelia from 17 wooded locations in Umbria, Central Italy, from August 1986 to November 1989. The data are measured concentrations (Bq/kg) of cesium in the soil and in the mushrooms.

Usage

```
ex0818
```

Format

A data frame with 17 observations on the following 2 variables.

`Mushroom` Cesium concentrations in mushrooms (in Bq/kg)

`Soil` Cesium concentrations in soil (in Bq/kg)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Borio, R., Chiocchini, S., Cicioni, R., Degli Esposti, P., Rongoni, A., Sabatini, P., Scampoli, P., Antonini, A. and Salvadori, P. (1991). Uptake of Radiocesium by Mushrooms, *Science of the Total Environment* **106**(3): 183–190.

Examples

```
str(ex0818)
```

ex0820 *Election Fraud*

Description

The data are observations on the difference between Democratic and Republican vote counts, by (a) absentee ballot and (b) voting machine, for 21 elections in Philadelphia's senatorial districts over the last 10 years.

Usage

```
ex0820
```

Format

A data frame with 21 observations on the following 2 variables.

`Absentee` Democratic minus Republican vote count by absentee ballot

`Machines` Democratic minus Republican vote count by voting machine

Details

In a special election to fill a Pennsylvania State Senate seat in 1993, the Democrat, William Stinson, received 19,127 machine–counted votes and the Republican, Bruce Marks, received 19,691. In addition, there were 1,391 absentee ballots for Stinson and 366 absentee ballots for Marks, so that the total tally showed Stinson the winner by 461 votes. The large disparity between the machine–counted and absentee votes, and the resulting reversal of the outcome due to the absentee ballots caused some concern about possible illegal influence on the absentee votes. To see whether the discrepancy in absentee votes was larger than could be explained by chance, an econometrician considered the data given in this data frame (read from a graph in *The New York Times*, 11 April 1994).

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

See Also

ex1115

Examples

```
str(ex0820)
```

ex0822 *Ecosystem Decay*

Description

Data are the number of butterfly species in 16 islands of forest of various sizes in otherwise cleared areas in Brazil.

Usage

```
ex0822
```

Format

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

`Area` area (ha) of forest patch

`Species` number of butterfly species

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Lovejoy, T.E., Rankin, J.M., Bierregaard, Jr., R.O., Brown, Jr., K.S., Emmons, L.H. and van der Voort, M. (1984). Ecosystem decay of Amazon forest remnants *in* Nitecki, M.H. (ed.) *Extinctions*, University of Chicago Press.

Examples

```
str(ex0822)
```

ex0823 *Wine Consumption and Heart Disease*

Description

The data are the average wine consumption rates (in liters per person per year) and number of ischemic heart disease deaths (per 1000 men aged 55 to 64 years) for 18 industrialized countries.

Usage

```
data(ex0823)
```

Format

A data frame with 18 observations on the following 3 variables.

`Country` a character vector indicating the country

`Wine` consumption of wine (liters per person per year)

`Mortality` heart disease mortality rate (deaths per 1,000)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

St. Leger A.S., Cochrane, A.L. and Moore, F. (1979). Factors Associated with Cardiac Mortality in Developed Countries with Particular Reference to the Consumption of Wine, *Lancet*: 1017–1020.

Examples

```
str(ex0823)
```

---

ex0824 *Respiratory Rates for Children*

Description

A high respiratory rate is a potential diagnostic indicator of respiratory infection in children. To judge whether a respiratory rate is "high" however, a physician must have a clear picture of the distribution of normal rates. To this end, Italian researchers measured the respiratory rates of 618 children between the ages of 15 days and 3 years.

Usage

```
ex0824
```

Format

A data frame with 618 observations on the following 2 variables.

`Age` age in months of child

`Rate` respiratory rate (breaths per minute)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Rusconi, F., Castagneto, M., Porta, N., Gagliardi, L., Leo, G., Pellegatta, A., Razon, S. and Braga, M. (1994). Reference Values for Respiratory Rate in the First 3 Years of Life, *Pediatrics* **94**(3): 350–355.

Examples

```
str(ex0824)
```

`ex0825` *The Dramatic U.S. Presidential Election of 2000*

Description

Data set shows the number of votes for Buchanan and Bush in all 67 counties in Florida during the U.S. presidential election of November 7, 2000.

Usage

```
ex0825
```

Format

A data frame with 67 observations on the following 3 variables.

`County` a character vector indicating the county

`Buchanan2000` votes cast for P. Buchanan

`Bush2000` votes cast for G.W. Bush

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

See Also

`ex1222`

Examples

```
str(ex0825)
```

`ex0914` *Pace of Life and Heart Disease*

Description

In four regions of the US (Northeast, Midwest, South and West), in three different sized metropolitan regions, researchers measured indicators of pace of life.

Usage

```
ex0914
```

Format

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

`Bank` bank clerk speed

`Walk` pedestrian walking speed

`Talk` postal clerk talking speed

`Heart` age adjusted death rate due to heart disease

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Levine, R.V. (1990). The Pace of Life, *American Scientist* **78**: 450–459.

Examples

```
str(ex0914)
```

---

ex0915 *Rainfall and Corn Yield*

Description

Data on corn yield and rainfall in six U.S. corn–producing states (Iowa, Nebraska, Illinois, Indiana, Missouri and Ohio), recorded for each year from 1890 to 1927.

Usage

```
ex0915
```

Format

`Year` year of observation (1890–1927)

`Yield` average corn yield for the six states (in bu/acre)

`Rainfall` average rainfall in the six states (in in/year)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Ezekiel, M. and Fox, K.A. (1959). *Methods of Correlation and Regression Analysis*, John Wiley \& Sons, New York.

Examples

```
str(ex0915)
```

ex0918 *Speed of Evolution*

Description

Researchers studied the development of a fly (Drosophila subobscura) that had been accidentally introduced from the Old World into North America around 1980.

Usage

```
ex0918
```

Format

A data frame with 21 observations on the following 8 variables.

`Continent` a factor with levels `"NA"` and `"EU"`

`Latitude` latitude (degrees)

`Females` average wing size ($10^3 \times$log mm) of female flies on log scale

`SE.F` standard error of wing size ($10^3 \times$log mm) of female flies on log scale

`Males` average wing size ($10^3 \times$log mm) of male flies on log scale

`SE.M` standard error of wing size ($10^3 \times$log mm) of male flies on log scale

`Ratio` average basal length to wing size ratios of female flies

`SE.R` standard error of average basal length to wing size ratio of female flies

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Huey, R.B., Gilchrist, G.W., Carlson, M.L., Berrigan, D. and Serra, L. (2000). Rapid Evolution of a Geographic Cline in Size in an Introduced Fly, *Science* **287**(5451): 308–309.

Examples

```
str(ex0918)
```

ex0920 *Winning Speeds at the Kentucky Derby*

Description

Data set contains the year of the Kentucky Derby, the winning horse, the condition of the track and the average speed of the winner for years 1896–2000.

Usage

```
ex0920
```

Format

A data frame with 105 observations on the following 4 variables.

`Year` year of Kentucky Derby

`Winner` a character vector with the name of the winning horse

`Condition` a factor with levels `"fast"`, `"good"` and `"slow"`

`Speed` average speed of the winner (feet per second)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

http://www.kentuckyderby.com

Examples

```
str(ex0920)
```

ex1014 *Toxic Effects of Copper and Zinc*

Description

Researchers randomly allocated 25 beakers containing minnow larvae to receive one of 25 treatment combinations of 5 levels of zinc and 5 levels of copper.

Usage

```
ex1014
```

Format

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

`Copper` amount of copper received (in ppm)

`Zinc` amount of zinc received (in ppm)

`Protein` protein in minnow larvae exposed to copper and zinc (μg/larva)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Ryan, D.A., Hubert, J.J., Carter, E.M., Sprague, J.B. and Parrott, J. (1992). A Reduced-Rank Multivariate Regression Approach to Aquatic Joint Toxicity Experiments, *Biometrics* **48**(1): 155–162.

Examples

```
str(ex1014)
```

---

ex1026 *Thinning of Ozone Layer*

Description

Depletion of the ozone layer allows the most damaging ultraviolet radiation to reach the Earth's surface. To measure the relationship, researchers sampled the ocean column at various depths at 17 locations around Antarctica during the austral spring of 1990.

Usage

```
ex1026
```

Format

A data frame with 17 observations on the following 3 variables.

`Inhibit` percent inhibition of primary phytoplankton production in water

`UVB` UVB exposure

`Surface` a factor with levels `"Deep"` and `"Surface"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Smith, R.C., Prézelin, B.B., Baker, K.S., Bidigare, R.R., Boucher, N.P., Coley, T., Karentz, D., MacIntyre, S., Matlick, H.A., Menzies, D., Ondrusek, M., Wan, Z. and Waters, K.J. (1992). Ozone Depletion: Ultraviolet Radiation and Phytoplankton Biology in Antarctic Waters, *Science* **255**(5047): 952–959.

Examples

```
str(ex1026)
```

---

ex1027 *Factors Affecting Extinction*

Description

Data are measurements on breeding pairs of land-bird species collected from 16 islands around Britain over the course of several decades. For each species, the data set contains an average time of extinction on those islands where it appeared, the average number of nesting pairs, the size of the species and the migratory status of the species.

Usage

```
ex1027
```

Format

A data frame with 62 observations on the following 5 variables.

`Species` a character vector indicating the species

`Time` average extinction time in years

`Pairs` average number of nesting pairs

`Size` a factor with levels `"L"` and `"S"`

`Status` a factor with levels `"M"` and `"R"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Pimm, S.L., Jones, H.L., and Diamond, J. (1988). On the Risk of Extinction, *American Naturalist* **132**(6): 757–785.

Examples

```
str(ex1027)
```

ex1028 *El Nino and Hurricanes*

Description

Data set with the numbers of Atlantic Basin tropical storms and hurricanes for each year from 1950–1997. The variable storm index is an index of overall intensity of hurricane season. Also listed are whether the year was a cold, warm or neutral El Nino year and a variable indicating whether West Africa was wet or dry that year.

Usage

```
ex1028
```

Format

A data frame with 48 observations on the following 7 variables.

`Year` year

`ElNino` a factor with levels `"cold"`, `"neutral"` and `"warm"`

`Temperature` numeric variable with values -1 if `ElNino` is `"cold"`, 0 if `"neutral"` and 1 if `"warm"`

`WestAfrica` numeric variable indicating whether West Africa was wet (1) or dry (0)

`Storms` number of storms

`Hurricanes` number of hurricanes

`StormIndex` index of overall intensity of hurricane season

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Data were gathered by William Gray of Colorado State University and reported on USA Today weather page: http://www.usatoday.com/weather/whurnum.htm

Examples

```
str(ex1028)
```

ex1029 *Wage and Race*

Description

Data set contains weekly wages in 1987 for a sample of 25,632 males between the age of 18 and 70 who worked full-time along with their years of education, years of experience, indicator variable for whether they were black, indicator variable for whether they worked in or near a city, and a code for the region in the US where they worked.

Usage

```
ex1029
```

Format

A data frame with 25631 observations on the following 6 variables.

`Wage` weekly wage in dollars

`Education` education in years

`Experience` experience in years

`Black` a factor with levels `"Yes"` and `"No"`; indicator for whether the person is black

`SMSA` a factor with levels `"Yes"` and `"No"`; indicator for whether the person worked in or near a city

`Region` a factor with levels `"MW"`, `"NE"`, `"S"` and `"W"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Bierens, H.J. and Ginther, D.K. (2001). Integrated Conditional Moment Testing of Quantile Regression Models, *Empirical Economics* **26**(1): 307–324

http://econ.la.psu.edu/~hbierens/QUANTILE.PDF

http://econ.la.psu.edu/~hbierens/MEDIAN.HTM

Examples

```
str(ex1029)
```

ex1115 *Election Fraud*

Description

The data are observations on the difference between Democratic and Republican vote counts, by (a) absentee ballot and (b) voting machine, for 22 elections in Philadelphia's senatorial districts over the last 10 years.

Usage

```
ex1115
```

Format

A data frame with 22 observations on the following 2 variables.

`Absentee` Democratic minus Republican vote count by absentee ballot

`Machines` Democratic minus Republican vote count by voting machine

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

See Also

ex0820

Examples

```
str(ex1115)
```

ex1120 *Was Tyrannosaurus Rex Warm-Blooded?*

Description

Data are the isotopic composition of structural bone carbonate (X) and the isotopic composition of the coexisting calcite cements (Y) in 18 bone samples from a specimen of the dinosaur Tyrannosaurus rex. Evidence that the mean of Y is positively associated with X was used in an argument that the metabolic rate of this dinosaur resembled warm-blooded more than cold-blooded animals.

Usage

```
ex1120
```

Format

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

`Carbonat` isotopic composition of bone carbonate

`Calcite` isotopic composition of calcite cements

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Barrick, R.E. and Showers, W.J. (1994). Thermophysiology of *Tyrannosaurus rex*: Evidence from Oxygen Isotopes, *Science* **265**(5169): 222–224.

See Also

`ex0523`

Examples

```
str(ex1120)
```

---

ex1122 *Deforestation and Debt*

Description

It has been theorized that developing countries cut down their forests to pay off foreign debt. Data are debt, deforestation, and population from 11 Latin American nations.

Usage

```
ex1122
```

Format

A data frame with 11 observations on the following 4 variables.

`Country` a character vector indicating the country

`Debt` debt (millions of dollars)

`Deforest` deforestation (thousands of ha)

`Pop` population (thousands of people)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Gullison, R.R. and Losos, E.C. (1992). The Role of Foreign Debt in Deforestation in Latin America, *Conservation Biology* **7**(1): 140–7.

Examples

```
str(ex1122)
```

`ex1123` *Air Pollution and Mortality*

Description

Does pollution kill people? Data in one early study designed to explore this issue from 5 Standard Metropolitan Statistical Areas in the U.S between 1959–1961.

Usage

```
ex1123
```

Format

A data frame with 60 observations on the following 7 variables.

`City` a character vector indicating the city

`Mort` total age-adjusted mortality from all causes

`Precip` mean annual precipitation (inches)

`Educ` median number of school years completed for persons 25 years or older

`Nonwhite` percentage of population that is nonwhite

`NOx` relative pollution potential of oxides of nitrogen

`SO2` relative pollution potential of sulfur dioxide

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

McDonald, G.C. and Ayers, J.A. (1978). Some Applications of the "Chernoff Faces": A Technique for Graphically Representing Multivariate Data *in* Wang, P.C.C. (ed.) *Graphical Representation of Multivariate Data*, Academic Press.

See Also

`ex1217`

Examples

```
str(ex1123)
```

ex1124 *Natal Dispersal Distances of Mammals*

Description

An assessment of the factors affecting dispersal distances is important for understanding population spread, recolonization and gene flow which are central issues for conservation of many vertebrate species. Researchers gathered data on body weight, diet type and maximum natal dispersal distance for various animals.

Usage

```
ex1124
```

Format

A data frame with 64 observations on the following 4 variables.

`Species` a character vector indicating the species

`Bodymass` bodymass (kg)

`Maxdist` maximum dispersal distance (km)

`Type` a factor with levels `"Carnivore"`, `"Herbivore"` and `"Omnivore"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Sutherland, G.D., Harestad, A.S., Price, K. and Lertzman, K.P. (2000). Scaling of Natal Dispersal Distances in Terrestrial Birds and Mammals, *Conservation Ecology* **4**(1): 16.

Examples

```
str(ex1124)
```

ex1217 *Pollution and Mortality*

Description

Complete data set for problem introduced in ex1123. Data from early study designed to explore the relationship between air pollution and mortality.

Usage

```
ex1217
```

Format

A data frame with 60 observations on the following 17 variables.

`City` a character vector indicating the city

`Mort` total age-adjusted mortality from all causes

`Precip` mean annual precipitation (inches)

`Humidity` percent relative humidity (annual average at 1:00pm)

`Jantemp` mean January temperature (degrees F)

`Julytemp` mean July temperature (degrees F)

`Over65` percentage of the population aged 65 years or over

`House` population per household

`Educ` median number of school years completed for persons 25 years or older

`Sound` percentage of the housing that is sound with all facilities

`Density` population density (in persons per square mile of urbanized area)

`Nonwhite` percentage of population that is nonwhite

`Whitecol` percentage of employment in white collar occupations

`Poor` percentage of households with annual income under $3,000 in 1960

`HC` relative pollution potential of hydrocarbons

`NOx` relative pollution potential of oxides of nitrogen

`SO2` relative pollution potential of sulfur dioxide

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

McDonald, G.C. and Ayers, J.A. (1978). Some Applications of the "Chernoff Faces": A Technique for Graphically Representing Multivariate Data *in* Wang, P.C.C. (ed.) *Graphical Representation of Multivariate Data*, Academic Press.

See Also

`ex1123`

Examples

```
str(ex1217)
```

`ex1220` *Galapagos Islands*

Description

The number of species on an island is known to be related to the island's area. Of interest is what other variables are also related to the number of species, after island area is accounted for, and whether the answer differs for native and non native species.

Usage

```
ex1220
```

Format

A data frame with 30 observations on the following 8 variables.

`Island` a character vector indicating the island

`Total` total number of observed species

`Native` number of native species

`Area` area (km^2)

`Elev` elevation (m)

`DistNear` distance from nearest island (km)

`DistSC` distance from Santa Cruz (km)

`AreaNear` area of nearest island (km^2)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Johnson, M.P. and Raven, P.H. (1973). Species Number and Endemism: The Galapagos Archipelago Revisited, *Science* **179**(4076): 893–895.

Examples

```
str(ex1220)
```

ex1221 *River Nitrogen*

Description

The rise in abundance of algae in coastal waters is thought to be due to increases in nutrients such as nitrate and other forms of nitrogen. Researchers gathered data to gauge the evidence that nitrates in the discharges of rivers around the world are associated with human population density.

Usage

```
ex1221
```

Format

A data frame with 42 observations on the following 11 variables.

`River` a character vector indicating the river

`Country` a factor variable with 26 levels

`Discharge` the estimated annual average discharge of the river into an ocean (m^3 per second)

`Runoff` estimated annual average runoff from the watershed (liters/(sec$\times$ km^2))

`Area` watershed area (km^2)

`Density` density of people (people/km^2)

`NO3` nitrate concentration (μM/l)

`Export` nitrate export (product of runoff times nitrate concentration)

`Dep` deposition (proportional to product of nitratrate precipitation times precipitation)

`NPrec` nitrate precipitation (μmol NO_3/(sec$\times km^2$))

`Prec` precipitation (cm/year)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Cole, J.L., Peierls, B.L., Caraco, N.F. and Pace, M.L. (1993). Nitrogen Loading of Rivers as a Human-driven Process, *in* McDonnell, M.J. and Pickett, S.T.A. (eds.) *Humans as Components of Ecosystems: The Ecology of Subtle Human Effects and Populated Areas*, Springer-Verlag.

Examples

```
str(ex1221)
```

ex1222 *Bush Gore Ballot Controversy*

Description

This data set contains the vote counts by county in Florida for Buchanan and for four other presidential candidates in 2000, along with the total vote counts in 2000, the presidential vote counts for three presidential candidates in 1996, the vote count for Buchanan in his only other campaign in Florida—the 1996 Republican primary, the registration in Buchanan's Reform party and the total political party registration in the county.

Usage

```
ex1222
```

Format

A data frame with 67 observations on the following 13 variables.

`County` a character vector indicating the county

`Buchanan2000` votes cast for Buchanan in 2000 presidential election

`Gore2000` votes cast for Gore in 2000 presidential election

`Bush2000` votes cast for Bush in 2000 presidential election

`Nader2000` votes cast for Nader in 2000 presidential election

`Browne2000` votes cast for Browne in 2000 presidential election

`Total2000` total vostes cast in 2000 presidential election

`Clinton96` votes cast for Clinton in 1996 presidential election

`Dole96` votes cast for Dole in 1996 presidential election

`Perot96` votes cast for Perot in 1996 presidential election

`Buchanan96p` votes cast for Buchanan in 1996 Republican primary

`ReformReg` the registration in Buchanan's Reform party

`TotalReg` the total political party registration

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

See Also

`ex0825`

Examples

```
str(ex1222)
```

ex1317 *Dinosaur Extinctions—An Observational Study*

Description

About 65 million years ago, the dinosaurs suffered a mass extinction virtually overnight (in geologic time). Among many clues, one that all scientists regard as crucial is a layer of iridium-rich dust that was deposited over much of the earth at that time. The theory is that an event like a volcanic eruption or meteor impact caused a massive dust cloud that blanketed the earth for years killing off animals and their food sources. Dataset has Iridium depths by type of deposit.

Usage

```
ex1317
```

Format

A data frame with 28 observations on the following 3 variables.

`Iridium` Iridium in samples (ppt)

`Strata` a factor with levels `"Limestone"` and `"Shale"`

`Depth` a factor with six levels: `"1"`, `"2"`, ..., `"6"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Alvarez, W. and Asaro, F. (1990). What Caused the Mass Extinction? An Extraterrestrial Impact, *Scientific American* **263**(4): 76–84.

Courtillot, E. (1990). What Caused the Mass Extinction? A Volcanic Eruption. *Scientific American* **263**(4): 85–92.

Examples

```
str(ex1317)
```

ex1319 *Nature—Nurture*

Description

A 1989 study investigated the effect of heredity and environment on intelligence. Data are the IQ scores for adopted children whose biological and adoptive parents were categorized either in the highest or the lowest socioeconomic status category.

Usage

```
ex1319
```

Format

A data frame with 38 observations on the following 3 variables.

`IQ` IQ scores of adopted children

`Adoptive` a factor with levels `"High"` and `"Low"`; the socioeconomic status of the adoptive parents

`Biologic` a factor with levels `"High"` and `"Low"`; the socioeconomic status of the biological parents

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Capron, C. and Duyme, M. (1991). Children's IQ's and SES of Biological and Adoptive Parents in a Balanced Cross-fostering Study, *European Bulletin of Cognitive Psychology* **11**(3): 323–348.

See Also

ex1605

Examples

```
str(ex1319)
```

---

ex1320 — *Gender Differences in Performance on Mathematics Achievement Tests*

Description

Data set on 861 ACT Assessment Mathematics Usage Test scores from 1987. The test was given to a sample of high school seniors who met one of three profiles of high school mathematics course work: (a) Algebra I only; (b) two Algebra courses and Geometry; and (c) two Algebra courses, Geometry, Trigonometry, Advanced Mathematics and Beginning Calculus.

These data were generated from summary statistics for one particular form of the test as reported by Doolittle (1989).

Usage

```
ex1320
```

Format

A data frame with 861 observations on the following 3 variables.

`Sex` a factor with levels `"female"` and `"male"`

`Background` a factor with levels `"a"`, `"b"` and `"c"`

`Score` ACT mathematics test score

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Doolittle, A.E. (1989). Gender Differences in Performance on Mathematics Achievement Items, *Applied Measurement in Education* **2**(2): 161–177.

http://www.act.org/research/reports/pdf/ACT_RR87-16.pdf

Examples

```
str(ex1320)
```

---

ex1414 *Blood Brain Barrier*

Description

Researchers designed an experiment to investigate how delivery of brain cancer antibody is influenced by tumor size, antibody molecular weight, blood-brain barrier disruption, and delivery route.

Usage

```
ex1414
```

Format

A data frame with 36 observations on the following 6 variables.

`Agent` a factor with levels `"AIB"`, `"DEX7"` and `"MTX"`

`Treatment` a factor with levels `"BD"` and `"NS"`

`Route` a factor with levels `"IA"` and `"IV"`

`Days` days after inoculation

`BAT` concentration of antibody in the part of the brain around the tumor

`LH` concentration of antibody in the unaffected part of the brain

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Barnett, P.A., Roman-Goldstain, S., Ramsey, F., McCormick, C.I., Sexton, G., Szumowski, J. and Neuwelt, E.A. (1995). Differential Permeability and Quantitative MR Imaging of a Human Lung Carcinoma Brain Xenograft in the Nude Rat, *American Journal of Pathology* **146**(2): 436–449.

See Also

ex1415

Examples

```
str(ex1414)
```

---

ex1415 *Second Replicate of the Barrier Disruption Study*

Description

Researchers designed an experiment to investigate how delivery of brain cancer antibody is influenced by tumor size, antibody molecular weight, blood-brain barrier disruption, and delivery route. The data for the first replicate of this study is in ex1414. This is the second replicate for the study.

Usage

```
ex1415
```

Format

A data frame with 36 observations on the following 6 variables.

`Agent` a factor with levels `"AIB"`, `"DEX7"` and `"MTX"`

`Treatment` a factor with levels `"BD"` and `"NS"`

`Route` a factor with levels `"IA"` and `"IV"`

`Days` days after inoculation

`BAT` concentration of antibody in the part of the brain around the tumor

`LH` concentration of antibody in the unaffected part of the brain

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Barnett, P.A., Roman-Goldstain, S., Ramsey, F., McCormick, C.I., Sexton, G., Szumowski, J. and Neuwelt, E.A. (1995). Differential Permeability and Quantitative MR Imaging of a Human Lung Carcinoma Brain Xenograft in the Nude Rat, *American Journal of Pathology* **146**(2): 436–449.

See Also

ex1414

Examples

```
str(ex1415)
```

`ex1417` *Tennessee Corn Yield Trials*

Description

Corn yield trials were performed at four locations in Tennessee in 1999. Data shows the average yields for six hybrids at each of four locations.

Usage

```
ex1417
```

Format

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

`Location` a factor with five levels: `"Ames.irr"`, `"Ames.un"`, `"Crossvill"`, `"Knoxville"` and `"Milan"`

`Hybrid` a factor with six levels: `"AsgrowRX799"`, `"Beck5912W"`, `"Cargill7821"`, `"FFR739W"`, `"NorthrupKing"` and `"Pioneer"`

`Yield` average yield (bushels per acre)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

University of Tennessee Agricultural Experiment Station.

Examples

```
str(ex1417)
```

ex1509 *Sunspot Counts for 1749–1948*

Description

Time series data set of annual counts of sunspots.

Usage

```
ex1509
```

Format

A data frame with 200 observations on the following 2 variables.

`Year` year

`Spots` number of sunspots

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Waldmeier, M. (1961). *The Sunspot Activity in the Years 1610–1960*, Federal Observatory, Zurich.

Examples

```
str(ex1509)
```

ex1512 *Melanoma and Sunspot Activity—An Observational Study*

Description

Several factors suggest that the incidence of melanoma is related to solar radiation. Data has the age-adjusted melanoma incidence among males from Connecticut Tumor Registry, 1936–1972.

Usage

```
ex1512
```

Format

A data frame with 37 observations on the following 3 variables.

`Year` year

`Melanoma` male melanoma incidence in number of cases per 100,000 population

`Sunspot` sunspot relative number

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Houghton, A., Munster, E.W. and Viola, M.V. (1978). Increased Incidence of Malignant Melanoma After Peaks of Sunspot Activity, *Lancet*: 759–760.

Examples

```
str(ex1512)
```

`ex1513` *Lynx Trappings and Sunspots*

Description

Data on the annual numbers of lynx trapped in the Mackenzie River district of northwest Canada from 1821–1934.

Usage

```
ex1513
```

Format

A data frame with 114 observations on the following 3 variables.

`Year` year

`Lynx` number of lynx trapped

`Spots` number of sunspots

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Elston, C. and Nicholson, M. (1942). The Ten Year Cycle in Numbers of the Lynx in Canada, *Journal of Animal Ecology* **11**(2): 215–244.

Examples

```
str(ex1513)
```

ex1514 *Trends in Firearm and Motor Vehicle Deaths in the U.S.*

Description

Data shows the number of deaths due to firearms and the number of deaths due to motor vehicle accidents in the United States between 1968 and 1993.

Usage

```
ex1514
```

Format

A data frame with 26 observations on the following 3 variables.

`Year` year

`FirearmDeaths` deaths due to firearms (in thousands per year)

`MotorVehicleDeaths` deaths due to motor vehicles (in thousands per year)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Data read from a Centers for Disease Control and Prevention graph reported in The Oregonian, June 17, 1997.

Examples

```
str(ex1514)
```

ex1515 *S&P 500*

Description

Data on the value of a $1 U.S. stock investment in 1871 at the end of each year, based on the Standard and Poor (S&P) 500 Composite stock index.

Usage

```
ex1515
```

Format

A data frame with 129 observations on the following 2 variables.

`Year` year

`SPReturn` S&P composite stock index ($)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
str(ex1515)
```

---

ex1605 *Nature—Nurture*

Description

Data are a subset from an observational, longitudinal, study on adopted children. Is child's intelligence related to intelligence of the biological mother and the intelligence of the adoptive mother?

Usage

```
ex1605
```

Format

A data frame with 62 observations on the following 6 variables.

`AMED` adoptive mother's years of education

`BMIQ` biological mother's score on IQ test

`Age2IQ` IQ of child at age 2

`Age4IQ` IQ of child at age 4

`Age8IQ` IQ of child at age 8

`Age13IQ` IQ of child at age 13

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Skodak, M. and Skeels, H.M. (1949). A Final Follow-up Study of One Hundred Adopted Children, *Journal of Genetic Psychology* **75**: 85–125.

See Also

`ex1319`

Examples

```
str(ex1605)
```

ex1611 *Religious Competition*

Description

Adam Smith, in Wealth of Nations, observed that even religious monopolies become weak when they are not challenged by competition. Data to illustrate this point is from 21 countries in which the percentages of Catholics in the populations varied from a low 1.2% to a high 97.6%.

Usage

```
ex1611
```

Format

A data frame with 21 observations on the following 4 variables.

`Country` a character vector indicating the country

`PctCath` percent Catholics in the population

`P2PRatio` priest to parishioner ratio

`PctIndig` percent clergy indigenous

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Gill, A.J. (1994). Rendering unto Caesar? Religious Competition and Catholic Political Strategy in Latin America, 1962–79, *American Journal of Political Science* **38**(2): 403–425.

Examples

```
str(ex1611)
```

ex1612 *Wastewater*

Description

Samples of effluent were divided and sent to two laboratories for testing. Data are measurements of biochemical oxygen demand and suspended solid measurements obtained for 2 sample splits from the two laboratories.

Usage

```
ex1612
```

Format

A data frame with 11 observations on the following 4 variables.

`ComBOD` biochemical oxygen demand measurements from commercial laboratory

`ComSS` suspended solids measurements from commercial laboratory

`StaBOD` biochemical oxygen demand measurements from state laboratory

`StaSS` suspended solids measurements from state laboratory

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Johnson, R.A. and Wichern, D.W. (1988). *Applied Multivariate Statistical Analysis*, Prentice-Hall.

Examples

```
str(ex1612)
```

---

ex1613 *Flea Beetle Distinction*

Description

Data are the measurements from two very similar species of flea beetle.

Usage

```
ex1613
```

Format

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

`Jnt1` measurement of first joint in micrometers

`Jnt2` measurement of second joint in micrometers

`Species` a factor with levels `"conc"` and `"heik"`

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Lubischew, A.A. (1962). On the Use of Discriminant Functions in Taxonomy, *Biometrics* **18**: 455–477.

Examples

```
str(ex1613)
```

ex1614 *Pschoimmunology*

Description

Recent studies in the field of psychoimmunology suggest a link exists between behavioral events and the functioning of one's immune system. Data shows the results of a study on 12 subjects who were monitored during three distinct activities. The first activity consisted of neutral activity such as reporting tasks. During the second activity, subjects listened to audiotape exercises relating to images of heaviness, warmth in the body, relaxation, suggestions to remember happy events, etc. The third activity included a nonaudio tape follow up stimulus consisting of continued relaxation as in activity 2 and a verbal discussion of the positive aspects of the audiotape.

Usage

```
ex1614
```

Format

A data frame with 12 observations on the following 3 variables.

`PhaseA` Interleukin-1 levels (counts per minute) from blood samples taken during activity A

`PhaseB` Interleukin-1 levels (counts per minute) from blood samples taken during activity B

`PhaseC` Interleukin-1 levels (counts per minute) from blood samples taken during activity C

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Keppel, W. (1993). Effects of Behavioral Stimuli on Plasma Interleukin-1 Activity in Humans at Rest, *Journal of Clinical Psychology* **49**(6): 777–785.

Examples

```
str(ex1614)
```

ex1615 *Trends in SAT Scores*

Description

Data shows a partial listing of a data set with ratios of average math to average verbal SAT scores in the United States and the District of Columbia for 1989 and 1996–1999.

Usage

```
ex1615
```

Format

A data frame with 51 observations on the following 6 variables.

`State` a character vector indicating the state

`M/V:89` average MATH SAT scores divided by average VERBAL SAT score in 1989

`M/V:96` average MATH SAT scores divided by average VERBAL SAT score in 1996

`M/V:97` average MATH SAT scores divided by average VERBAL SAT score in 1997

`M/V:98` average MATH SAT scores divided by average VERBAL SAT score in 1998

`M/V:99` average MATH SAT scores divided by average VERBAL SAT score in 1999

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
str(ex1615)
```

---

ex1708 *Pig Fat*

Description

Actual pig fat and measurements of pig fat from magnetic resonance images at 13 locations for 12 pigs.

Usage

```
ex1708
```

Format

A data frame with 12 observations on the following 14 variables.

`Fat` actual pig fat (in percent)

`M1` magnetic resonance image at location 1

`M2` magnetic resonance image at location 2

`M3` magnetic resonance image at location 3

`M4` magnetic resonance image at location 4

`M5` magnetic resonance image at location 5

`M6` magnetic resonance image at location 6

`M7` magnetic resonance image at location 7

`M8` magnetic resonance image at location 8

`M9` magnetic resonance image at location 9

`M10` magnetic resonance image at location 10

`M11` magnetic resonance image at location 11

`M12` magnetic resonance image at location 12

`M13` magnetic resonance image at location 13

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Glasbey, C.A and Flowler, P.A. (1992). Regression Models Fitted Using Conditional Independence to Estimate Pig Fatness from Magnetic Resonance Images, *The Statistician* **41**(2): 179–184.

Examples

```
str(ex1708)
```

---

ex1713 *Church Distinctiveness*

Description

Data show measures that differ among denominations of American Protestant and Catholic churches.

Usage

```
ex1713
```

Format

A data frame with 18 observations on the following 6 variables.

`Denomination` a character vector indicating the church denomination

`Distinct` distinctiveness (strictness of discipline on a seven point scale)

`Attend` average percentage of weeks that individuals attended a church meeting (% weekly)

`NonChurch` average number of secular organisations to which members belong

`StrongPct` average percentage of members that describe themselves as being strong church members (%)

`AnnInc` average income of members (US$)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Iannaccone, L.R. (1994). Why Strict Churches Are Strong, *American Journal of Sociology* **99**(5): 1180–1211.

Examples

```
str(ex1713)
```

ex1714 *Insurance*

Description

In the 1970's the U.S. Commission on Civil Rights investigated charges that insurance companies were attempting to redefine Chicago "neighborhoods" in order to cancel existing homeowner insurance policies or refuse to issue new ones. Dataset has data on homeowner and residential fire insurance policy issuances from 47 zip codes in the Chicago area.

Usage

```
ex1714
```

Format

A data frame with 47 observations on the following 8 variables.

`Zip` last 2 digits of zip code

`Fire` fires per 1000 housing units

`Theft` thefts per 1000 population

`Age` percentage of housing units built prior to 1940

`Income` median family income

`Race` percentage minority

`Vol` number of new policies per 100 housing units

`Invol` number of FAIR plan policies and renewals per 100 housing units

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Andrews, D.F. and Herzberg, A.M. (1985). *Data: A Collection of Problems from many Fields for the Student and Research Worker*, Springer-Verlag.

Examples

```
str(ex1714)
```

ex1914 *Mantel-Haenszel Test for Censored survival Times: Lymphoma and Radiation Data*

Description

Survival times for two groups of lymphoma patients.

Usage

```
ex1914
```

Format

A data frame with 34 observations on the following 4 variables.

`Months` months after diagnosis

`Group` a factor with levels `"no"` and `"radiation"`

`Survived` number of patients known to survive beyond this month

`Died` number of patients known to die after this many months

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Neuwelt, E.A., Goldman, D.L., Dahlborg, S.A., Crossen, J., Ramsey, F., Roman-Goldstein, S., Braziel, R. and Dana, B. (1991). Primary CNS Lymphoma Treated with Osmotic Blood-brain Barrier Disruption: Prolonged Survival and Preservation of Cognitive Function, *Journal of Clinical Oncology* **9**(9): 1580–1590.

Examples

```
str(ex1914)
```

ex1916 *Vitamin C and Colds*

Description

Fictitious data set based on results of an experiment where subjects were randomly divided into two groups and given a placebo or vitamin c to take during the cold season. At the end of the cold season, the subjects were interviewed by a physician who determined whether they had or had not suffered a cold during the period. Skeptics interviewed the 800 subjects to determine who knew and who did not know to which group they had been assigned. Vitamin C has a bitter taste and those familiar with it could recognize whether their pills contained it.

Usage

```
ex1916
```

Format

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

`Knew` a factor with levels `"no"` and `"yes"`

`Treatment` a factor with levels `"placebo"` and `"vitC"`

`Cold` number of people who got a cold

`NoCold` number of people who did not get a cold

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
str(ex1916)
```

---

ex1917 *Alcohol Consumption and Breast Cancer—A Retrospective Study*

Description

Dataset from a study which investigated the added risk of breast cancer due to alcohol consumption. A sample of confirmed breast cancer patients were compared with a sample of cancer free women who were close in age and from the same neighborhood as the cases. Data was collected on the alcohol consumption and body mass of both sets of women.

Usage

```
ex1917
```

Format

A data frame with 6 observations on the following 4 variables.

`Bodymass` a factor with levels `"high"`, `"low"` and `"medium"`

`Drinking` a factor with levels `"high"` and `"low"`

`Cases` number of women with breast cancer

`Controls` number of women without breast cancer

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Rosenberg, L., Palmer, J.R., Miller, D.R., Clarke, E.A. and Shapiro, S. (1990). A Case-Control Study of Alcoholic Beverage Consumption and Breast Cancer, *American Journal of Epidemiology* **131**(1): 6–14.

Examples

```
str(ex1917)
```

`ex1918` *The Donner Party*

Description

In 1846 the Donner party became stranded while crossing the Sierra Nevada Mountains near Lake Tahoe. The data frame has the counts for male and female survivors for six age groups.

Usage

```
ex1918
```

Format

A data frame with 12 observations on the following 4 variables.

`Age` a factor with six levels: `"15-19"`, `"20-29"`, `"30-39"`, `"40-49"`, `"50-59"` and `"60-69"`

`Sex` a factor with levels `"female"` and `"male"`

`Lived` number that lived

`Died` number that died

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Grayson, D.K. (1990). Donner Party Deaths: A Demographic Assessment, *Journal of Anthropological Research* **46**: 223–242.

See Also

`case2001`

Examples

```
str(ex1918)
```

ex1919 *Tire-Related Fatal Accidents and Ford Sports Utility Vehicles*

Description

Data shows the numbers of compact sports utility vehicles involved in fatal accidents in the U.S. between 1995 and 1999, categorized according to travel speed, make of car (Ford or other), and cause of accident (tire-related or other).

Usage

```
ex1919
```

Format

A data frame with 8 observations on the following 4 variables.

`Speed` a factor with levels `"0-40"`, `"41-55"`, `"56-65"` and `">65"`

`Make` a factor with levels `"Ford"` and `"Other"`

`Other` cause of accident was other than tire-related

`Tire` cause of accident was tire-related

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

See Also

ex2018

Examples

```
str(ex1919)
```

ex2011 *Space Shuttle*

Description

This data frame contains the launch temperatures (degrees Fahrenheit) and an indicator of O-ring failures for 24 space shuttle launches prior to the space shuttle *Challenger* disaster of January 28, 1986.

Usage

```
ex2011
```

Format

A data frame with 24 observations on the following 2 variables.

`Temp` Launch temperature (in degrees Fahrenheit)

`Failure` Indicator of O-ring failure

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

See Also

`case0401`, `ex2223`

Examples

```
str(ex2011)
```

---

ex2012 *Muscular Dystrophy*

Description

Duchenne Muscular Dystrophy (DMD) is a genetically transmitted disease, passed from a mother to her children. Boys with the disease usually die at a young age; but affected girls usually do not suffer symptoms, may unknowingly carry the disease and may pass it to their offspring. It is believed that about 1 in 3,300 women are DMD carriers. A woman might suspect she is a carrier when a related male child develops the disease. Doctors must rely on some kind of test to detect the presence of the disease. This data frame contains data on two enzymes in the blood, creatine kinase (CK) and hemopexin (H) for 38 known DMD carriers and 82 women who are not carriers. It is desired to use these data to obtain an equation for indicating whether a women is a likely carrier.

Usage

```
ex2012
```

Format

A data frame with 120 observations on the following 3 variables.

`Group` Indicator whether the woman has DMD (`"Case"`) or not (`"Control"`)

`CK` Creatine kinase reading

`H` Hemopexin reading

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Andrews, D.F. and Herzberg, A.M. (1985). *Data: A Collection of Problems From Many Fields For The Student And Research Worker*, Springer-Verlag, New York.

Examples

```
str(ex2012)
```

`ex2015` *Spotted Owl Habitat*

Description

A study examined the association between nesting locations of the Northern Spotted Owl and availability of mature forests. Wildlife biologists identified 30 nest sites. The researchers selected 30 other sites at random coordinates in the same forest. On the basis of aerial photographs, the percentage of mature forest (older than 80 years) was measured in various rings around each of the 60 sites.

Usage

```
ex2015
```

Format

A data frame with 60 observations on the following 8 variables.

`Site` Site, a factor with levels `"Random"` and `"Nest"`

`PctRing1` Percentage of mature forest in ring with outer radius 0.91 km

`PctRing2` Percentage of mature forest in ring with outer radius 1.18 km

`PctRing3` Percentage of mature forest in ring with outer radius 1.40 km

`PctRing4` Percentage of mature forest in ring with outer radius 1.60 km

`PctRing5` Percentage of mature forest in ring with outer radius 1.77 km

`PctRing6` Percentage of mature forest in ring with outer radius 2.41 km

`PctRing7` Percentage of mature forest in ring with outer radius 3.38 km

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Ripple W.J., Johnson, D.H., Thershey, K.T. and Meslow E.C. (1991). Old–growth and Mature Forests Near Spotted Owl Nests in Western Oregon, *Journal of Wildlife Management* **55**(2): 316–318.

Examples

```
str(ex2015)
```

`ex2016` *Bumpus Natural Selection Data*

Description

Hermon Bumpus analysed various characteristics of some house sparrows that were found on the ground after a severe winter storm in 1898. Some of the sparrows survived and some perished. This data set contains the survival status, age, the length from tip of beak to tip of tail (in mm), the alar extent (length from tip to tip of the extended wings, in mm), the weight in grams, the length of the head in mm, the length of the humerus (arm bone, in inches), the length of the femur (thigh bones, in inches), the length of the tibio–tarsus (leg bone, in inches), the breadth of the skull in inches and the length of the sternum in inches.

Usage

```
ex2016
```

Format

A data frame with 87 observations on the following 11 variables.

`Status` Survival status, factor with levels `"Perished"` and `"Survived"`

`AG` Age, factor with levels `"adult"` and `"juvenile"`

`TL` total length (in mm)

`AE` alar extent (in mm)

`WT` weight (in grams)

`BH` length of beak and head (in mm)

`HL` length of humerus (in inches)

`FL` length of femur (in inches)

`TT` length of tibio–tarsus (in inches)

`SK` width of skull (in inches)

`KL` length of keel of sternum (in inches)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

See Also

`case0201`, `ex0221`

Examples

```
str(ex2016)
```

ex2017 *Catholic stance*

Description

The Catholic church has explicitly opposed authoritarian rule in some (but not all) Latin American countries. Although such action could be explained as a desire to counter repression or to increase the quality of life of its parishioners, A.J. Gill supplies evidence that the underlying reason may be competition from evangelical Protestant denominations. He compiled the data given in this data frame.

Usage

```
ex2017
```

Format

A data frame with 12 observations on the following 5 variables.

`Stance` Catholic church stance, factor with levels `"Pro"` and `"Anti"`

`Country` Latin American country

`PQLI` Physical Quality of Life Index in the mid-1970s; Average of live expectancy at age 1, infant mortality and literacy at age 15+.

`Repress` Average civil rights score for the period of authoritarian rule until 1979

`Compete` Percentage increase of competitive religious groups during the period 1900–1970

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Gill, A.J. (1994). Rendering unto Caesar? Religious Competition and Catholic Strategy in Latin America, 1962–1979, *American Journal of Political Science* **38**(2): 403–425.

Examples

```
str(ex2017)
```

`ex2018` *Fatal Car Accidents Involving Tire Failures on Ford Explorers*

Description

This data frame contains data on 1995 and later model compact sports utility vehicles involved in fatal accidents in the United States between 1995 and 1999, excluding those that were struck by another car and excluding accidents that, according to police reports, involved alcohol.

Usage

```
ex2018
```

Format

A data frame with 2321 observations on the following 4 variables.

`Make` Type of sports utility vehicle, factor with levels `"Other"` and `"Ford"`

`Vehicle.age` Vehicle age (in years); surrogate for age of tires

`Passengers` Number of passengers

`Cause` Cause of fatal accident, factor with levels `"Not_Tire"` and `"Tire"`

Details

The Ford Explorer is a popular sports utility vehicle made in the United States and sold throughout the world. Early in its production concern arose over a potential accident risk associated with tires of the prescribed size when the vehicle was carrying heavy loads, but the risk was thought to be acceptable if a low tire pressure was recommended. The problem was apparently exacerbated by a particular type of Firestone tire that was overly prone to separation, especially in warm temperatures. This type of tire was a common one used on Explorers in model years 1995 and later. By the end of 1999 more than 30 lawsuits had been filed over accidents that were thought to be associated with this problem. U.S. federal data on fatal car accidents were analysed at that time, showing that the odds of a fatal accident being associated with tire failure were three times as great for Explorers as for other sports utility vehicles.

Additional data from 1999 and additional variables may be used to further explore the odds ratio. It is of interest to see whether the odds that a fatal accident is tire-related depend on whether the vehicle is a Ford, after accounting for age of the car and number of passengers. Since the Ford tire problem may be due to the load carried, there is some interest in seeing whether the odds associated with a Ford depend on the number of passengers.

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

See Also

`ex1919`

Examples

```
str(ex2018)
```

ex2115 *Belief Accessibility*

Description

The study the effect of *context questions* prior to *target questions*, researchers conducted a poll involving 1,054 subjects selected randomly from the Chicago phone directory. To include possibly unlisted phones, selected numbers were randomly altered in the last position. This data frame contains the responses to one of the questions asked concerning continuing U.S. aid to the Nicaraguan Contra rebels. Eight different versions of the interview were given, representing all possible combinations of three factors at each of two levels. The experimental factors were `Context`, `Mode` and `Level`.

`Context` refers to the type of context questions preceding the question about Nicaraguan aid. Some subjects received a context question about Vietnam, designed to elicit reticence about having the U.S. become involved in another foreign war in a third–world country. The other context question was about Cuba, designed to elicit anti–communist sentiments.

`Mode` refers to whether the target question immediately followed the context question or whether there were other questions scattered in between.

`Level` refers to two versions of the context question. In the `"high"` level the question was worded to elicit a higher level of agreement than in the `"low"` level wording.

Usage

```
ex2115
```

Format

A data frame with 8 observations on the following 5 variables.

`Context` Factor referring to the context of the question preceding the target question about U.S. aid to the Nicaraguan Contra rebels

`Mode` Factor with levels `"not"` and `"scattered"`, `"scattered"` is used if the target question was not asked directly after the context question

`Level` Factor with levels `"low"` and `"high"`, refers to the wording of the question

`M` Number of people interviewed

`Percent` Percentage in favour of Contra aid

Details

Increasingly, politicians look to public opinion surveys to shape their public stances. Does this represent the ultimate in democracy? Or are seemingly scientific polls being rigged by the manner of questioning? Psychologists believe that opinions—expressed as answers to questions—are usually generated at the time the question is asked. Answers are based on a quick sampling of relevant beliefs held by the subject, rather than a systematic canvas of all such beliefs. Furthermore, this sampling of beliefs tends to overrepresent whatever beliefs happen to be most accessible at the time the question is asked. This aspect of delivering opinions can be abused by the pollster. Here, for example, is one sequence of questions:

(1) "Do you believe the Bill of Rights protects personal freedom?"

(2) "Are you in favor of a ban on handguns?"

Here is another:

(1) "Do you think something should be done to reduce violent crime?"

(2) "Are you in favor of a ban on handguns?"

The proportion of yes answers to question 2 may be quite different depending on which question 1 is asked first.

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Tourangeau, R., Rasinski, K.A., Bradburn, N. and D'Andrade, R. (1989). Belief Accessibility and Context Effects in Attitude Measurement, *Journal of Experimental Social Psychology* **25**: 401–421.

Examples

```
str(ex2115)
```

`ex2116` *Aflatoxicol and Liver Tumors in Trout*

Description

An experiment at the Marine/Freshwater Biomedical Sciences Center at Oregon State University investigated the carcinogenic effects of aflatoxicol, a metabolite of Aflatoxin B1, which is a toxic by-product produced by a mold that infects cottonseed meal, peanuts and grains. Twenty tanks of rainbow trout embryos were exposed to one of five doses of Aflatoxicol for one hour. The data represent the numbers of fish in each tank and the numbers of these that had liver tumours after one year.

Usage

```
ex2116
```

Format

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

`Dose` Dose (in ppm)

`Tumor` Number of trout with liver tumours

`Total` Number of trout in tank

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
str(ex2116)
```

`ex2117` *Effect of Stress During Conception on Odds of a Male Birth*

Description

The probability of a male birth in humans is about .51. It has previously been noticed that lower proportions of male births are observed when offspring is conceived at times of exposure to smog, floods or earthquakes. Danish researchers hypothesised that sources of stress associated with severe life events may also have some bearing on the sex ratio. To investigate this theory they obtained the sexes of all 3,072 children who were born in Denmark between 1 January 1980 and 31 December 1992 to women who experienced the following kind of severe life events in the year of the birth or the year prior to the birth: death or admission to hospital for cancer or heart attack of their partner or of their other children. They also obtained sexes on a sample of 20,337 births to mothers who did not experience these life stress episodes. This data frame contains the data that were collected. Noticed that for one group the exposure is listed as taking place during the first trimester of pregnancy. The rationale for this is that the stress associated with the cancer or heart attack of a family member may well have started before the recorded time of death or hospital admission.

Usage

```
ex2117
```

Format

A data frame with 5 observations on the following 4 variables.

`Group` Indicator for groups to which mothers belong

`Time` Indicator for time at which severe life event occurred

`Number` Number of births

`PctBoys` Percentage of boys born

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Hansen, D., Møller, H. and Olsen, J. (1999). Severe Periconceptional Life Events and the Sex Ratio in Offspring: Follow Up Study based on Five National Registers, *British Medical Journal* **319**(7209): 548–549.

Examples

```
str(ex2117)
```

ex2118 *HIV and Circumcision*

Description

Researchers in Kenya identified a cohort of more that 1,000 prostitutes who were known to be a major reservoir of sexually transmitted diseases in 1985. It was determined that more than 85% of them were infected with human immunodeficiency virus (HIV) in February, 1986. The researchers identified men who acquired a sexually-transmitted disease from this group of women after the men sought treatment at a free clinic. The data frame contains data on the subset of those men who did not test positive for HIV on their first visit and who agreed to participate in the study. The men are categorised according to whether they later tested positive for HIV during the study period, whether they had one or multiple sexual contacts with the prostitutes and whether they were circumcised.

Usage

```
ex2118
```

Format

A data frame with 4 observations on the following 5 variables.

`Contact` Whether men had single or multiple contact with prostitutes

`Circumcised` Whether the men are circumcised, factor with levels `"no"` and `"yes"`

`HIV` Number of men that tested positive for HIV

`Number` Number of men

`NoHIV` Number of men that did not test positive for HIV (should be `Number-HIV`)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Cameron, D.W., D'Costa, L.J., Maitha, G.M., Cheang, M., Piot, P., Simonsen, J.N., Ronald, A.R., Gakinya, M.N., Ndinya-Achola, J.O., Brunham, R.C. and Plummer, F. A. (1989). Female to Male Transmission of Human Immunodeficiency Virus Type I: Risk Factors for Seroconversion in Men, *The Lancet* **334**(8660): 403–407.

Examples

```
str(ex2118)
```

ex2119 *Meta–Analysis of Breast Cancer and Lactation Studies*

Description

This data frame gives the results of 10 separate case–control studies on the association of breast cancer and whether a woman had breast–fed children.

Usage

```
ex2119
```

Format

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

`Study` Factor indicating the study from which data was taken

`Lactate` Whether women had breast–fed children (lactated)

`Cancer` Number of women with breast cancer

`NoCancer` Number of women without breast cancer

Details

Meta–analysis refers to the analysis of analyses. When the main results of studies can be cast into 2×2 tables of counts, it is natural to combine individual odds ratios with a logistic regression model that includes a factor to account for different odds from the different studies. In addition, the odds ratio itself might differ slightly among studies because of different effects on different populations or different research techniques. One approach for dealing with this is to suppose an underlying common odds ratio and to model between–study variability as extra–binomial variation.

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Data gathered from various sources by Karolyn Kolassa as part of a Master's project, Oregon State University.

Examples

```
str(ex2119)
```

ex22.20 *Cancer Death of Atomic Bomb Survivors*

Description

The data in this data frame are the number of cancer deaths among survivors of the atomic bombs dropped on Japan during World War II, categorised by time (years) after the bomb that death occurred and the amount of radiation exposure that the survivors received from the blast (Data from D.A. Pierce, personal communication.) Also listed in each cell is the *person-years at risk*, in 100s. This is the sum total of all years spent by all persons in the category.

Usage

```
ex22.20
```

Format

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

`Exposure` Estimated exposure to radiation (in rads)

`Years` Years after exposure, factor with 7 levels

`Deaths` Number of cancer deaths

`Risk` Person-years at risk (in 100s)

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
str(ex22.20)
```

ex2216 *Murder–Suicides by Deliberate Plane Crash*

Description

Some sociologist suspect that highly publicised suicides may trigger additional suicides. In one investigation of this hypothesis, D.P. Phillips collected information about 17 airplane crashes that were known (because of notes left behind) to be murder–suicides. For each of these crashes, Phillips reported an index of the news coverage (circulation of nine newspapers devoting space to the crash multiplied by length of coverage) and the number of multiple-fatality plane crashes during the week following the publicised crash. This data frame contains the collected data.

Usage

```
ex2216
```

Format

A data frame with 17 observations on the following 2 variables.

`Index` Index for the amount of newspaper coverage given the murder–suicide

`Crashes` Multiple-fatality crashes in the week following a murder–suicide by plane crash

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Phillips, D.P. (1978). Airplane Accident Fatalities Increase Just After Newspaper Stories About Murder and Suicide, *Science* **201**: 748–750.

Examples

```
str(ex2216)
```

---

ex2222 *Emulating Jane Austen's Writing Style*

Description

When she died in 1817, the English novelist Jane Austen had not yet finished the novel *Sanditon*, but she did leave notes on how she intended to conclude the book. The novel was completed by a ghost writer, who attempted to emulate Austen's style. In 1978, a researcher reported counts of some words found in chapters of books written by Austen and in chapters written by the emulator. These data are given in this data frame.

Usage

```
ex2222
```

Format

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

`Count` Number of occurrences of a word in various chapters of books written by Jane Austen and the ghost writer

`Book` Title of books used

`Word` Words used

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Morton, A.Q. (1978). *Literary Detection: How to Prove Authorship and Fraud in Literature and Documents*, Charles Scribner's Sons, New York.

Examples

```
str(ex2222)
```

`ex2223` *Space Shuttle O-Ring Failures*

Description

On January 27, 1986, the night before the space shuttle *Challenger* exploded, an engineer recommended to the National Aeronautics and Space Administration (NASA) that the shuttle not be launched in the cold weather. The forecasted temperature for the *Challenger* launch was 31 degrees Fahrenheit—the coldest launch ever. After an intense 3-hour telephone conference, officials decided to proceed with the launch. This data frame contains the launch temperatures and the number of O-ring problems in 24 shuttle launches prior to the *Challenger*.

Usage

```
ex2223
```

Format

A data frame with 24 observations on the following 2 variables.

`Temp` Launch temperatures (in degrees Fahrenheit)

`Incident` Numbers of O-ring incidents

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

See Also

`case0401`, `ex2011`

Examples

```
str(ex2223)
```

`ex2224` *Valve Failure in Nuclear Reactors*

Description

This data frame contains data on characteristics and numbers of *failures* observed in valve types from one pressurised water reactor.

Usage

```
ex2224
```

Format

A data frame with 90 observations on the following 7 variables.

`System` System, factor with 5 levels

`Operator` Operator type, factor with 4 levels

`Valve` Valve type, factor with 6 levels

`Size` Head size, factor with 3 levels (less than 2 inches, 2–10 inches and 10–30 inches)

`Mode` Operation mode, factor with 2 levels

`Failures` Number of failures observed

`Time` Lengths of observation time

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Moore, L.M. and Beckman, R.J. (1988). Appropriate One-Sided Tolerance Bounds on the Number of Failures using Poisson Regression, *Technometrics* **30**: 283–290.

Examples

```
str(ex2224)
```

ex2225 *Body Size and Reproductive Success in a Population of Male Bullfrogs*

Description

As an example of field observation in evidence of theories of sexual selection, S.J. Arnold and M.J. Wade presented the following data set on size and number of mates observed in 38 bullfrogs.

Usage

```
ex2225
```

Format

A data frame with 38 observations on the following 2 variables.

`Bodysize` Body size (in mm)

`Mates` Number of mates

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Arnold, S.J. and Wade, M.J. (1984). On the Measurement of Natural and Sexual Selection: Aplications, *Evolution* **38**: 720–734.

Examples

```
str(ex2225)
```

ex2414 *Amphibian Crisis and UV-B*

Description

Data frame contains the percentage of unsuccessful hatching from enclosures containing 150 eggs each in a study to investigate whether UV-B is responsible for low hatch rates.

Usage

```
ex2414
```

Format

A data frame with 71 observations on the following 4 variables.

`Percent` percentage of frog eggs failing to hatch

`Treat` factor variable with levels `"NoFilter"`, `"UV-BTransmitting"` and `"UV-BBlocking"`

`Location` factor variable with levels `"ThreeCreeks"`, `"SparksLake"`, `"SmallLake"` and `"LostLake"`

`Phtolyas` Photolyase activity

Source

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

References

Blaustein, A.R., Hoffman, P.D., Hokit, D.G., Kiesecker, J.M., Walls, S.C. and Hays, J.B. (1994). UV Repair and Resistance to Solar UV-B in Amphibian Eggs: A Link to Population Declines? *Proceedings of the National Academy of Science, USA* **91**: 1791–1795.

Examples

```
str(ex2414)
```

`Sleuth2Manual` *Manual of the R Sleuth2 package*

Description

If the option "pdfviewer" is set, this command will display the PDF version of the help pages.

Usage

```
Sleuth2Manual()
```

Author(s)

Berwin A Turlach <berwin@maths.uwa.edu.au>

References

Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Examples

```
## Not run: Sleuth2Manual()
```

Index