

Package ‘Sleuth3’

August 4, 2012

Title Data sets from Ramsey and Schafer’s “Statistical Sleuth (3rd ed)”

Version 0.0-5

Date 2012-08-03

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. (2013), “The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)”, Cengage Learning.

Maintainer Berwin A Turlach <Berwin.Turlach@gmail.com>

LazyData yes

Depends R (>= 2.12.0)

Suggests lattice, multcomp, car, leaps

License GPL (>= 2)

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

R topics documented:

Sleuth3-package	2
case0101	2
case0102	3
case0201	4
case0202	5
case0301	6
case0302	8
case0401	9
case0402	10
case0501	11
case0502	13
case0601	14
case0602	16
case0701	17
case0702	18
case0801	20
case0802	21

case0901	22
case0902	23
case1001	25
case1002	26
case1101	28
case1102	30
case1201	32
case1202	33
case1301	36
case1302	37
case1401	39
case1402	40
Sleuth3Manual	42
Index	43

Sleuth3-package	<i>The R Sleuth3 package</i>
-----------------	------------------------------

Description

Data sets from Ramsey and Schafer’s "Statistical Sleuth (3rd ed)"

Details

This package contains a variety of datasets. For a complete list, use `library(help="Sleuth3")` or `Sleuth3Manual()`.

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.Turlach@gmail.com>

case0101	<i>Motivation and Creativity</i>
----------	----------------------------------

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, with levels "Extrinsic" and "Intrinsic"

Source

Ramsey, F.L. and Schafer, D.W. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

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

```
attach(case0101)
str(case0101)
boxplot(Score ~ Treatment) # Basic boxplots for each level of Treatment

boxplot(Score ~ Treatment, # Boxplots with labels
        ylab= "Average Creativity Score From 11 Judges (on a 40-point scale)",
        names=c("23 'Extrinsic' Group Students", "24 'Intrinsic' Group Students"),
        main= "Haiku Creativity Scores for 47 Creative Writing Students")

detach(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, with levels "Female" and "Male"

Source

Ramsey, F.L. and Schafer, D.W. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

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

```
attach(case0102)
str(case0102)

boxplot(Salary ~ Sex,
  ylab= "Starting Salary (U.S. Dollars)",
  names=c("61 Females", "32 Males"),
  main= "Harris Bank Entry Level Clerical Workers, 1969-1971")

hist(Salary[Sex=="Female"])
dev.new()
hist(Salary[Sex=="Male"])

t.test(Salary ~ Sex, var.equal=TRUE) # Equal var. version; 2-sided by default
t.test(Salary ~ Sex, var.equal=TRUE,
  alternative = "less") # 1-sided; that group 1 (females) mean is less

detach(case0102)
```

case0201

Peter and Rosemary Grant's Finch Beak Data

Description

In the 1980s, biologists Peter and Rosemary Grant caught and measured all the birds from more than 20 generations of finches on the Galapagos island of Daphne Major. In one of those years, 1977, a severe drought caused vegetation to wither, and the only remaining food source was a large, tough seed, which the finches ordinarily ignored. Were the birds with larger and stronger beaks for opening these tough seeds more likely to survive that year, and did they tend to pass this characteristic to their offspring? The data are beak depths (height of the beak at its base) of 89 finches caught the year before the drought (1976) and 89 finches captured the year after the drought (1978).

Usage

case0201

Format

A data frame with 178 observations on the following 2 variables.

Year Year the finch was caught, 1976 or 1978

Depth Beak depth of the finch (mm)

Source

Ramsey, F.L. and Schafer, D.W. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

References

Grant, P. (1986). *Ecology and Evolution of Darwin's Finches*, Princeton University Press, Princeton, N.J.

See Also

[ex2018](#)

Examples

```
attach(case0201)
str(case0201)

mean(Depth[Year==1978]) - mean(Depth[Year==1976])

yearFactor <- factor(Year) # Convert the numerical variable Year into a factor
# with 2 levels. 1976 is "group 1" (it comes first alphanumerically)
t.test(Depth ~ yearFactor, var.equal=TRUE) # 2-sample t-test; 2-sided by default
t.test(Depth ~ yearFactor, var.equal=TRUE,
       alternative = "less") # 1-sided; alternative: group 1 mean is less

boxplot(Depth ~ Year,
        ylab= "Beak Depth (mm)",
        names=c("89 Finches in 1976", "89 Finches in 1978"),
        main= "Beak Depths of Darwin Finches in 1976 and 1978")

## BOXPLOTS FOR PRESENTATION
boxplot(Depth ~ Year,
        ylab="Beak Depth (mm)", names=c("89 Finches in 1976", "89 Finches in 1978"),
        main="Beak Depths of Darwin Finches in 1976 and 1978", col="green",
        boxlwd=2, medlwd=2, whisklty=1, whisklwd=2, staplewex=.2, staplelwd=2,
        outlwd=2, outpch=21, outbg="green", outcex=1.5)

detach(case0201)
```

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 cm³) of several regions and subregions of the twins' brains.

Usage

case0202

Format

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

Unaffected volume of left hippocampus of unaffected twin (in cm³)

Affected volume of left hippocampus of affected twin (in cm³)

Source

Ramsey, F.L. and Schafer, D.W. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

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

```
attach(case0202)
str(case0202)

diff <- Unaffected-Affected
summary(diff)
t.test(diff) # Paired t-test is a one-sample t-test on differences
t.test(Unaffected,Affected,pair=TRUE) # Alternative coding for the same test

boxplot(diff,
  ylab="Difference in Hippocampus Volume (cubic cm)",
  xlab="15 Sets of Twins, One Affected with Schizophrenia",
  main="Hippocampus Difference: Unaffected Twin Minus Affected Twin")
abline(h=0,lty=2) # Draw a dashed (lty=2) horizontal line at 0

## BOXPLOT FOR PRESENTATION:
boxplot(diff,
  ylab="Difference in Hippocampus Volume (cubic cm)",
  xlab="15 Sets of Twins, One Affected with Schizophrenia",
  main="Hippocampus Difference: Unaffected Minus Affected Twin",
  col="green", boxlwd=2, medlwd=2, whisklty=1, whisklwd=2,
  staplewex=.2, staplelwd=2, outlwd=2, outpch=21, outbg="green",
  outcex=1.5)
abline(h=0,lty=2)

detach(case0202)
```

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 suitable 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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

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

```
attach(case0301)
str(case0301) #Seeded is level 1 of Treatment (it's first alphabetically)

boxplot(Rainfall ~ Treatment)
boxplot(log(Rainfall) ~ Treatment) # Boxplots of natural logs of Rainfall

t.test(log(Rainfall) ~ Treatment, var.equal=TRUE,
       alternative="greater") # 1-sided t-test; alternative: level 1 mean is greater

myTest <- t.test(log(Rainfall) ~ Treatment, var.equal=TRUE,
               alternative="two.sided") # 2-sided alternative to get confidence interval
exp(myTest$est[1] - myTest$est[2]) # Back-transform estimate on log scale
exp(myTest$conf) # Back transform endpoints of confidence interval

boxplot(log(Rainfall) ~ Treatment,
       ylab="Log of Rainfall Volume in Target Area (Acre Feet)",
       names=c("On 26 Seeded Days", "On 26 Unseeded Days"),
       main="Distributions of Rainfalls from Cloud Seeding Experiment")

## POLISHED BOXPLOTS FOR PRESENTATION:
opar <- par(no.readonly=TRUE) # Store device graphics parameters
par(mar=c(4,4,4,4)) # Change margins to allow more space on right
boxplot(log(Rainfall) ~ Treatment, ylab="Log Rainfall (Acre-Feet)",
       names=c("on 26 seeded days","on 26 unseeded days"),
       main="Boxplots of Rainfall on Log Scale", col="green", boxlwd=2,
       medlwd=2, whisklty=1, whisklwd=2, staplewex=.2, staplelwd=2,
       outlwd=2, outpch=21, outbg="green", outcex=1.5 )
myTicks <- c(1,5,10,100,500,1000,2000,3000) # some tick marks for original scale
axis(4, at=log(myTicks), label=myTicks) # Add original-scale axis on right
mtext("Rainfall (Acre Feet)", side=4, line=2.5) # Add right-side axis label
par(opar) # Restore previous graphics parameter settings
```

```
detach(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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

References

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

Examples

```
attach(case0302)
str(case0302)    # Note: Level 1 of Veteran is "Other" (first alphabetically)

boxplot(Dioxin ~ Veteran)

t.test(Dioxin ~ Veteran, var.equal=TRUE,
       alternative="less") # 1-sided t-test; alternative: group 1 mean is less
t.test(Dioxin ~ Veteran, alternative="less", var.equal=TRUE,
       subset=(Dioxin < 40)) # t-test on subset for which Dioxin < 40
t.test(Dioxin ~ Veteran, alternative="less", var.equal=TRUE,
       subset=(Dioxin < 20))
t.test(Dioxin ~ Veteran, var.equal=TRUE) # 2-sided--to get confidence interval

## HISTOGRAMS FOR PRESENTATION
opar <- par(no.readonly=TRUE) # Store device graphics parameter settings
par(mfrow=c(2,1), mar=c(3,3,1,1)) # 2 by 1 layout of plots; change margins
myBreaks <- (0:46) - .5 # Make breaks for histogram bins
```



```

hist(Dioxin[Veteran=="Other"], breaks=myBreaks, xlim=range(Dioxin),
     col="green", xlab="", ylab="", main="")
text(10,25,
     "Dioxin in 97 'Other' Veterans; Estimated mean = 4.19 ppt (95% CI: 3.72 to 4.65 ppt)",
     pos=4, cex=.75) # CI from 1-sample t-test & subset=(Veteran="Other")
hist(Dioxin[Veteran=="Vietnam"],breaks=myBreaks,xlim=range(Dioxin),
     col="green", xlab="", ylab="", main="")
text(10,160,
     "Dioxin in 646 Vietnam Veterans; Estimated mean = 4.26 ppt (95% CI: 4.06 to 4.64 ppt)",
     pos=4, cex=.75)
text(13,145,"[Estimated Difference in Means: 0.07 ppt (95% CI: -0.63 to 0.48 ppt)]",
     pos=4, cex=.75)
par(opar) # Restore previous graphics parameter settings

detach(case0302)

```

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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

References

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

See Also

[ex2011](#), [ex2223](#)

Examples

```
str(case0401)
attach(case0401)

mCool <- mean(Incidents[Launch=="Cool"])
mWarm <- mean(Incidents[Launch=="Warm"])
mDiff <- mCool - mWarm
c(mCool,mWarm,mDiff) # Show the values of these variables

## PERMUTATION TEST , VIA REPEATED RANDOM RE-GROUPING (ADVANCED)
numRep <- 50 # Number of random groupings. CHANGE TO LARGER NUMBER; eg 50,000.
rDiff <- rep(0,numRep) # Initialize this variable to contain numRep 0s.
for (rep in 1:numRep) { # Repeat the following commands numRep times:
  randomGroup <- rep("rWarm",24) # Set randomGroup to have 24 values "rWarm"
  randomGroup[sample(1:24,4)] <- "rCool" # Replace 4 at random with "rCool"
  mW <- mean(Incidents[randomGroup=="rWarm"]) # average of random "rWarm" group
  mC <- mean(Incidents[randomGroup=="rCool"]) # average of random "rCool" group
  rDiff[rep] <- mC-mW # Store difference in averages in 'rep' cell of rDiff
} # End of loop
hist(rDiff, # Histogram of difference in averages from numRep random groupings
     main="Approximate Permutation Distribution",
     xlab="Possible Values of Difference in Averages",
     ylab="Frequency of Occurrence")
abline(v=mDiff) # Draw a vertical line at the actually observed difference
pValue <- sum(rDiff >= 1.3)/numRep # 1-sided p-value
pValue
text(mDiff,75000, paste(" -->",round(pValue,4)), adj=-0.1)

detach(case0401)
```

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 allotted 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

Treatment factor variable with two levels—“Modified” and “Conventional”

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

Source

Ramsey, F.L. and Schafer, D.W. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

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) # level 1 of Treatment is "Conventional" (1st alphabetically)
attach(case0402)

boxplot(Time ~ Treatment)
median(Time[Treatment=="Conventional"])-median(Time[Treatment=="Modified"])

wilcox.test(Time ~ Treatment, exact=FALSE, correct=TRUE,
  alternative="greater") # Rank-sum test; alternative: group 1 is greater
wilcox.test(Time ~ Treatment, exact=FALSE, correct=TRUE,
  alternative="two.sided", conf.int=TRUE) # Use 2-sided to get confidence int.

## DOT PLOTS FOR PRESENTATION
xTreatment <- ifelse(Treatment=="Conventional",1,2) # Make numerical values
myPointCode <- ifelse(Censored==0,21,24)
plot(Time ~ jitter(xTreatment,.2), # Jitter the 1's and 2's for visibility
  ylab="Completion Time (Sec.)", xlab="Training Method (jittered)",
  main="Test Completion Times from Cognitive Load Experiment",
  axes=FALSE, pch=myPointCode, bg="green", cex=2, xlim=c(.5,2.5) )
axis(2) # Draw y-axis as usual
axis(1, tick=FALSE, at=c(1,2), # Draw x-axis without ticks
  labels=c("Conventional (n=14 Students)","Modified (n=14 Students)") )
legend(1.5,300, legend=c("Did not Complete in 300 sec","Completed in 300 sec."),
  pch=c(24,21), pt.cex=2, pt.bg="green")

detach(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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

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)
attach(case0501)

# Re-order levels for better boxplot organization:
myDiet <- factor(Diet, levels=c("NP", "N/N85", "N/R50", "R/R50", "lopro", "N/R40") )

myNames <- c("NP(49)", "N/N85(57)", "N/R50(71)", "R/R50(56)", "lopro(56)",
  "N/R40(60)") # Make these for boxplot labeling.
boxplot(Lifetime ~ myDiet, ylab= "Lifetime (months)", names=myNames,
  xlab="Treatment (and sample size)")
myAov1 <- aov(Lifetime ~ Diet) # One-way analysis of variance
plot(myAov1, which=1) # Plot residuals versus estimated means.
summary(myAov1)
pairwise.t.test(Lifetime,Diet, pool.SD=TRUE, p.adj="none") # All t-tests

## p-VALUES AND CONFIDENCE INTERVALS FOR SPECIFIED COMPARISONS OF MEANS
if(require(multcomp)){
  diet <- factor(Diet,labels=c("lopro", "NN85", "NR40", "NR50", "NP", "RR50"))
  myAov2 <- aov(Lifetime ~ diet - 1)
  myComparisons <- glht(myAov2,
    linfct=c("dietNR50 - dietNN85 = 0",
      "dietNR40 - dietNR50 = 0",
      "dietRR50 - dietNR50 = 0",
      "dietlopro - dietNR50 = 0",
      "dietNN85 - dietNP = 0") )
  summary(myComparisons,test=adjusted("none")) # No multiple comparison adjust.
  confint(myComparisons, calpha = univariate_calpha()) # No adjustment
}

## EXAMPLE 5: BOXPLOTS FOR PRESENTATION
boxplot(Lifetime ~ myDiet, ylab= "Lifetime (months)", names=myNames,
  main= "Lifetimes of Mice on 6 Diet Regimens",
  xlab="Diet (and sample size)", col="green", boxlwd=2, medlwd=2, whisklty=1,
```

```
whisklwd=2, staplewex=.2, staplelwd=2, outlwd=2, outpch=21, outbg="green",
outcex=1.5)

detach(case0501)
```

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

Source

Ramsey, F.L. and Schafer, D.W. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

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)
attach(case0502)

# Make new factor level names (with sample sizes) for boxplots
myNames <- c("A (5)", "B (6)", "C (9)", "D (2)", "E (6)", "F (9)", "Spock's (9)")

boxplot(Percent ~ Judge, ylab = "Percent of Women on Judges' Venires",
```

```

names = myNames, xlab = "Judge (and number of venires)",
main = "Percent Women on Venires of 7 Massachusetts Judges")
myAov1 <- aov(Percent ~ Judge)
plot(myAov1, which=1) # Residual plot
summary(myAov1) # Initial screening. Any evidence of judge differences? (yes)

## ANALYSIS 1. TWO-SAMPLE t-TEST (ASSUMING NON-SPOCK JUDGES HAVE A COMMON MEAN)
SpockOrOther <- factor(ifelse(Judge=="Spock's","Spock","Other"))
aovFull <- aov(Percent ~ Judge)
aovReduced <- aov(Percent ~ SpockOrOther)
anova(aovReduced,aovFull) #Any evidence that 7 mean fits better than the 2 mean?
t.test(Percent ~ SpockOrOther, var.equal=TRUE) # Evidence that 2 means differ?

## ANALYSIS 2. COMPARE SPOCK MEAN TO AVERAGE OF OTHER MEANS
myAov3 <- aov(Percent ~ Judge - 1)
myContrast <- rbind(c(1/6, 1/6, 1/6, 1/6, 1/6, 1/6, - 1))
if(require(multcomp)){ # use multcomp library
  myComparison <- glht(myAov3, linfct=myContrast)
  summary(myComparison, test=adjusted("none"))
  confint(myComparison)
}

## BOXPLOTS FOR PRESENTATION
boxplot(Percent ~ Judge, ylab= "Percent of Women on Judges' Venires",
names=myNames, xlab="Judge (and number of venires)",
main= "Percent Women on Venires of 7 Massachusetts Judges",
col="green", boxlwd=2, medlwd=2, whisklty=1, whisklwd=2,
staplewex=.2, staplelwd=2, outlwd=2, outpch=21, outbg="green",
outcex=1.5)

detach(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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

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)
attach(case0601)

## EXPLORATION
myHandicap <- factor(Handicap,
  levels=c("None", "Amputee", "Crutches", "Hearing", "Wheelchair"))
boxplot(Score ~ myHandicap,
  ylab= "Qualification Score Assigned by Student to Interviewee",
  xlab= "Treatment Group--Handicap Portrayed (14 Students in each Group)",
  main= "Handicap Discrimination Experiment on 70 Undergraduate Students")
myAov <- aov(Score ~ myHandicap)
plot(myAov, which=1) # Plot residuals versus estimated means
summary(myAov)

## COMPARE MEAN QUALIFICATION SCORE OF EVERY HANDICAP GROUP TO "NONE"
if(require(multcomp)){ # Use the multcomp library
  myDunnnett <- glht(myAov, linfct = mcp(myHandicap = "Dunnnett"))
  summary(myDunnnett)
  confint(myDunnnett, level=.95)
  opar <- par(no.readonly=TRUE) # Save current graphics parameter settings
  par(mar=c(4.1,8.1,4.1,1.1)) # Change margins
  plot(myDunnnett,
    xlab="Difference in Mean Qualification Score (and Dunnett-adjusted CIs)")
  par(opar) # Restore original graphics parameter settings
}

## COMPARE EVERY MEAN TO EVERY OTHER MEAN
if(require(multcomp)){ # Use the multcomp library
  myTukey <- glht(myAov, linfct = mcp(myHandicap = "Tukey"))
  summary(myTukey)
}

## TEST THE CONTRAST OF DISPLAY 6.4
myAov2 <- aov(Score ~ myHandicap - 1)
myContrast <- rbind(c(0, -1/2, 1/2, -1/2, 1/2))
if(require(multcomp)){ # Use the multcomp library
  myComparison <- glht(myAov2, linfct=myContrast)
  summary(myComparison, test=adjusted("none"))
  confint(myComparison)
}

# BOXPLOTS FOR PRESENTATION
boxplot(Score ~ myHandicap,
  ylab= "Qualification Score Assigned by Student to Video Job Applicant",
```

```
xlab="Handicap Portrayed by Job Applicant in Video (14 Students in each Group)",
main= "Handicap Discrimination Experiment on 70 Undergraduate Students",
col="green", boxlwd=2, medlwd=2, whisklty=1, whisklwd=2, staplewex=.2,
staplelwd=2, outlwd=2, outpch=21, outbg="green", outcex=1.5)

detach(case0601)
```

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—"Pair1", "Pair2", "Pair3", "Pair4", "Pair5" and "Pair6"

Length Body size of the males

Source

Ramsey, F.L. and Schafer, D.W. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

References

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

Examples

```
str(case0602)
attach(case0602)

## EXPLORATION
plot(Proportion ~ Length,
     xlab="Length of the Two Males",
     ylab="Proportion of Time Female Spent with Yellow-Sword Male",
     main="Proportion of Time Spent with Yellow Rather than Transparent Sword Male")
abline(h=.5) # Draw a horizontal line at 0.5 (i.e. the "no preference" line)
```



```

myAov <- aov(Proportion ~ Pair)
plot(myAov, which=1) # Residual plot
summary(myAov)

# Explore possibility of linear effect, as in Display 6.5
myAov2 <- aov(Proportion ~ Pair - 1) # Show the estimated means.
myContrast <- rbind(c(5, -3, 1, 3, -9, 3))
if(require(multcomp)){ # Use the multcomp library
  myComparison <- glht(myAov2, linfct=myContrast)
  summary(myComparison, test=adjusted("none"))
}

# Simpler exploration of linear effect, via regression (Ch. 7)
myLm <- lm(Proportion ~ Length)
summary(myLm)

# ONE-SAMPLE t-TEST THAT MEAN PROPORTION = 0.5, IGNORING MALE PAIR EFFECT
t.test(Proportion, mu=.5, alternative="greater") # Get 1-sided p-value
t.test(Proportion, alternative="two.sided") # Get C.I.

## SCATTERPLOT FOR PRESENTATION
plot(Proportion ~ Length,
     xlab="Length of the Two Males (mm)",
     ylab="Proportion of Time Female Spent with Yellow-Sword Male",
     main="Female Preference for Yellow Rather than Transparent Sword Male",
     pch=21, lwd=2, bg="green", cex=1.5 )
abline(h=.5,lty=2,col="blue",lwd=2)
text(29.5,.52,"0.5 (no preference)", col="blue")

detach(case0602)

```

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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

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)
attach(case0701)

## EXPLORATION
plot(Distance ~ Velocity)
myLm <- lm(Distance ~ Velocity)
abline(myLm)

myResiduals <- myLm$res
myFits <- myLm$fit
plot(myResiduals ~ myFits) # Plot residuals versus estimated means.
abline(h=0) # Draw a horizontal line at 0.
# OR, use this shortcut...
plot(myLm, which=1) # Residual plot (red curve is a scatterplot smoother)

## INFERENCE
summary(myLm)
confint(myLm, level=.95)
myLm2 <- lm(Distance ~ Velocity - 1) # Drop the intercept.
summary(myLm2)
confint(myLm2)

## DISPLAY FOR PRESENTATION
plot(Distance ~ Velocity, xlab="Recession Velocity (km/sec)",
     ylab="Distance from Earth (megaparsecs)",
     main="Measured Distance Versus Velocity for 24 Extra-Galactic Nebulae",
     pch=21, lwd=2, bg="green", cex=1.5 )
abline(myLm, lty=2, col="blue", lwd=2)
abline(myLm2, lty=3, col="red", lwd=2)
legend(-250, 2.05,
      c("unrestricted regression line", "regression through the origin"),
      lty=c(2, 3), lwd=c(2, 2), col=c("blue", "red"))

detach(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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

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)
attach(case0702)

# EXPLORATION
plot(pH ~ Time)
myLm <- lm(pH ~ Time)
abline(myLm, col="blue", lwd=2)
lines(lowess(Time,pH), col="red", lty=2, lwd=2) # Add scatterplot smoother
plot(myLm, which=1) # Residual plot

logTime <- log(Time)
plot(pH ~ logTime)
myLm2 <- lm(pH ~ logTime)
abline(myLm2)
plot(myLm2, which=1)

## PREDICTION BAND ABOUT REGRESSION LINE
xToPredict <- seq(1,8,length=100) # sequence from 1 to 8 of length 100
logXToPredict <- log(xToPredict)
newData <- data.frame(logTime = logXToPredict)
myPredict <- predict(myLm2,newData,
  interval="prediction", level=.90)
plot(pH ~ logTime)
abline(myLm2)
lines(myPredict[,3]~ logXToPredict, lty=2)
lines(myPredict[,2] ~ logXToPredict, lty=2)
# Find smallest time at which the upper endpoint of a 90% prediction
# interval is less than or equal to 6:
minTime <- min(xToPredict[myPredict[,3] <= 6.0])
minTime
```

```

abline(v=log(minTime),col="red")

# DISPLAY FOR PRESENTATION
plot(pH ~ Time, xlab="Time After Slaughter (Hours); log scale",
     ylab="pH in Muscle", main="pH and Time after Slaughter for 10 Steers",
     log="x", pch=21, lwd=2, bg="green", cex=2 )
lines(xToPredict,myPredict[,1], col="blue", lwd=2)
lines(xToPredict, myPredict[,3], lty=2, col="blue", lwd=2)
lines(xToPredict, myPredict[,2], lty=2, col="blue", lwd=2)
legend(3,7, c("Estimated Regression Line", "90% Prediction Band"),
      lty=c(1,2), col="blue", lwd=c(2,2))
abline(h=6, lty=3, col="purple", lwd=2)
text(1.5,6.05,"Desired pH", col="purple")
lines(c(minTime,minTime),c(5,6.15), col="purple", lwd=2)
text(minTime,6.2,"4.9 hours",col="purple",cex=1.25)

detach(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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

References

Wilson, E.O., 1992, **The Diversity of Life**, W. W. norton, N.Y.

Examples

```

str(case0801)
attach(case0801)

## EXPLORATION
logSpecies <- log(Species)
logArea <- log(Area)

```

```

plot(logSpecies ~ logArea, xlab="Log of Island Area",
     ylab="Log of Number of Species",
     main="Number of Reptile and Amphibian Species on 7 Islands")
myLm <- lm(logSpecies ~ logArea)
abline(myLm)

## INFERENCE AND INTERPRETATION
summary(myLm)
slope      <- myLm$coef[2]
slopeConf <- confint(myLm,2)
100*(2^(slope)-1) # Back-transform estimated slope
100*(2^(slopeConf)-1) # Back-transform confidence interval
# Interpretation: Associated with each doubling of island area is a 19% increase
# in the median number of bird species (95% CI: 16% to 21% increase).

## DISPLAY FOR PRESENTATION
plot(Species ~ Area, xlab="Island Area (Square Miles); Log Scale",
     ylab="Number of Species; Log Scale",
     main="Number of Reptile and Amphibian Species on 7 Islands",
     log="xy", pch=21, lwd=2, bg="green",cex=2 )
dummyArea <- c(min(Area),max(Area))
beta <- myLm$coef
meanLogSpecies <- beta[1] + beta[2]*log(dummyArea)
medianSpecies <- exp(meanLogSpecies)
lines(medianSpecies ~ dummyArea,lwd=2,col="blue")
island <- c(" Cuba"," Hispaniola"," Jamaica", " Puerto Rico",
           " Montserrat"," Saba"," Redonda")
for (i in 1:7) {
  offset <- ifelse(Area[i] < 10000, -.2, 1.5)
  text(Area[i],Species[i],island[i],col="dark green",adj=offset,cex=.75) }

detach(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 responses 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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

References

Nelson, W.B., 1970, G.E. Co. Technical Report 71-C-011, Schenectady, N.Y.

Examples

```
str(case0802)
attach(case0802)

## EXPLORATION
plot(Time ~ Voltage)
myLm <- lm(Time ~ Voltage)
plot(myLm, which=1) # Residual plot
logTime <- log(Time)
plot(logTime ~ Voltage)
myLm <- lm(logTime ~ Voltage)
abline(myLm)
plot(myLm, which=1) # Residual plot
myOneWay <- lm(logTime ~ factor(Voltage))
anova(myLm, myOneWay) # Lack of fit test for simple regression (seems okay)

## INFERENCE AND INTERPREATION
beta <- myLm$coef
100*(1 - exp(beta[2])) # Back-transform estimated slope
100*(1 - exp(confint(myLm, "Voltage"))))
# Interpretation: Associated with each 1 kV increase in voltage is a 39.8%
# decrease in median breakdown time (95% CI: 32.5% decrease to 46.3% decrease).

## DISPLAY FOR PRESENTATION
options(scipen=50) # Do this to avoid scientific notation on y-axis
plot(Time ~ Voltage, log="y", xlab="Voltage (kV)",
      ylab="Breakdown Time (min.); Log Scale",
      main="Breakdown Time of Insulating Fluid as a Function of Voltage Applied",
      pch=21, lwd=2, bg="green", cex=1.75 )
dummyVoltage <- c(min(Voltage), max(Voltage))
meanLogTime <- beta[1] + beta[2]*dummyVoltage
medianTime <- exp(meanLogTime)
lines(medianTime ~ dummyVoltage, lwd=2, col="blue")

detach(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 timing of the onset of the light 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; 1=Late, 2=Early

Intensity light intensity (in $\mu\text{mol}/\text{m}^2/\text{sec}$)

Source

Ramsey, F.L. and Schafer, D.W. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

Examples

```
str(case0901)
attach(case0901)

## EXPLORATION
plot(Flowers ~ Intensity, pch=ifelse(Time ==1, 19, 21))
myLm <- lm(Flowers ~ Intensity + factor(Time) + Intensity:factor(Time))
plot(myLm, which=1)
summary(myLm) # Note p-value for interaction term

# INFERENCE
myLm2 <- lm(Flowers ~ Intensity + factor(Time))
summary(myLm2)
confint(myLm2)

# DISPLAY FOR PRESENTATION
plot(Flowers ~ jitter(Intensity,.3),
     xlab=expression("Light Intensity ("*mu*"mol/"*m^2*" /sec)"), # Include symbols
     ylab="Average Number of Flowers per Plant",
     main="Effect of Light Intensity and Timing on Meadowfoam Flowering",
     pch=ifelse(Time ==1, 21, 22), bg=ifelse(Time==1, "orange","green"),
     cex=1.7, lwd=2)
beta <- myLm2$coef
abline(beta[1],beta[2],lwd=2, lty=2)
abline(beta[1]+beta[3],beta[2],lwd=2,lty=3)
legend(700,79,c("Early Start","Late Start"),
     pch=c(22,21),lwd=2,pt.bg=c("green","orange"),pt.cex=1.7,lty=c(3,2))

detach(case0901)
```

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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

See Also

[case0902](#)

Examples

```
str(case0902)
attach(case0902)

## EXPLORATION
myMatrix <- cbind(Brain, Body, Litter, Gestation)
if(require(car)){ # Use the car library
  scatterplotMatrix(myMatrix, # Matrix of scatterplots
    smooth=F, # Omit scatterplot smoother on plots
    diagonal="histogram") # Draw histograms on diagonals
}
myLm <- lm(Brain ~ Body + Litter + Gestation)
plot(myLm, which=1)
logBrain <- log(Brain)
logBody <- log(Body)
logGestation <- log(Gestation)
myMatrix2 <- cbind(logBrain, logBody, Litter, logGestation)
if(require(car)){ # Use the car library
  scatterplotMatrix(myMatrix2, smooth=F, diagonal="histogram")
}
myLm2 <- lm(logBrain ~ logBody + Litter + logGestation)
plot(myLm2, which=1) # Residual plot.

if(require(car)){ # Use the car library
  crPlots(myLm2) # Partial residual plots (Sleuth Ch.11)
}
plot(logBrain ~ logBody)
identify(logBrain ~ logBody, labels=Species) # Identify points on scatterplot
# Place the cursor over a point of interest, then left-click.
# Continue with other points if desired. When finished, pres Esc.

## INFERENCE
```



```

summary(myLm2)
confint(myLm2)

# DISPLAYS FOR PRESENTATION
myLm3 <- lm(logBrain ~ logBody + logGestation)
beta <- myLm3$coef
logBrainAdjusted <- logBrain - beta[2]*logBody
y <- exp(logBrainAdjusted)
ymod <- 100*y/median(y)
plot(ymod ~ Gestation, log="xy",
     xlab="Average Gestation Length (Days); Log Scale",
     ylab="Brain Weight Adjusted for Body Weight, as a Percentage of the Median",
     main="Brain Weight Adjusted for Body Weight, Versus Gestation Length, for 96 Mammal Species",
     pch=21,bg="green",cex=1.3)
identify(ymod ~ Gestation,labels=Species, cex=.7) # Identify points, as desired
# Press Esc to complete identify.
abline(h=100,lty=2) # Draw horizontal line at 100%

myLm4 <- lm(logBrain ~ logBody + Litter)
beta <- myLm4$coef
logBrainAdjusted <- logBrain - beta[2]*logBody
y2 <- exp(logBrainAdjusted)
y2mod <- 100*y2/median(y2)
plot(y2mod ~ Litter, log="y", xlab="Average Litter Size",
     ylab="Brain Weight Adjusted for Body Weight, as a Percentage of the Median",
     main="Brain Weight Adjusted for Body Weight, Versus Litter Size, for 96 Mammal Species",
     pch=21,bg="green",cex=1.3)
identify(y2mod ~ Litter,labels=Species, cex=.7)
abline(h=100,lty=2)

detach(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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

Examples

```
str(case1001)
attach(case1001)

## EXPLORATION
plot(Distance ~ Height)
myLm <- lm(Distance ~ Height)
plot(myLm, which=1)
height2 <- Height^2
myLm2 <- lm(Distance ~ Height + height2)
plot(myLm2, which=1)
summary(myLm2) # Note p-value for quadratic term (it's small)
height3 <- Height^3
myLm3 <- update(myLm2, ~ . + height3)
plot(myLm3, which=1)
summary(myLm3) # Note p-value for cubic term (it's small)
height4 <- Height^4
myLm4 <- update(myLm3, ~ . + height4)
summary(myLm4) # Note p-value for quartic term (it's not small)

## DISPLAY FOR PRESENTATION
plot(Distance ~ Height, xlab="Initial Height (Punti)",
     ylab="Horizontal Distance Traveled (Punti)",
     main="Galileo's Falling Body Experiment",
     pch=21, bg="green", lwd=2, cex=2)
dummyHeight <- seq(min(Height), max(Height), length=100)
betaQ <- myLm2$coef
quadraticCurve <- betaQ[1] + betaQ[2]*dummyHeight + betaQ[3]*dummyHeight^2
lines(quadraticCurve ~ dummyHeight, col="blue", lwd=3)
betaC <- myLm3$coef # coefficients of cubic model
cubicCurve <- betaC[1] + betaC[2]*dummyHeight + betaC[3]*dummyHeight^2 +
  betaC[4]*dummyHeight^3
lines(cubicCurve ~ dummyHeight, lty=3, col="red", lwd=3)
legend(590, 290, legend=c(expression("Quadratic Fit " * R^2 * " = 99.0%"),
  expression("Cubic Fit " * R^2 * " = 99.9%")),
  lty=c(1, 3), col=c("blue", "red"), lwd=c(3, 3))

detach(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.

Mass mass (in grams)

Type a factor with 3 levels indicating the type of flying vertebrate: non-echolocating bats, non-echolocating birds, echolocating bats

Energy in-flight energy expenditure (in W)

Source

Ramsey, F.L. and Schafer, D.W. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

References

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

Examples

```
str(case1002)
attach(case1002)

## EXPLORATION
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"))

logEnergy <- log(Energy)
logMass <- log(Mass)
myLm2 <- lm(logEnergy ~ logMass + Type + logMass:Type)
plot(myLm2, which=1)
myLm3 <- update(myLm2, ~ . - logMass:Type)
anova(myLm3, myLm2) # Test for interaction with extra ss F-test

## INFERENCE AND INTERPRETATION
myLm4 <- update(myLm3, ~ . - Type) # Reduced model...with no effect of Type
anova(myLm4, myLm3) # Test for Type effect
myType <- factor(Type,
  levels=c("non-echolocating bats", "echolocating bats", "non-echolocating birds"))
myLm3a <- lm(logEnergy ~ logMass + myType)
summary(myLm3a)
100*(exp(myLm3a$coef[3]) - 1)
100*(exp(confint(myLm3a,3))-1)
# Conclusion: Adjusted for body mass, the median energy expenditure for
# echo-locating bats exceeds that for echo-locating bats by an estimated
# 8.2% (95% confidence interval: 29.6% LESS to 66.3% MORE)

# DISPLAY FOR PRESENTATION
myPlotCode <- ifelse(Type=="non-echolocating birds",24,21)
myPointColor <- ifelse(Type=="echolocating bats","green","white")
plot(Energy ~ Mass, log="xy", xlab="Body Mass (g); Log Scale ",
```

```

ylab="In-Flight Energy Expenditure (W); Log Scale",
main="In-Flight Energy Expenditure Study",
pch=myPlotCode,bg=myPointColor,lwd=2, cex=1.5)
dummyMass <- seq(5,800,length=50)
beta      <- myLm3$coef
curve1    <- exp(beta[1] + beta[2]*log(dummyMass))
curve2    <- exp(beta[1] + beta[2]*log(dummyMass) + beta[3])
curve3    <- exp(beta[1] + beta[2]*log(dummyMass) + beta[4])
lines(curve1 ~ dummyMass)
lines(curve2 ~ dummyMass, lty=2)
lines(curve3 ~ dummyMass, lty=3)
legend(100,3,
      c("Echolocating Bats","Non-Echolocating Bats","Non-Echolocating Birds"),
      pch=c(21,21,24),lwd=2,pt.cex=c(1.5,1.5,1.5),pt.lwd=c(2,2,2),
      pt.bg=c("green","white","white"),lty=c(1,2,3))

detach(case1002)

```

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 $\mu\text{mol}/\text{min}/\text{g}$ of tissue)

Sex sex of the subject

Alcohol whether the subject is alcoholic or not

Source

Ramsey, F.L. and Schafer, D.W. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

Examples

```

str(case1101)
attach(case1101)

## EXPLORATION
library(lattice)
xyplot(Metabol~Gastric|Sex*Alcohol, case1101)

myPch <- ifelse(Sex=="Female",24,21)
myBg <- ifelse(Alcohol=="Alcoholic","gray","white")
plot(Metabol~Gastric, pch=myPch,bg=myBg,cex=1.5)
legend(1,12, pch=c(24,24,21,21), pt.cex=c(1.5,1.5,1.5,1.5),
      pt.bg=c("white","gray", "white", "gray"),
      c("Non-alcoholic Females", "Alcoholic Females",
        "Non-alcoholic Males", "Alcoholic Males"))
identify(Metabol ~ Gastric)
# Left click on outliers to show case number; Esc when finished.

myLm1 <- lm(Metabol ~ Gastric + Sex + Gastric:Sex)
plot(myLm1, which=1)
plot(myLm1, which=4) # Show Cook's Distance; note cases 31 and 32.
plot(myLm1, which=5) # Note leverage and studentized residual for cases 31 and 32.
subject <- 1:32 # Create ID number from 1 to 32

# Refit model without cases 31 and 32:
myLm2 <- update(myLm1, ~ ., subset = (subject !=31 & subject !=32))
plot(myLm2,which=1)
plot(myLm2,which=4)
plot(myLm2,which=5)
summary(myLm1)
summary(myLm2) # Significance of interaction terms hinges on cases 31 and 32.

myLm3 <- update(myLm2, ~ . - Gastric:Sex) #Drop interaction (without 31,32).
summary(myLm3)
if(require(car)){ # Use the car library
  crPlots(myLm3) # Show partial residual (component + residual) plots.
}

## INFERENCE AND INTERPRETATION
summary(myLm3)
confint(myLm3,2:3)

## DISPLAY FOR PRESENTATION
myCol <- ifelse(Sex=="Male","blue","red")
plot(Metabol ~ Gastric,
     xlab=expression("Gastric Alcohol Dehydrogenase Activity in Stomach ("*mu*"mol/min/g of Tissue)"),
     ylab="First-pass Metabolism in the Stomach (mmol/liter-hour)",
     main="First-Pass Alcohol Metabolism and Enzyme Activity for 18 Females and 14 Males",
     pch=myPch, bg=myBg,cex=1.75, col=myCol, lwd=1)
legend(0.8,12.2, c("Females", "Males"), lty=c(1,2),
      pch=c(24,21), pt.cex=c(1.75,1.75), col=c("red", "blue"))
dummyGastric <- seq(min(Gastric),3,length=100)
beta <- myLm3$coef
curveF <- beta[1] + beta[2]*dummyGastric
curveM <- beta[1] + beta[2]*dummyGastric + beta[3]
lines(curveF ~ dummyGastric, col="red")

```

```
lines(curveM ~ dummyGastric, col="blue",lty=2)
text(.8,10,"gray indicates alcoholic",cex = .8, adj=0)

detach(case1101)
```

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)

Treatment 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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

Examples

```
str(case1102)
attach(case1102)

## EXPLORATION
logRatio <- log(Brain/Liver)
logTime <- log(Time)
myMatrix <- cbind(logRatio, Days, Weight, Loss, Tumor, logTime)
if(require(car)){ # Use the car library
  scatterplotMatrix(myMatrix,groups=Treatment,
    smooth=F, diagonal="histogram", col=c("green","blue"),pch=c(16,17),cex=1.5)
}
```

```

myLm1 <- lm(logRatio ~ Treatment + logTime + Days + Sex + Weight + Loss + Tumor)
plot(myLm1, which=1)
if(require(car)){ # Use the car library
  crPlots(myLm1) # Draw partial residual plots.
}

myLm2 <- lm(logRatio ~ Treatment + factor(Time) +
  Days + Sex + Weight + Loss + Tumor) # Include Time as a factor.
anova(myLm1, myLm2)
if(require(car)){ # Use the car library
  crPlots(myLm2) # Draw partial residual plots.
}

summary(myLm2) # Use backward elimination
myLm3 <- update(myLm2, ~ . - Days)
summary(myLm3)
myLm4 <- update(myLm3, ~ . - Sex)
summary(myLm4)
myLm5 <- update(myLm4, ~ . - Weight)
summary(myLm5)
myLm6 <- update(myLm5, ~ . - Tumor)
summary(myLm6)
myLm7 <- update(myLm6, ~ . - Loss)
summary(myLm7) # Final model for inference

## INFERENCE AND INTERPRETATION
myTreatment <- factor(Treatment, levels=c("NS", "BD")) # Change level ordering
myLm7a <- lm(logRatio ~ factor(Time) + myTreatment)
summary(myLm7a)
beta <- myLm7a$coef
exp(beta[5])
exp(confint(myLm7a, 5))
# Interpretation: The median ratio of brain to liver tumor counts for barrier-
# disrupted rats is estimated to be 2.2 times the median ratio for control rats
# (95% CI: 1.5 times to 3.2 times as large).

## DISPLAY FOR PRESENTATION
ratio <- Brain/Liver
jTime <- exp(jitter(logTime, .2)) # Back-transform a jittered version of logTime
plot(ratio ~ jTime, log="xy",
  xlab="Sacrifice Time (Hours), jittered; Log Scale",
  ylab="Effectiveness: Brain Tumor Count Relative To Liver Tumor Count; Log Scale",
  main="Blood Brain Barrier Disruption Effectiveness in 34 Rats",
  pch= ifelse(Treatment=="BD", 21, 24), bg=ifelse(Treatment=="BD", "green", "orange"),
  lwd=2, cex=2)
dummyTime <- c(0.5, 3, 24, 72)
controlTerm <- beta[1] + beta[2]*(dummyTime==3) +
  beta[3]*(dummyTime==24) + beta[4]*(dummyTime==72)
controlCurve <- exp(controlTerm)
lines(controlCurve ~ dummyTime, lty=1, lwd=2)
BDTerm <- controlTerm + beta[5]
BDCurve <- exp(BDTerm)
lines(BDCurve ~ dummyTime, lty=2, lwd=2)
legend(0.5, 10, c("Barrier disruption", "Saline control"), pch=c(21, 22),
  pt.bg=c("green", "orange"), pt.lwd=c(2, 2), pt.cex=c(2, 2), lty=c(2, 1), lwd=c(2, 2))

```

```
detach(case1102)
```

case1201	<i>State Average SAT Scores</i>
----------	---------------------------------

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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

Examples

```
str(case1201)
attach(case1201)

## EXPLORATION
logTakers <- log(Takers)
myMatrix <- cbind(SAT, logTakers, Income, Years, Public, Expend, Rank)
if(require(car)){ # Use the car library
  scatterplotMatrix(myMatrix, diagonal="histogram", smooth=F)
}
State[Public < 50] # Identify state with low Public (Louisiana)
State[Expend > 40] # Alaska
myLm1 <- lm(SAT ~ logTakers + Income + Years + Public + Expend + Rank)
plot(myLm1, which=1)
plot(myLm1, which=4) # Cook's Distance
State[29] # Identify State number 29? ([1] Alaska)
```



```

plot(myLm1,which=5)
if(require(car)){ # Use the car library
  crPlots(myLm1) # Partial residual plot
}
myLm2 <- update(myLm1, ~ . ,subset=(State != "Alaska"))
plot(myLm2,which=1)
plot(myLm2,which=4)
if(require(car)){ # Use the car library
  crPlots(myLm2) # Partial residual plot
}
## RANK STATES ON SAT SCORES, ADJUSTED FOR Takers AND Rank
myLm3 <- lm(SAT ~ logTakers + Rank)
myResiduals <- myLm3$res
myOrder <- order(myResiduals)
State[myOrder]

## DISPLAY FOR PRESENTATION
dotchart(myResiduals[myOrder], labels=State[myOrder],
  xlab="SAT Scores, Adjusted for Percent Takers and HS Ranks (Deviation From Average)",
  main="States Ranked by SAT Scores Adjusted for Percent of HS Students That Took the SAT and Average High School",
  bg="green", cex=.8)
abline(v=0, col="gray")

## VARIABLE SELECTION (FOR RANKING STATES AFTER ACCOUNTING FOR ALL VARIABLES)
expendSquared <- Expend^2
if(require(leaps)){ # Use the leaps library
  mySubsets <- regsubsets(SAT ~ logTakers + Income+ Years + Public + Expend +
    Rank + expendSquared, nvmax=8, data=case1201, subset=(State != "Alaska"))
  mySummary <- summary(mySubsets)
  p <- apply(mySummary$which, 1, sum)
  plot(p, mySummary$bic, ylab = "BIC")
  cbind(p,mySummary$bic)
  mySummary$which[4,]
  myLm4 <- lm(SAT ~ logTakers + Years + Expend + Rank, subset=(State != "Alaska"))
  summary(myLm4)

## DISPLAY FOR PRESENTATION
myResiduals2 <- myLm4$res
myOrder2 <- order(myResiduals2)
newState <- State[State != "Alaska"]
newState[myOrder2]
dotchart(myResiduals2[myOrder2], labels=State[myOrder2],
  xlab="SAT Scores, Adjusted for Demographic and Expenditure Variables (Deviation From Average)",
  main="States Ranked by SAT Scores Adjusted for Demographics of Takers and Education Expenditure",
  bg="green", cex = .8)
abline(v=0, col="gray")
}

detach(case1201)

```

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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

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)
attach(case1202)

## EXPLORATION
logSal <- log(Bsal)
myMatrix <- cbind(logSal, Senior, Age, Educ, Exper)
if(require(car)){ # Use the car library
  scatterplotMatrix(myMatrix, smooth=F, diagonal="histogram",
                    groups=Sex, col=c("red", "blue") )
}
myLm1 <- lm(logSal ~ Senior + Age + Educ + Exper + Sex)
plot(myLm1, which=1)
plot(myLm1, which=4) # Cook's Distance
if(require(car)){ # Use the car library
  crPlots(myLm1) # Partial residual plots
```

```

}
ageSquared    <- Age^2
ageCubed      <- Age^3
experSquared  <- Exper^2
experCubed    <- Exper^3
myLm2 <- lm(logSal ~ Senior + Age + ageSquared + ageCubed +
  Educ + Exper + experSquared + experCubed + Sex)
plot(myLm2, which=1) # Residual plot
plot(myLm1, which=4) # Cook's distance

if(require(leaps)){ # Use the leaps library
  mySubsets <- regsubsets(logSal ~ (Senior + Age + Educ + Exper +
    ageSquared + experSquared)^2, nvmax=25, data=case1202)
  mySummary <- summary(mySubsets)
  p <- apply(mySummary$which, 1, sum)
  plot(mySummary$bic ~ p, ylab = "BIC")
  cbind(p, mySummary$bic)
  mySummary$which[8,] # Note that Age:ageSquared = ageCubed
}
myLm3 <- lm(logSal ~ Age + Educ + ageSquared + Senior:Educ +
  Age:Exper + ageCubed + Educ:Exper + Exper:ageSquared)
summary(myLm3)

myLm4 <- update(myLm3, ~ . + Sex)
summary(myLm4)
myLm5 <- update(myLm4, ~ . + Sex:Age + Sex:Educ + Sex:Senior +
  Sex:Exper + Sex:ageSquared)
anova(myLm4, myLm5)

## INFERENCE AND INTERPRETATION
summary(myLm4)
beta <- myLm4$coef
exp(beta[6])
exp(confint(myLm4,6))
# Conclusion: The median beginning salary for males was estimated to be 12%
# higher than the median salary for females with similar values of the available
# qualification variables (95% confidence interval: 7% to 17% higher).

## DISPLAY FOR PRESENTATION
years <- Exper/12 # Change months to years
plot(Bsal ~ years, log="y", xlab="Previous Work Experience (Years)",
  ylab="Beginning Salary (Dollars); Log Scale",
  main="Beginning Salaries and Experience for 61 Female and 32 Male Employees",
  pch= ifelse(Sex=="Male",24,21), bg = "gray",
  col= ifelse(Sex=="Male","blue","red"), lwd=2, cex=1.8 )
myLm6 <- lm(logSal ~ Exper + experSquared + experCubed + Sex)
beta <- myLm6$coef
dummyExper <- seq(min(Exper),max(Exper),length=50)
curveF <- beta[1] + beta[2]*dummyExper + beta[3]*dummyExper^2 +
  beta[4]*dummyExper^3
curveM <- curveF + beta[5]
dummyYears <- dummyExper/12
lines(exp(curveF) ~ dummyYears, lty=1, lwd=2,col="red")
lines(exp(curveM) ~ dummyYears, lty = 2, lwd=2, col="blue")
legend(28,8150, c("Male","Female"),pch=c(24,21), pt.cex=1.8, pt.lwd=2,
  pt.bg=c("gray","gray"), col=c("blue","red"), lty=c(2,1), lwd=2)

```

```
detach(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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

References

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

Examples

```
str(case1301)
attach(case1301)

## EXPLORATION AND MODEL DEVELOPMENT
plot(Cover ~ Treat,xlab="Animals Present",ylab="Remaining Seaweed Coverage (%)")
myLm1 <- lm(Cover ~ Block + Treat + Block:Treat)
plot(myLm1,which=1)
ratio <- Cover/(100 - Cover)
logRatio <- log(ratio)
myLm2 <- lm(logRatio ~ Block + Treat + Block:Treat)
plot(myLm2, which=1)
myLm3 <- lm(logRatio ~ Block + Treat)
anova(myLm3, myLm2) # Test for interaction with extra ss F-test
if(require(car)){ # Use the car library
```

```

crPlots(myLm3) # Partial residual plots
myLm4 <- lm(logRatio ~ Treat)
anova(myLm4, myLm3) # Test for Block effect
myLm5 <- lm(logRatio ~ Block)
anova(myLm5, myLm3) # Test for Treatment effect
lmp <- factor(ifelse(Treat %in% c("L", "Lf", "LfF"), "yes", "no"))
sml <- factor(ifelse(Treat %in% c("f", "fF", "Lf", "LfF"), "yes", "no"))
big <- factor(ifelse(Treat %in% c("fF", "LfF"), "yes", "no"))
myLm6 <- lm(logRatio ~ Block + lmp + sml + big)
crPlots(myLm6)
myLm7 <- lm(logRatio ~ Block + (lmp + sml + big)^2)
anova(myLm6, myLm7) # Test for interactions of lmp, sml, and big

## INFERENCE AND INTERPRETATION
summary(myLm6) # Get p-values for lmp, sml, and big effects; R^2 = .8522
beta <- myLm6$coef
exp(beta[9:11])
exp(confint(myLm6, 9:11) )

myLm7 <- update(myLm6, ~ . - lmp)
summary(myLm7) # R^2 = .4568; Note .8522-.4580 = 39.54# (explained by limpets)
myLm8 <- update(myLm6, ~ . - big)
summary(myLm8) # R^2 = .8225; Note .8522-.8255= 2.97# (explained by big fish)
myLm9 <- update(myLm6, ~ . - sml)
summary(myLm9) # R^2: .8400; Note .8522-.8400 = 1.22# (explained by small fish)

## DISPLAY FOR PRESENTATION
myYLab <- "Adjusted Seaweed Regeneration (Log Scale; Deviation from Average)"
crPlots(myLm6, ylab=myYLab, ylim=c(-2.2,2.2),
  main="Effects of Blocks and Treatments on Log Regeneration Ratio, Adjusted for Other Factors")
}

detach(case1301)

```

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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

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)
attach(case1302)

## EXPLORATION AND MODEL DEVELOPMENT
plot(Score ~ as.numeric(Company),cex=1.5, pch=21,
     bg=ifelse(Treat=="Pygmalion","blue","light gray"))
myLm1 <- lm(Score ~ Company + Treat + Company:Treat) # Fit with interaction.
plot(myLm1,which=1) # Plot residuals.
myLm2 <- update(myLm1, ~ . - Company:Treat) # Refit, without interaction.
anova(myLm2, myLm1) # Show extra-ss-F-test p-value (for interaction effect).
if(require(car)){ # Use the car library
  crPlots(myLm2)
}

## INFERENCE
myLm3 <- update(myLm2, ~ . - Company) # Fit reduced model without Company.
anova(myLm3, myLm2) # Test for Company effect.
summary(myLm2) # Show estimate and p-value for Pygmalion effect.
confint(myLm2,11) # Show 95% CI for Pygmalion effect.

## DISPLAY FOR PRESENTATION
beta <- myLm2$coef
partialRes <- myLm2$res + beta[11]*ifelse(Treat=="Pygmalion",1,0) # partial res
boxplot(partialRes ~ Treat, # Boxplots of partial residuals for each treatment
       ylab="Average Test Score, Adjusted for Company Effect (Deviation from Company Average)",
       names=c("19 Control Platoons","10 Pygmalion Treated Platoons"),
       col="green", boxlwd=2, medlwd=2,whisklty=1, whisklwd=2, staplewex=.2,
       staplelwd=2, outlwd=2, outpch=21, outbg="green", outcex=1.5 )

detach(case1302)
```

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 "Booe", "Cindy", "Bruno" and "Thelma"

Sign a factor indicating word taught, with 10 levels

Source

Ramsey, F.L. and Schafer, D.W. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

References

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

Examples

```
str(case1401)
attach(case1401)

## EXPLORATION AND MODEL DEVELOPMENT
plot(Minutes ~ Sign)
myLm1 <- lm(Minutes ~ Chimp + Sign)
plot(myLm1, which=1) # Plot residuals (indicates a need for transformation).
logMinutes <- log(Minutes)
myLm2 <- lm(logMinutes ~ Chimp + Sign)
plot(myLm2, which=1) # This looks fine.
if(require(car)){ # Use the car library
  crPlots(myLm2) # Partial residual plots
}

## INFERENCE AND INTERPRETATION
myLm3 <- update(myLm2, ~ . - Chimp) # Fit reduced model without Chimp.
anova(myLm3, myLm2) # Test for Chimp effect.
myLm4 <- update(myLm2, ~ . - Sign) # Fit reduced model without Sign.
anova(myLm4, myLm2) # Test for Sign effect.
# Fit 2-way model without intercept to order signs from easiest to hardest
```

```

myAov1 <- aov(logMinutes ~ Sign + Chimp - 1)
sort(myAov1$coef[1:10]) # Show the ordering of Signs
orderedSign <- factor(Sign, levels=c("listen", "drink", "shoe", "key", "more",
  "food", "fruit", "hat", "look", "string")) # Re-order signs, easiest 1st
myAov2 <- aov(logMinutes ~ orderedSign + Chimp - 1) # Refit
opar <- par(no.readonly=TRUE) # Store current graphics parameters settings
par(mar=c(4.1, 7.1, 4.1, 2.1)) # Adjust margins to allow room for y-axis labels
if(require(multcomp)){ # Use the multcomp library
  myMultComp <- glht(myAov2, linfct = mcp(orderedSign = "Tukey"))
  plot(myMultComp) # Plot Tukey-adjusted confidence intervals.
  summary(myMultComp) # Show Tukey-adjusted p-values pairwise comparisons
  confint(myMultComp) # Show Tukey-adjusted 95% confidence intervals
}
par(opar) # Restore original graphics parameters settings

## DISPLAY FOR PRESENTATION
myYLab <- "Log Learning Time, Adjusted for Chimp Effect"
myXLab <- "Sign Learned"
if(require(car)){ # Use the car library
  crPlots(myAov2, ylab=myYLab, xlab=myXLab,
    main="Learning Times by Sign, Adjusted for Chimp Effects",
    layout=c(1,1)) # Click on graph area to show next page (Just use first one.)
}

detach(case1401)

```

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. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

References

Heggstad, 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)
attach(case1402)

## EXPLORATION AND MODEL DEVELOPMENT; FORREST CULTIVAR
logForrest <- log(Forrest)
# Fit model without interactions first--to examine partial residual plots.
myLm1 <- lm(logForrest ~ Stress + SO2 + O3)
if(require(car)){ # Use the car library
  crPlots(myLm1) # Partial res plots => linear effects of SO2 and O3 look ok.
}
myLm2 <- lm(logForrest ~ (Stress + SO2 + O3)^2) # all 2-factor interactions.
plot(myLm1, which=1) # Residual plot looks ok.
anova(myLm1, myLm2) # Test for interactive effects.

## INFERENCE AND INTERPRETATION; FORREST CULTIVAR
summary(myLm1)
betaF <- myLm1$coef
# Effect of 0.01 increase in SO2 (note coeff is negative):
100*(1 - exp(0.01*betaF[3]))
#1.655701; a 1.65% decrease in median yield
100*(1-exp(0.01*confint(myLm1,"SO2")))
#3.772277 -0.5074294: 95% onfidence interval for effect of 0.01 increase in SO2
# Effect of 0.01 increase in O3 (note coeff is negative):
100*(1 - exp(0.01*betaF[4]))
# 5.585979 I.e. a 5.6% decrease in median yield
100*(1-exp(0.01*confint(myLm1,"O3")))
#7.445966 3.688613: 95% confidence interval for effect of 0.01 increase in O3
# Effect of water stress (note coeff is positive for effect of well-watered):
100*(1 - exp(-betaF[2])) # Get estimate for negative of this beta
#3.220556: a 3.2% decrease in median yield due to water stress
100*(1-exp(-confint(myLm1,2)))
#-7.875521 13.17529: 95% confidence interval

## DISPLAY FOR PRESENTATION; FORREST CULTIVAR
jO3 <- jitter(O3, factor=.25) # Jitter for plotting; jittering factor 0.25.
jS <- jitter(SO2, factor=.25) # Jitter SO2 for plotting.
cexfac <- 1.75 # Use character expansion factor of 1.75 for plotting symbols.
opar <- par(no.readonly=TRUE) # Store current graphics parameters settings
par(mfrow=c(1,2)) # Make a panel of 2 plots in 1 row.
myPointCode <- ifelse(Stress=="Well-watered", 21, 24)
myPointColor <- ifelse(Stress=="Well-watered", "green", "orange")
par(mar=c(4.1, 4.1, 2.1, 0.1))
plot(Forrest ~ jO3, log="y", ylab="Yield of Forrest Cultivar (kg/ha)",
     xlab=expression(paste(italic("Ozone"), mu, "L/L", jittered))),
     pch=myPointCode, lwd=2, bg=myPointColor, cex=cexfac)
legend(.02, 2400, c("Well-watered", "Water Stressed"), pch=c(21, 24),
      pt.cex=cexfac, pt.bg=c("green", "orange"), pt.lwd=2, lty=c(3, 1), lwd=c(2, 2))
dummy0 <- seq(min(O3), max(O3), length=2)
curve1 <- exp(betaF[1] + betaF[3]*mean(SO2) + betaF[4]*dummy0)
```

```

curve2    <- exp(betaF[1] + betaF[2] + betaF[3]*mean(S02)+ betaF[4]*dummy0)
lines(curve1 ~ dummy0,lwd=2)
lines(curve2 ~ dummy0,lwd=2,lty=3)

# PLOT FORREST VS S02
par(mar=c(4.1,2.1,2.1,2.1)) # Change margins
plot(Forrest ~ jS, log="y", ylab="",
xlab=expression(paste(italic("Sulfur Dioxide ("),mu,"L/L), jittered))),
  yaxt="n", pch=myPointCode, lwd=2, bg=myPointColor, cex=cexfac)
dummyS    <- seq(min(S02),max(S02),length=2)
curve1    <- exp(betaF[1] + betaF[3]*dummyS + betaF[4]*mean(O3))
curve2    <- exp(betaF[1] + betaF[2] + betaF[3]*dummyS + betaF[4]*mean(O3))
lines(curve1 ~ dummyS,lwd=2)
lines(curve2 ~ dummyS,lwd=2,lty=3)
par(opar) # Restore previous graphics parameter settings

detach(case1402)

```

Sleuth3Manual

Manual of the R Sleuth3 package

Description

If the option “pdfviewer” is set, this command will display the PDF version of the help pages.

Usage

```
Sleuth3Manual()
```

Author(s)

Berwin A Turlach <Berwin.Turlach@gmail.com>

References

Ramsey, F.L. and Schafer, D.W. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Cengage Learning.

Examples

```
## Not run: Sleuth3Manual()
```

Index

*Topic **datasets**

case0101, [2](#)
case0102, [3](#)
case0201, [4](#)
case0202, [5](#)
case0301, [6](#)
case0302, [8](#)
case0401, [9](#)
case0402, [10](#)
case0501, [11](#)
case0502, [13](#)
case0601, [14](#)
case0602, [16](#)
case0701, [17](#)
case0702, [18](#)
case0801, [20](#)
case0802, [21](#)
case0901, [22](#)
case0902, [23](#)
case1001, [25](#)
case1002, [26](#)
case1101, [28](#)
case1102, [30](#)
case1201, [32](#)
case1202, [33](#)
case1301, [36](#)
case1302, [37](#)
case1401, [39](#)
case1402, [40](#)

*Topic **documentation**

Sleuth3Manual, [42](#)

*Topic **package**

Sleuth3-package, [2](#)

case0101, [2](#)
case0102, [3](#), [34](#)
case0201, [4](#)
case0202, [5](#)
case0301, [6](#)
case0302, [8](#)
case0401, [9](#)
case0402, [10](#)
case0501, [11](#)
case0502, [13](#)

case0601, [14](#)
case0602, [16](#)
case0701, [17](#)
case0702, [18](#)
case0801, [20](#)
case0802, [21](#)
case0901, [22](#)
case0902, [23](#), [24](#)
case1001, [25](#)
case1002, [26](#)
case1101, [28](#)
case1102, [30](#)
case1201, [32](#)
case1202, [4](#), [33](#)
case1301, [36](#)
case1302, [37](#)
case1401, [39](#)
case1402, [40](#)

ex0727, [18](#)
ex0816, [19](#)
ex2011, [9](#)
ex2018, [5](#)
ex2223, [9](#)

Sleuth3 (Sleuth3-package), [2](#)
Sleuth3-package, [2](#)
Sleuth3Manual, [42](#)