

Package ‘Sleuth3’

June 22, 2012

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

Version 0.0-3

Date 2012-06-22

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. (2012), “The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)”, Duxbury.

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

LazyData yes

Depends R (>= 2.12.0), multcomp

Suggests lattice

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	7
case0302	8
case0401	10
case0402	11
case0501	13
case0502	14
Sleuth3Manual	16
Index	17

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

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="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 one 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. (2012). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd 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

```
attach(case0101)           # Attach the data frame case0101 to the R search path.
str(case0101)              # Show the structure of the data frame case0101.

boxplot(Score ~ Treatment) # Draw a boxplot of Score for each level of Treatment.

boxplot(Score ~ Treatment, # Draw boxplots, with ...
        ylab= "Average Creativity Score From 11 Judges (on a 40-point scale)", # ...y-axis label,
        names=c("23 'Extrinsic' Group Students", "24 'Intrinsic' Group Students"), # ...names below the boxplots, and
        main= "Haiku Creativity Scores for 47 Creative Writing Students") # ...main title.

detach(case0101)           # Detach the data frame from the R search path.
```

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. (2012). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd 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

```
attach(case0102)           # Attach the data frame case0102 to the R search path.
str(case0102)              # Show the structure of the data frame case0102.

boxplot(Salary ~ Sex)      # Draw boxplots of Salary for each level of Sex.

boxplot(Salary ~ Sex,      # Draw boxplots, with...
        ylab= "Starting Salary (U.S. Dollars)",      # ...y-axis label,
        names=c("61 Females", "32 Males"),          # ...names below the boxplots, and
        main= "Harris Bank Entry Level Clerical Workers, 1969-1971") # ...main title.

hist(Salary[Sex=="Female"]) # Draw a histogram of Salary for females only.
if(interactive()) dev.new()  # Open a new graphic window.
hist(Salary[Sex=="Male"])    # Draw a histogram of Salary for males only.

t.test(Salary ~ Sex,       # Perform the two-sample t-test...
       var.equal=TRUE)     # ...using the equal variance version.
t.test(Salary ~ Sex,       # Perform the two-sample t-test...
       var.equal=TRUE,     # ...using the equal variance version...
       alternative = "less") # ...and with alternative hypothesis that Group 1
detach(case0102)           # Detach the data frame from the R search path.
```

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. (2012). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Duxbury.

References

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

See Also

[ex2018](#)

Examples

```
attach(case0201)          # Attach the data frame case0201 to the R search path.
str(case0201)             # Show the structure of the data frame case0201.

boxplot(Depth ~ Year)      # Draw boxplots of Depth for each level of Year.

mean(Depth[Year==1978]) - mean(Depth[Year==1976])    # Calculate the average Depth in Year 1978 minus the

yearFactor <- factor(Year)          # Create a factor with 2 levels ("1976" and "1978") from the
t.test(Depth ~ yearFactor,          # Perform the two-sample t-test (Syntax: numerical variable
      var.equal=TRUE)              # ...using the equal variance version ("T" means TRUE).
t.test(Depth ~ yearFactor,          # Perform the two-sample t-test,
      var.equal=TRUE,              # ...using the equal-variance version, and
      alternative = "less")        # ...with alternative hypothesis that the Group 1 (1976, which

boxplot(Depth ~ Year,              # Draw boxplots of Depth for each level of Year, with ...
      ylab= "Beak Depth (mm)",      # ...y-axis label
      names=c("89 Finches in 1976", "89 Finches in 1978"),    # ...names below the boxplots, and
      main= "Beak Depths of Darwin Finches in 1976 and 1978")    # ...main title.

## BOXPLOTS FOR PRESENTATION
boxplot(Depth ~ Year,              # Draw boxplots of Depth for each level of Year, with
      ylab="Beak Depth (mm)",      # ...y-axis label,
      names=c("89 Finches in 1976", "89 Finches in 1978"),    # ...names below the two boxplots,
      main="Beak Depths of Darwin Finches in 1976 and 1978",    # ...main title,
      col="green",                # ...green fill color,
      boxlwd=2,                   # ...line widths in box equal to twice their default value,
      medlwd=2,                   # ...line widths of horizontal lines at medians equal twice their d
      whisklty=1,                 # ...solid (=1) (not dashed =2) whiskers (lines extending from box
      whisklwd=2,                 # ...line width of whiskers equal to twice their default value,
      staplewex=.2,              # ...length of staple (horizontal line at end of whisker) 0.2 time
      staplelwd=2,               # ...line width of staple equal to twice its default value,
      outlwd=2,                  # ...line width of outliers (plotted points beyond whiskers) equal
      outpch=21,                 # ...plotting character for outlier set to be a circle (21 is the co
      outbg="green",             # ...background (fill) color for the outlier set to green, and
      outcex=1.5 )               # ...outlier size set to be 1.5 times the default value.

detach(case0201)              # Detach the data frame from the R search path.
```

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.

Unaffected 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. (2012). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd 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

```
attach(case0202)                                # Attach the data frame case0201 to the R search path.
str(case0202)                                   # Show the structure of the data frame case0202.

diff <- Unaffected-Affected                     # Calculate the difference from each pair.
summary(diff)                                  # Show summary statistics for the variable "diff".
boxplot(diff)                                  # Draw a boxplot of diff. (Note syntax: boxplot(oneVariable))
abline(h=0)                                    # Draw a horizontal line at 0 on the plot.

t.test(diff)                                   # Perform a one-sample t-test on differences (Syntax: t.test(x, mu=0))
t.test(Unaffected,Affected,pair=TRUE)          # Perform a paired t-test on Unaffected and Affected

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

## Example of a polished boxplot for presentation:
boxplot(diff,                                  # Draw a boxplot of diff, with
  ylab="Difference in Hippocampus Volume (cubic cm)", # ...y-axis label,
  xlab="15 Sets of Twins, One Affected with Schizophrenia", # ...x-axis label,
  main="Hippocampus Difference: Unaffected Minus Affected Twin", # ...main title,
  col="green", # ...green fill color,
```

```

    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)

```

```

# ...line widths in box equal to twice their default value,
# ...line widths of horizontal lines at medians equal twice th
# ...solid (=1) (not dashed =2) whiskers (lines extending fro
# ...line width of whiskers equal to twice their default valu
# ...length of staple (horizontal line at end of whisker) 0.
# ...line width of staple equal to twice its default value,
# ...line width of outliers (plotted points beyond whiskers)
# ...plotting character for outlier set to be a circle (21 is
# ...background (fill) color for the outlier set to green, a
# ...outlier size set to be 1.5 times the default value.
# Draw a dashed (lty=2) horizontal line (h=...) at 0 for r

# Detach the data frame from the R search path.

```

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 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. (2012). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd 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

```
attach(case0301)
str(case0301)
```

```

# Attach the data frame case0301 to the R search path.
# Show the structure of the data frame.

```

```
boxplot(Rainfall ~ Treatment)
boxplot(log(Rainfall) ~ Treatment)
```

```

# Draw boxplots of Rainfall for each level of Treatment.
# Draw boxplots of the natural logarithm of Rainfall f

```

```

t.test(log(Rainfall) ~ Treatment,                # Perform the two-sample t-test of the null hypothesis
alternative="greater",                          # ...alternative hypothesis that group 1 mean (Seeded, which
var.equal=TRUE)                                # ...use the equal variance version.

myTest <- t.test(log(Rainfall) ~ Treatment,      # Perform two-sample t-test and store results in object
alternative="two.sided",                       # ...alternative that group means differ (for getting a2
var.equal=TRUE)                                # ...use the equal-variance version.
exp(myTest$est[1] - myTest$est[2])              # Back-transform the estimated difference in means on original scale
exp(myTest$conf)                               # Back-transform the endpoints of the confidence interval.

boxplot(log(Rainfall) ~ Treatment,              # Draw boxplots of log(Rainfall) for each level of Treatment
ylab="Log of Rainfall Volume in Target Area (Acre Feet)", # ...y-axis label,
names=c("On 26 Seeded Days", "On 26 Unseeded Days"), # ...names below boxes, and
main="Distributions of Rainfalls from Cloud Seeding Experiment") # ...main title.

## Example of polished display for presentation:
boxplot(log(Rainfall) ~ Treatment,              # Draw boxplots of log(Rainfall) for each level of Treatment
ylab="Log Rainfall (Acre-Feet)",                # ...y-axis label,
names=c("on 26 seeded days", "on 26 unseeded days"), # ...names below boxes,
main="Boxplots of Rainfall on Log Scale",        # ...main title
col="green",                                    # ...green fill color,
boxlwd=2,                                       # ...line widths in box equal to twice their default value,
medlwd=2,                                     # ...line widths of horizontal lines at medians equal to twice their default value,
whisklty=1,                                   # ...solid (=1) (not dashed =2) whiskers (lines extending from boxes)
whisklwd=2,                                   # ...line width of whiskers equal to twice their default value,
staplewex=.2,                                 # ...length of staple (horizontal line at end of whisker) 0.2 times default value,
staplelwd=2,                                  # ...line width of staple equal to twice its default value,
outlwd=2,                                     # ...line width of outliers (plotted points beyond whiskers) equal to twice default value,
outpch=21,                                    # ...plotting character for outlier set to be a circle (21 is the default)
outbg="green",                                # ...background (fill) color for the outlier set to green, and
outcex=1.5)                                   # ...outlier size set to be 1.5 times the default value.
myTicks <- c(1,5, 10,100,500,1000,2000,3000)    # Create values for y-axis ticks on the original scale
axis(4, at=log(myTicks),                      # Add tick marks at log(myTicks) to the right hand side of plot
label=myTicks)                                # ...use MyTicks as labels.
mtext("Rainfall (Acre Feet)",                 # Add this text outside of the plotting region,
side=4,                                       # ...on side 4 (right hand side), and
line=2.5)                                    # ...2.5 lines outside the plot boundary.

detach(case0301)                             # Detach case0301 from R Search path.

```

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. (2012). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Duxbury.

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)                                # Attach data frame case0302 to R search path
str(case0302)                                   # Show the structure of the data frame.

boxplot(Dioxin ~ Veteran)                       # Draw boxplots of Dioxin for each level

t.test(Dioxin ~ Veteran,                        # Perform two-sample t-test of the null
       alternative="less",                     # ...with the alternative hypothesis that
       var.equal=TRUE)                        # ...use the equal variance version of the
t.test(Dioxin ~ Veteran,                       # Perform the two-sample t-test,
       subset=(Dioxin < 40),                  # ...applied to the subset of observations
       alternative="less",                     # ...with alternative that group 1 mean is
       var.equal=TRUE)                        # ...using the equal variance version of the
t.test(Dioxin ~ Veteran,                       # Perform the two-sample t-test, with
       subset=(Dioxin < 20),                  # ...the subset of observations for which
       alternative="less",                     # ...with alternative that group 1 mean is
       var.equal=TRUE)                        # ...using the equal variance version of the
t.test(Dioxin ~ Veteran,                       # Perform the two-sample t-test (with de
       var.equal=TRUE)                        # ...using the equal variance version of the

## Example of polished histograms for presentation
par(mfrow=c(2,1),                             # Note: histograms seem to reveal
     mar=c(2,2,1,1))                          # Set graphical parameters, with mfrow=c(
myBreaks <- (0:46) - .5                        # ...with margins of 2 lines below and to t
hist(Dioxin[Veteran=="Other"],                 # Create a variable of breaks at -.5, 1.5
     breaks=myBreaks,                         # Draw a histogram of Dioxin for Other
     xlim=range(Dioxin),                     # ...using myBreaks to define histogram bi
     col="green",                            # ...with the x-axis spanning the entire
     xlab="",                                # ...with green fill color,
     ylab="",                                # ...with no x-axis label,
     main="")                                # ...with no y-axis label,
text(6,25,                                     # ...and no main title.
     "Dioxin in 97 'Other' Veterans; Estimated mean = 4.19 ppt (95% CI: 3.72 to 4.65 ppt)", # Write the following text at x=6, y=25:
     pos=4,                                   # ...so that text is left justified at that po
     cex=.75)                                # ...using character expansion factor 0.75.
hist(Dioxin[Veteran=="Vietnam"],              # Draw a histogram of Dioxin for Viet
     breaks=myBreaks,                         # ...using myBreaks to define histogram bi
     xlim=range(Dioxin),                     # ...with the x-axis spanning the entire
```

```

col="green",                                # ...with green fill color,
xlab="",                                   # ...with no x-axis label,
ylab="",                                   # ...with no y-axis label,
main="")                                   # ...and no main title.
text(6,145,                                # Write the following text at x=6, y=145:
     "Dioxin in 646 Vietnam Veterans; Estimated mean = 4.26 ppt (95% CI: 4.06 to 4.64 ppt)",
     pos=4,                                # ...so that text is left justified at that po
     cex=.75)                              # ...using character expansion factor 0.75.
text(25,70,                                # Write the following text at x=25, y = 70:
     "Estimated Difference in Means: 0.07 ppt (95% CI: -0.63 to 0.48 ppt)")

detach(case0302)                            # Detach case0302 from the R search path.

```

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. (2012). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd 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)                                # Show the structure of the data frame.
attach(case0401)                             # Attach the data frame to the R search path.

mCool <- mean(Incidents[Launch=="Cool"])      # Compute average number of incidents for Cool Launches.
mWarm <- mean(Incidents[Launch=="Warm"])      # Compute average number of incidents for Warm Launches.
mDiff <- mCool - mWarm                       # Compute the difference in averages.
c(mCool,mWarm,mDiff)                         # Show the results.

## PERMUTATION TEST BASED ON DIFFERENCE IN AVERAGES, VIA REPEATED RANDOM RE-GROUPING (ADVANCED)
numRep <- 50000                                # Specify the number of random regroupings (of the 24 scores into groups)
rDiff <- rep(0,numRep)                         # Create a variable rDiff to store the differences in averages for the permutation test
for (rep in 1:numRep) {                       # Do the following commands in parentheses numRep (50,000) times:
  randomGroup <- rep("rWarm",24)              # ...make a character variable of size 24, initially all with value "rWarm"
  randomGroup[sample(1:24,4)] <- "rCool"       # ...change 4 of these, at random, to "rCool",
  mW <- mean(Incidents[randomGroup=="rWarm"]) # ...calculate the average number of Incidents from the 20 rWarm Launches
  mC <- mean(Incidents[randomGroup=="rCool"]) # ...calculate the average number of Incidents from the 4 rCool Launches
  rDiff[rep] <- mC-mW                         # ...store the difference in averages from this random re-grouping in rDiff[rep]
}
hist(rDiff,                                  # Draw a histogram of the differences from random re-groupings (which are stored in rDiff)
  main="Approximate Permutation Distribution", # ...title,
  xlab="Possible Values of Difference in Averages", # ...x-axis label, and
  ylab="Frequency of Occurrence")              # ...y-axis label.
abline(v=mDiff)                              # Draw a vertical line at the actually observed difference in averages
pValue <- sum(rDiff >= 1.3)/numRep             # Calculate a one-sided p-value from the permutation test.
pValue                                         # Show the result.
text(mDiff,75000,                             # Include text on the histogram at x=mDiff and y=75000,
  paste("-->",round(pValue,4)),                # ... including an arrow and the p-value rounded to 4 digits, and
  adj=-0.1)                                  # ... adjusted to start a little to the right of mDiff

detach(case0401)                             # Detach the data frame from the R search path.

```

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. (2012). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd 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) # Show the structure of the data frame.
attach(case0402) # Attach the data frame to the R search path.

boxplot(Time ~ Treatment) # Draw a box plot of Time for each level of the factor Treatment.
median(Time[Treatment=="Conventional"])-median(Time[Treatment=="Modified"]) # Compute the difference in median completion times.

wilcox.test(Time ~ Treatment, # Compute the Wilcoxon rank-sum test (same syntax as wilcox.test)
  exact=FALSE, # ...using the normal approximation, not the exact test.
  correct=TRUE, # ...the continuity correction to improve the normal approximation.
  alternative="greater") # ...with alternative hypothesis that group 1 ("Conventional") has higher times.
wilcox.test(Time ~ Treatment, # Compute the Wilcoxon rank-sum test,
  exact=FALSE, # ...using the normal approximation,
  correct=TRUE, # ...using continuity correction,
  alternative="two.sided", # ...using 2-sided alternative (to get 2-sided p-value)
  conf.int=TRUE) # ...including a confidence interval.

## DOT PLOTS FOR PRESENTATION
xTreatment <- ifelse(Treatment=="Conventional",1,2) # Make a numerical code for Treatment.
myPointCode <- ifelse(Censored==0,21,24) # Create a plotting code equal to 20 if uncensored, 21 if censored.
plot(Time ~ jitter(xTreatment,.2), # Draw a scatterplot of Time vers xTreatment
  ylab="Completion Time (Sec.)", # ... with y-axis label,
  xlab="Training Method", # ... with x-axis label,
  main="Test Completion Times from Cognitive Load Experiment", # ... with title,
  axes=FALSE, # ... with axes suppressed (to be added later),
  pch=myPointCode, # ... using point codes in myPointCode,
  bg="green", # ... with point code fill color green,
  cex=2, # ... with character expansion of point symbols twice
  xlim=c(.5,2.5) ) # ... with x-axis extending between .5 and 2.5.
axis(2) # Add a y-axis, using default values.
axis(1, # Add an x-axis, with
  tick=FALSE, # ... no tick marks,
  at=c(1,2), # ... labels at x = 1 and x = 2, and
  labels=c("Conventional (n=14 Students)","Modified (n=14 Students)") ) # ... using these labels.
legend(1.5,300, # Add a plot legend with upper left corner at x = 1.5, y = 300.
  legend=c("Did not Complete in 300 sec","Completed in 300 sec."), # ... these descriptions,
  pch=c(24,21), # ... these plotting characters,
  pt.cex=2, # ... point character expansion factor 2, and
  pt.bg="green") # ... point fill character green.

detach(case0402) # Detach the data frame from the R search path.
```

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. (2012). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd 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)
attach(case0501)

myDiet <- factor(Diet, levels=c("NP", "N/N85", "N/R50", "R/R50", "lopro", "N/R40")) # Re-order levels to a more m
myNames <- c("NP(49)", "N/N85(57)", "N/R50(71)", "R/R50(56)", "lopro(56)", "N/R40(60)") # Specify boxplot label na
boxplot(Lifetime ~ myDiet,                                     # Draw box plot of Lifetime for each level of m
       ylab= "Lifetime (months)",                             # ...y-axis label,
       names=myNames,                                         # ...names to go under the boxplots, and
       xlab="Treatment (and sample size)")                   # ...x-axis label.
myAov1 <- aov(Lifetime ~ Diet)                                # Perform analysis of variance calculations.
```

```

plot(myAov1, which=1)
summary(myAov1)
pairwise.t.test(Lifetime,Diet,
  pool.SD=TRUE,
  p.adj="none")

## p-VALUES AND CONFIDENCE INTERVALS FOR SPECIFIED COMPARISONS OF MEANS, USING GLHT IN MULTCOMP PACKAGE
diet <- factor(Diet,labels=c("lopro", "NN85", "NR40", "NR50", "NP", "RR50")) # Rename levels to omit "/" (ne
myAov2 <- aov(Lifetime ~ diet - 1)
library(multcomp)
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"))
confint(myComparisons, calpha = univariate_calpha())

## 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)

```

Plot residuals versus fitted values.
 # Show analysis of variance table and F-test p-values.
 # Perform t-tests for all possible pairs of means.
 # ...using the equal variance (i.e. equal SD) version.
 # ...with no multiple comparison adjustment.

 # Perform analysis of variance calculations,
 # Attach multcomp library (after installing the package)
 # Perform multiple comparisons on myAov3 results.
 # ...on the specified contrasts

 # Show the p-values, without adjustment.
 # Show confidence intervals, without adjustment.

 # Draw boxplots of Lifetime for each level of myDiet.
 # ...y-axis label,
 # ...names under the boxplots,
 # ...main title,

 # ...green fill color,
 # ...line widths in box equal to twice their default.
 # ...line widths of horizontal lines at medians equal to twice their default.
 # ...solid (=1) (not dashed =2) whiskers (lines extend to 1.5 times the box height).
 # ...line width of whiskers equal to twice their default.
 # ...length of staple (horizontal line at end of whiskers) equal to twice its default.
 # ...line width of staple equal to twice its default.
 # ...line width of outliers (plotted points beyond whiskers) equal to twice its default.
 # ...plotting character for outlier set to be a circle.
 # ...background (fill) color for the outlier set to be green.
 # ...outlier size set to be 1.5 times the default value.

 # Detach data frame from R search path.

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. (2012). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd 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) # Show structure of data frame.
attach(case0502) # Attach data frame to R search path.

## EXPLORATION
myNames <- c("A (5)", "B (6)", "C (9)", "D (2)", "E (6)", "F (9)", "Spock's (9)") # Make labels for boxplots.
boxplot(Percent ~ Judge, # Draw boxplots of Percent for each level of
  ylab = "Percent of Women on Judges' Venires", # ...y-axis label,
  names = myNames, # ...names under the boxplots,
  xlab = "Judge (and number of venires)", # ...x-axis label, and
  main = "Percent Women on Venires of 7 Massachusetts Judges") # ...main title.
myAov1 <- aov(Percent ~ Judge) # Perform analysis of variance calculation
plot(myAov1, which=1) # Draw a plot of residuals versus fitted values
summary(myAov1) # Show the F-test for hypothesis that all 7 judges have the same mean

## ANALYSIS 1. TWO-SAMPLE t-TEST (ASSUMING NON-SPOCK JUDGES HAVE A COMMON MEAN)
SpockOrOther <- ifelse(Judge=="Spock's", "Spock", "Other") # Create a factor with 2 levels, 0 for Spock, 1 for other
aovFull <- aov(Percent ~ Judge) # Fit the full model with 7 means for 7 judges
aovReduced <- aov(Percent ~ SpockOrOther) # Fit the reduced model with 2 means; 1 for Spock, 1 for other
anova(aovReduced, aovFull) # Perform the extra sum of squares F-test to compare the two models
t.test(Percent ~ SpockOrOther, # Perform a two-sample t-test, comparing Spock and other
  var.equal=TRUE) # ...using equal variance version.

## ANALYSIS 2. COMPARE SPOCK MEAN TO AVERAGE OF OTHER MEANS (NOT ASSUMING NON-SPOCK JUDGES HAVE A COMMON MEAN)
library(multcomp) # Attach multcomp library (after installing the package)
myAov3 <- aov(Percent ~ Judge - 1) # Fit with the model with separate means (1 for Spock, 0 for other)
myContrast <- rbind(c(1/6, 1/6, 1/6, 1/6, 1/6, 1/6, - 1)) # Create a one-row matrix with coefficients
myComparison <- glht(myAov3, linfct=myContrast) # Perform "general linear hypothesis testing"
summary(myComparison, test=adjusted("none")) # Show a p-value for a test that the contrast is zero
confint(myComparison) # Show a confidence interval for the contrast

## BOXPLOTS FOR PRESENTATION
boxplot(Percent ~ Judge, # Draw boxplots of Percent for each level of
  ylab = "Percent of Women on Judges' Venires", # ...y-axis label,
```

```

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)

```

```

# ...names under the boxplots,
# ...x-axis label,
# ...main title
# ...green fill color,
# ...line widths in box equal to twice their default,
# ...line widths of horizontal lines at medians equal to twice their default,
# ...solid (=1) (not dashed =2) whiskers (lines equal to twice their default),
# ...line width of whiskers equal to twice their default,
# ...length of staple (horizontal line at end of box) equal to twice its default,
# ...line width of staple equal to twice its default,
# ...line width of outliers (plotted points beyond whiskers) equal to twice their default,
# ...plotting character for outlier set to be a circle,
# ...background (fill) color for the outlier set to be green,
# ...outlier size set to be 1.5 times the default value.

# Detach data frame from R search path.

```

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. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (3rd ed)*, Duxbury.

Examples

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


Index

*Topic **datasets**

- case0101, [2](#)
- case0102, [3](#)
- case0201, [4](#)
- case0202, [5](#)
- case0301, [7](#)
- case0302, [8](#)
- case0401, [10](#)
- case0402, [11](#)
- case0501, [13](#)
- case0502, [14](#)

*Topic **documentation**

- Sleuth3Manual, [16](#)

*Topic **package**

- Sleuth3-package, [2](#)

- case0101, [2](#)
- case0102, [3](#)
- case0201, [4](#)
- case0202, [5](#)
- case0301, [7](#)
- case0302, [8](#)
- case0401, [10](#)
- case0402, [11](#)
- case0501, [13](#)
- case0502, [14](#)
- case1202, [3](#)

- ex2011, [10](#)
- ex2018, [5](#)
- ex2223, [10](#)

- Sleuth3 (Sleuth3-package), [2](#)
- Sleuth3-package, [2](#)
- Sleuth3Manual, [16](#)