Package ‘Sleuth3’
March 7, 2013

Title       Data sets from Ramsey and Schafer’s “Statistical Sleuth (3rd ed)”
Version     0.1-6
Date        2013-03-17
Author      Original by F.L. Ramsey and D.W. Schafer, modifications by Daniel W. Schafer, Jeannie Sif- neos and Berwin A. Turlach
Maintainer  Berwin A Turlach <Berwin.Turlach@gmail.com>
LazyData    yes
Depends     R (>= 2.15.0)
Suggests    MASS, lattice, multcomp, car, leaps, CCA, Hmisc
License     GPL (>= 2)
URL         http://r-forge.r-project.org/projects/sleuth2/

R topics documented:

<p>| Sleuth3-package | case0101 | case0102 | case0201 | case0202 | case0301 | case0302 | case0401 | case0402 | case0501 | case0502 | case0601 | case0602 | case0701 | case0702 | case0801 | case0802 |
|-----------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|</p>
<table>
<thead>
<tr>
<th>Case/Example</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>case0901</td>
<td>25</td>
</tr>
<tr>
<td>case0902</td>
<td>26</td>
</tr>
<tr>
<td>case1001</td>
<td>28</td>
</tr>
<tr>
<td>case1002</td>
<td>29</td>
</tr>
<tr>
<td>case1101</td>
<td>31</td>
</tr>
<tr>
<td>case1102</td>
<td>33</td>
</tr>
<tr>
<td>case1201</td>
<td>35</td>
</tr>
<tr>
<td>case1202</td>
<td>37</td>
</tr>
<tr>
<td>case1301</td>
<td>39</td>
</tr>
<tr>
<td>case1302</td>
<td>40</td>
</tr>
<tr>
<td>case1401</td>
<td>42</td>
</tr>
<tr>
<td>case1402</td>
<td>43</td>
</tr>
<tr>
<td>case1501</td>
<td>45</td>
</tr>
<tr>
<td>case1502</td>
<td>47</td>
</tr>
<tr>
<td>case1601</td>
<td>48</td>
</tr>
<tr>
<td>case1602</td>
<td>50</td>
</tr>
<tr>
<td>case1701</td>
<td>52</td>
</tr>
<tr>
<td>case1702</td>
<td>54</td>
</tr>
<tr>
<td>case1801</td>
<td>57</td>
</tr>
<tr>
<td>case1802</td>
<td>59</td>
</tr>
<tr>
<td>case1803</td>
<td>60</td>
</tr>
<tr>
<td>case1901</td>
<td>61</td>
</tr>
<tr>
<td>case1902</td>
<td>62</td>
</tr>
<tr>
<td>case2001</td>
<td>64</td>
</tr>
<tr>
<td>case2002</td>
<td>66</td>
</tr>
<tr>
<td>case2101</td>
<td>68</td>
</tr>
<tr>
<td>case2102</td>
<td>70</td>
</tr>
<tr>
<td>case2201</td>
<td>71</td>
</tr>
<tr>
<td>case2202</td>
<td>73</td>
</tr>
<tr>
<td>ex0112</td>
<td>74</td>
</tr>
<tr>
<td>ex0116</td>
<td>75</td>
</tr>
<tr>
<td>ex0125</td>
<td>76</td>
</tr>
<tr>
<td>ex0126</td>
<td>76</td>
</tr>
<tr>
<td>ex0127</td>
<td>77</td>
</tr>
<tr>
<td>ex0211</td>
<td>78</td>
</tr>
<tr>
<td>ex0218</td>
<td>79</td>
</tr>
<tr>
<td>ex0221</td>
<td>80</td>
</tr>
<tr>
<td>ex0222</td>
<td>81</td>
</tr>
<tr>
<td>ex0223</td>
<td>82</td>
</tr>
<tr>
<td>ex0321</td>
<td>83</td>
</tr>
<tr>
<td>ex0323</td>
<td>83</td>
</tr>
<tr>
<td>ex0327</td>
<td>84</td>
</tr>
<tr>
<td>ex0330</td>
<td>85</td>
</tr>
<tr>
<td>ex0331</td>
<td>86</td>
</tr>
<tr>
<td>ex0332</td>
<td>86</td>
</tr>
<tr>
<td>ex0333</td>
<td>87</td>
</tr>
<tr>
<td>ex0428</td>
<td>88</td>
</tr>
<tr>
<td>ex0429</td>
<td>89</td>
</tr>
<tr>
<td>ex0430</td>
<td>89</td>
</tr>
<tr>
<td>ex0431</td>
<td>90</td>
</tr>
<tr>
<td>ex0432</td>
<td>91</td>
</tr>
<tr>
<td>ex0518</td>
<td>91</td>
</tr>
</tbody>
</table>
R topics documented:

ex1320 ................................................................. 136
ex1321 ................................................................. 137
ex1416 ................................................................. 138
ex1417 ................................................................. 139
ex1419 ................................................................. 140
ex1420 ................................................................. 141
ex1507 ................................................................. 142
ex1509 ................................................................. 142
ex1514 ................................................................. 143
ex1515 ................................................................. 144
ex1516 ................................................................. 144
ex1517 ................................................................. 145
ex1518 ................................................................. 146
ex1519 ................................................................. 146
ex1605 ................................................................. 147
ex1611 ................................................................. 148
ex1612 ................................................................. 149
ex1613 ................................................................. 149
ex1614 ................................................................. 150
ex1615 ................................................................. 151
ex1620 ................................................................. 152
ex1708 ................................................................. 152
ex1715 ................................................................. 153
ex1716 ................................................................. 154
ex1914 ................................................................. 155
ex1916 ................................................................. 156
ex1917 ................................................................. 156
ex1918 ................................................................. 157
ex1919 ................................................................. 158
ex1921 ................................................................. 159
ex1922 ................................................................. 159
ex1923 ................................................................. 160
ex2011 ................................................................. 161
ex2012 ................................................................. 162
ex2015 ................................................................. 163
ex2016 ................................................................. 164
ex2017 ................................................................. 165
ex2018 ................................................................. 166
ex2019 ................................................................. 167
ex2113 ................................................................. 167
ex2115 ................................................................. 168
ex2116 ................................................................. 170
ex2117 ................................................................. 170
ex2118 ................................................................. 171
ex2119 ................................................................. 172
ex2120 ................................................................. 173
ex2121 ................................................................. 174
ex2220 ................................................................. 175
ex2222 ................................................................. 175
ex2223 ................................................................. 176
ex2224 ................................................................. 177
ex2225 ................................................................. 178
### 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

#### Motivation and Creativity

### Description

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

References


Examples

```r
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 4/zero.noslash-point scale)",
  names=c("23 'Extrinsic' Group Students","24 'Intrinsic' Group Students"),
  main= "Haiku Creativity Scores for 47 Creative Writing Students")

detach(case0101)
```

---

### Case 0102

**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**


**References**


**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


References

See Also

ex0218

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)

---

Anatomical Abnormalities Associated with Schizophrenia

Description

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

Usage

case0202

Format

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

Unaffected  volume of left hippocampus of unaffected twin (in cm$^3$)

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


References


Examples

```r
attach(case/zero.noslash2/zero.noslash2)
str(case/zero.noslash2)

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(case/zero.noslash2)
```

---

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


References


Examples

```r
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)
```
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


References


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=(2,1), mar=c(3,3,1,1))  # 2 by 1 layout of plots; change margins

myBreaks <- (c(0:46) - .5)  # Make breaks for histogram bins

hist(Dioxin[Veteran=="Other"], breaks=myBreaks, xlab="", ylab="", main="")
"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(case32)

case41

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

case41

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


References


See Also

ex211, ex2223
Examples

```r
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)
```

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

```r
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
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

References

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

## 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,
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


References


Examples

```r
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(case/zero.noslash5/zero.noslash2)

---

### 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


References


Examples

```r
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
  myDunnett <- glht(myAov, linfct = mcp(myHandicap = "Dunnett"))
  summary(myDunnett)
  confint(myDunnett, 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(myDunnett,
       xlab="Difference in Mean Qualification Score (and Dunnet-adjusted CIs)"

## 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",
```
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

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

- **Percentage** The percentage 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

References

Examples
```
str(case0602)
attach(case0602)

## EXPLORATION
plot(Percentage ~ Length,
     xlab="Length of the Two Males",
     ylab="Percentage of Time Female Spent with Yellow-Sword Male",
     main="Percentage of Time Spent with Yellow Rather than Transparent Sword Male")
abline(h=50)  # Draw a horizontal line at 50% (i.e. the "no preference" line)
```
myAov <- aov(Percentage ~ Pair)
plot(myAov, which=1)  # Residual plot
summary(myAov)

# Explore possibility of linear effect, as in Display 6.5
myAov2 <- aov(Percentage ~ 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(Percentage ~ Length)
summary(myLm)

# ONE-SAMPLE t-TEST THAT MEAN PERCENTAGE = 50%, IGNORING MALE PAIR EFFECT
t.test(Percentage, mu=5/zero.noslash, alternative="greater")  # Get 1-sided p-value
t.test(Percentage, alternative="two.sided")  # Get C.I.

## SCATTERPLOT FOR PRESENTATION
plot(Percentage ~ Length,
     xlab="Length of the Two Males (mm)",
     ylab="Percentage 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/zero.noslash,lty=2,col="blue",lwd=2)
text(29.5,52,"5/zero.noslash% (no preference)", col="blue")

detach(case/zero.noslash6/zero.noslash2)

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
References


See Also

ex0725

Examples

```r
str(case0701)
attach(case0701)

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

myResiduals <- myLm$res
myFits <- myLm$fitted
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)

## INFEERENCE
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(-25,2.5,
       c("unrestricted regression line","regression through the origin"),
       lty=c(2,3), lwd=c(2,2), col=c("blue","red"))

detach(case0701)
```

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


References


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=.9)
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])
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)
lines(c(minTime, minTime), c(5, 6.15), col="purple", lwd=2)
legend(3.5, 6.25, "Desired pH", lty=1, cex=1.25)
detach(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**

```r
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**


**References**


**Examples**

```r
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
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(case812)

---

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

References

Examples
```r
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 INTERPRETATION
beta <- myLm$coef
1/(1 - exp(beta[2]))  # Back-transform estimated slope
1/(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=1, lwd=2, bg="green", cex=1.75 )
dummyVoltage <- c(min(Voltage),max(Voltage))
medianTime <- exp(meanLogTime)
lines(medianTime ~ dummyVoltage, lwd=2, col="blue")
```

detach(case0802)

---

**Effects of Light on Meadowfoam Flowering**

**Description**
Meadowfoam is a small plant found growing in moist meadows of the US Pacific Northwest. Researchers reported the results from one study in a series designed to find out how to elevate meadowfoam production to a profitable crop. In a controlled growth chamber, they focused on the effects of two light–related factors: light intensity and the timeing of the onset of the 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 $\mu$mol/m$^2$/sec)

Source

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")"),
    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(7,70,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)
```

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

See Also
ex0333

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=FALSE, # 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=FALSE, 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=1/zero.noslash/zero.noslash,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=100,1ty=2) # Draw horizontal line at 100%
detach(case/zero.noslash/zero.noslash9/zero.noslash2)

case1/zero.noslash/zero.noslash1

---

**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**

```r
case1001
```

**Format**

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

- **Distance**  horizontal distances (in punti)
- **Height**  initial height (in punti)
## Examples

```r
str(case1001)
attach(case1001)

## EXPLORATION
plot(Distance ~ Height)
myLm <- lm(Distance ~ Height)
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(59,29,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)
```

## 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

```r
case1002
```
Format

A data frame with 20 observations on the following 3 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


References


Examples

```r
str(case1/zero.noslash/zero.noslash2)
attach(case1/zero.noslash/zero.noslash2)

## EXPLORATION
plot(Energy~Mass, case1/zero.noslash/zero.noslash2, 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, 5, 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 ",

col=myPointColor)
```
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 °mol/min/g of tissue)
- **Sex**  sex of the subject
- **Alcohol**  whether the subject is alcoholic or not

Source

### EXPLORATION

library(lattice)

xyplot(Metabol~Gastric|Sex*Alcohol, case11/zero.noslash1)

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(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="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


lines(curveF ~ dummyGastric, col="red")
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


See Also

ex1416, ex1417
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=FALSE, 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",
     ..."
```r
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)
  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(case112)
```

---

**case1201**

*State Average SAT Scores*

**Description**

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

**Usage**

case1201

**Format**

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

- **State** US state
- **SAT** state averages of the total SAT (verbal + quantitative) scores
- **Takers** the percentage of the total eligible students (high school seniors) in the state who took the exam
- **Income** the median income of families of test-takers (in hundreds of dollars)
- **Years** the average number of years that the test-takers had formal studies in social sciences, natural sciences and humanities
- **Public** the percentage of the test-takers who attended public secondary schools
- **Expend** the total state expenditure on secondary schools (in hundreds of dollars per student)
- **Rank** the median percentile ranking of the test-takers within their secondary school classes

**Source**

Examples

```r
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=FALSE)
}
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 Adjusted SAT Scores",
  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"]
```
Sex discrimination in Employment

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

Usage
case1202

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bsal</td>
<td>Annual salary at time of hire</td>
</tr>
<tr>
<td>Sal77</td>
<td>Salary as of March 1975</td>
</tr>
<tr>
<td>Sex</td>
<td>Sex of employee</td>
</tr>
<tr>
<td>Senior</td>
<td>Seniority (months since first hired)</td>
</tr>
<tr>
<td>Age</td>
<td>Age of employee (in months)</td>
</tr>
<tr>
<td>Educ</td>
<td>Education (in years)</td>
</tr>
<tr>
<td>Exper</td>
<td>Work experience prior to employment with the bank (months)</td>
</tr>
</tbody>
</table>

Source

References

See Also
case0120
Examples

```r
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=FALSE, 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$which, 1, sum)
  cbind(p,mySummary$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)
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(28,815/zero,noslash, 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(case12/zero,noslash2)

case13/zero,noslash1
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


References

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/(1 - 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)
anoa(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)

myLm7 <- update(myLm6, ~ Block + (lmp + sml + big)^2)
anova(myLm7, myLm6) # 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))

myLm8 <- update(myLm6, ~ . - lmp)
summary(myLm8) # R^2 = .4568; Note .8522-.4580 = 39.54# (explained by limpets)

myLm9 <- update(myLm6, ~ . - big)
summary(myLm9) # R^2 = .8255; Note .8522-.8255 = 2.97# (explained by big fish)

myLm10 <- update(myLm6, ~ . - sml)
summary(myLm10) # 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)
**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**


**References**


**Examples**

```r
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,
staplelw=2, outlwd=2, outpch=21, outbg="green", outcex=1.5 )
detach(case13/zero.noslash2)

---

### 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**

```r
case14/zero.noslash1
```

**Format**

A data frame with 40 observations on the following 4 variables.

- **Minutes** learning time in minutes
- **Chimp** a factor indicating chimp, with four levels "Booee", "Cindy", "Bruno" and "Thelma"
- **Sign** a factor indicating word taught, with 10 levels
- **Order** the order in which the sign was taught

**Source**


**References**


**Examples**

```r
str(case14/zero.noslash1)
attach(case14/zero.noslash1)

## 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
}

## INFERENDE 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[1:1/zero.noslash]) # 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(case1402)

---

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

References

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")))
#7.449566 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)))
#-7.87521 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(2,24, 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))
dummyO <- seq(min(O3), max(O3), length=2)
curve1 <- exp(betaF[1] + betaF[3]*mean(SO2) + betaF[4]*dummyO)
lines(curve1 ~ dummyO,lwd=2)
lines(curve2 ~ dummyO,lwd=2,lty=3)
# PLOT FORREST VS SO2
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(SO2),max(SO2),length=2)
curve1 <- exp(betaF[1] + 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(case14)

case1501 Logging and Water Quality

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

Usage

case1501

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

Week week after the start of the study
Patch natural logarithm of nitrate level (ppm) in the logged watershed (ppm)
NoCut natural logarithm of nitrate level in the undisturbed watershed (ppm)

Source
References


Examples

```r
str(case1501)
attach(case1501)

## EXPLORATION
opar <- par(no.readonly=TRUE) # Store current graphics parameters settings
par(mfrow=c(2,1)) # Set graphics parameters: 2 row, 1 column layout
plot(NoCut ~ Week, type="b", ylab="Log of Nitrate Concentration; NoCut")
abline(h=mean(NoCut)) # Horizontal line at the mean
plot(Patch ~ Week, type="b", ylab="Log of Nitrate Concentration; Patch Cut")
abline(h=mean(Patch))
par(opar) # Restore previous graphics settings
lag.plot(NoCut,do.lines=FALSE) # Lag plot for NoCut
lag.plot(Patch,do.lines=FALSE) # Lag plot for Patch
pacf(NoCut) # partial autocorrelation function plot; noCut
pacf(Patch) # partial autocorrelation function plot; Patch

## INFERENCE (2-sample comparison, accounting for first serial correlation)
diff <- mean(Patch) - mean(NoCut)
nPatch <- length(Patch) # length of Patch series
nNoCut <- length(NoCut) # length of NoCut series
acfPatch <- acf(Patch, type="covariance") # auto covariances for Patch series
c0Patch <- acfPatch$acf[1]*nPatch/(nPatch-1) # variance; n-1 divisor (Patch)
c1Patch <- acfPatch$acf[2]*nPatch/(nPatch-1) # autocov; n-1 divisor (Patch)
acfNoCut <- acf(NoCut, type="covariance") # auto covariances for NoCut series
c0NoCut <- acfNoCut$acf[1]*nNoCut/(nNoCut - 1) # variance; n-1 divisor (NoCut)
c1NoCut <- acfNoCut$acf[2]*nNoCut/(nNoCut - 1) # autocov; n-1 divisor (NoCut)
dfPatch <- nPatch - 1 # DF (n-1); Patch
dfNoCut <- nNoCut - 1 # DF (n-1); NoCut
c0Pooled <- (dfPatch*c0Patch + dfNoCut*c0NoCut)/(dfPatch + dfNoCut)
c0Pooled # [1] 1.413295 = pooled estimate of variance
c1Pooled <- (dfPatch*c1Patch + dfNoCut*c1NoCut)/(dfPatch + dfNoCut)
c1Pooled # [1] 0.9103366 = pooled estimate of lag 1 covariance

# Pooled estimate of first serial correlation coefficient:
r1 <- c1Pooled/c0Pooled # [1] 0.6441233
SEdiff <- sqrt(1 + r1)/(1-r1))*sqrt(c0Pooled*(1/nPatch + 1/nNoCut))

# t-test and confidence interval
tStat <- diff/SEdiff # [1] 0.2713923
pValue <- 1 - pt(tStat,dfPatch + dfNoCut) # One-sided p-value
halfWidth <- qt(.975,dfPatch + dfNoCut)*SEdiff # half width of 95% CI
diff + c(-1,1)*halfWidth #95% CI -0.6557578 0.8648487

## GRAPHICAL DISPLAY FOR PRESENTATION
par(mfrow=c(1,1)) # Reset mfrow to a single plot per page
plot(exp(Patch) ~ Week, # Use exp(Patch) to show results in original units
     log="y", type="b", xlab="Weeks After Logging",
```
Description

The data are the temperatures (in degrees Celsius) averaged for the northern hemisphere over a full year, for years 1850 to 2010. The 161-year average temperature has been subtracted, so each observation is the temperature difference from the series average.

Usage

```
Usage
case1502
```

Format

A data frame with 161 observations on the following 2 variables.

- **Year** year in which yearly average temperature was computed, from 1850 to 2010
- **Temperature** northern hemisphere temperature minus the 161-year average (degrees Celsius)

Source


References


See Also

```
See Also
ex1519
```
Examples

str(case1502)
attach(case1502)

## EXPLORATION AND MODEL BUILDING
plot(Temperature ~ Year, type="b") # Type = "b" for both points and lines

yearSquared <- Year^2
yearCubed <- Year^3
myLm1 <- lm(Temperature ~ Year + yearSquared + yearCubed)
res1 <- myLm1$res
myPacf <- pacf(res1) # Partial autocorrelation from residuals
r1 <- myPacf$acf[1] #First serial correlation coefficient
n <- length(Temperature) # Series length = 161
v <- Temperature[2:n] - r1*Temperature[1:(n-1)] # Filtered response
ones <- rep(1-r1, n-1) # make a variable of all 1's
u1 <- Year[2:n] - r1*Year[1:(n-1)] # Filtered "ones"

u2 <- yearSquared[2:n] - r1*yearSquared[1:(n-1)] # Filtered X1
u3 <- yearCubed[2:n] - r1*yearCubed[1:(n-1)] # Filtered X2
myLm2 <- lm(v ~ u1 + u2 + u3 )
res2 <- myLm2$res
pacf(res2) # Looks fine; don’t worry about lag 4 marginal significance
plot(myLm2, which=1) # Residual plot
summary(myLm2) # Cubic term isn’t needed.
myLm3 <- update(myLm2, ~ . - u3) # Drop cubic term

## INFERENCE
summary(myLm3) # Everything remaining is statistically significant.

## GRAPHICAL DISPLAY FOR PRESENTATION
plot(Temperature ~ Year, xlab="Year",
ylab=expression(paste("Annual Average Temperature (Difference From Average), ",
degree,"C")),main="Annual Average Temperature in Northern Hemisphere; 1850-2010",
type="b", pch=21, lwd=2, bg="green", cex=1.5)
myFits <- myLm3$fit
lines(myFits ~ Year[2:161], col="blue", lwd=2)
legend(1850,0.6,"Quadratic Regression Fit, Adjusted for AR(1) Serial Correlation",
col="blue", lwd=2, box.lty=0)
detach(case1502)

---

Sites of Short- and Long-Term Memory

Description

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

case1601

Format

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

- **Monkey**  
  Monkey name

- **Treatment**  
  A treatment factor with levels "Control" and "Treated"

- **Week2**  
  Percentage of 20 objects taught 2 weeks prior to treatment that were correctly distinguished in the test

- **Week4**  
  Percentage of 20 objects taught 4 weeks prior to treatment that were correctly distinguished in the test

- **Week8**  
  Percentage of 20 objects taught 8 weeks prior to treatment that were correctly distinguished in the test

- **Week12**  
  Percentage of 20 objects taught 12 weeks prior to treatment that were correctly distinguished in the test

- **Week16**  
  Percentage of 20 objects taught 16 weeks prior to treatment that were correctly distinguished in the test

Source


References


Examples

```r
str(case1601)
attrib(case1601)

## EXPLORATION
short <- (Week2 + Week4)/2
long <- (Week8 + Week12 + Week16)/3
myPointCode <- ifelse(Treatment=="Control",15,16)
myPointColor <- ifelse(Treatment=="Control","orange","green")
plot(long ~ short, pch=myPointCode, col=myPointColor, cex=2)
abline(h=mean(long),lty=2)
abline(v=mean(short),lty=2)
identify(short,long,labels=Monkey)  # Identify outliers; press Esc when done

## INFERENCE USING HOTELLING's T-SQUARED TEST
myLm1 <- lm(cbind(short,long) ~ Treatment)  # Full model
myLm2 <- lm(cbind(short,long) ~ 1)  # Reduced model, with only intercept
anova(myLm2, myLm1, test="Hotelling")  # p-value for Treatment effect
# confidence intervals
n1 <- sum(Treatment=="Control")  # 7 control monkeys
n2 <- sum(Treatment=="Treated")  # 11 treated monkeys
multiplier <- sqrt(2*(n1+n2-2)/(n1+n2-3))*qf(.95,2,n1+n2-3))  # Sleuth p. 492
```
In a randomized, double-blind, crossover experiment, researchers randomly assigned 20 volunteer hospital employees to either a low-fiber or low-fiber treatment group. The subjects followed the diets for six weeks. After two weeks on their normal diet, all patients crossed over to the other treatment group for another six weeks. The total serum cholesterol (in mg/dl) was measured on each patient before the first treatment, at the end of the first six week treatment, and at the end of the second six week treatment.
Usage

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

Baseline total serum cholesterol before treatment
HiFiber total serum cholesterol after the high fiber diet
LoFiber total serum cholesterol after the low fiber diet
Order factor to identify order of treatment, with two levels "HL" and "LH"

Source


References


Examples

```r
str(case16/zero.noslash2)
attach(case16/zero.noslash2)
## EXPLORATION
highMinusBase <- HiFiber-Baseline
highMinusLow <- HiFiber-LoFiber
plot(highMinusBase ~ highMinusLow)
abline(h=0) # Horizontal line at 0
abline(v=0) # Vertical line at 0
# Hotelling 2-sample t-test for order effect on bivariate response:
myLm1 <- lm(cbind(highMinusBase,highMinusLow) ~ Order ) # Full model
myLm2 <- update(myLm1, ~ . - Order) # Reduced model without Order effect
anova(myLm2, myLm1, test="Hotelling") # p-value for Order effect

## INFERENCE: HOTELLING ONE-SAMPLE TEST THAT MEAN OF BIVARIATE RESPONSE IS (0,0)
myLm3 <- lm(cbind(highMinusBase, highMinusLow) ~ 1) # Full model
myLm4 <- update(myLm3, ~ . - 1) # Reduced model with both means = 0
anova(myLm4, myLm3, test="Hotelling") # test that the bivariate mean is (0,0)
# Confidence intervals
sumary(myLm3)
HighMinusBase <- myLm3$coef[[1]] # -13.85/zero.noslash
seHighMinusBase <- 3.533 # Standard error, read from summary(myLm3)
HighMinusLow <- myLm3$coef[[2]] # -0.850
seHighMinusLow <- 3.527 # Standard error, read from summary(myLm3)
n <- length(highMinusBase) # 20: sample size
multiplier <- sqrt(2*(n-1)/(n-2))*qf(.95,2,n-2)) # See Sleuth, page 495
HighMinusBase + c(-1,1)*multiplier*seHighMinusBase # 95% CI for High - Base
HighMinusLow + c(-1,2)*multiplier*seHighMinusLow # 95% CI for High - Low
```
## GRAPHICAL DISPLAY FOR PRESENTATION

```r
lowMinusBase <- LoFiber - Baseline
myPointCode <- ifelse(Order== "HL",21,22)
myPointColor <- ifelse(Order== "HL","green","light blue")
plot(highMinusBase ~ lowMinusBase,
     xlab="Cholesterol Change (from Baseline) After High Fiber Diet (mg/dl)",
     ylab="Cholesterol Change (From Baseline) After Low Fiber Diet (mg/dl)",
     main="Cholesterol Effects of High- and Low-Cholesterol Diets on 2/zero.noslash Subjects",
     ylim=c(-45,2/zero.noslash), pch=myPointCode, bg=myPointColor, cex=2.5, lwd=2)
abline(h=0) # Horizontal line at 0
abline(v=0) # Vertical line at 0
legend(-43,22,c("High Fiber Given First","Low Fiber Given First"), pch=c(21,22),
       pt.bg =c("green","light blue"),pt.cex =c(2.5,2.5), lw = c(2,2), lty=c(0,0))
detach(case16/zero.noslash2)
```

---

### Magnetic Force on Printer Rods

**Description**

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

**Usage**

```r
case17/zero.noslash1
```

**Format**

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

- `L1, L2, L3, L4, L5, L6, L7, L8, L9, L10, L11`: the magnetic force at each of the equally-spaces positions 1, 2, ..., 11 on the printer rod
- `Current`: electric current passing through the rod, with three levels "0", "250" and "500" (milliamperees)
- `Config`: a factor identifying the configuration, with two levels "0" and "1"
- `Material`: a factor identifying the type of metal from which the rod was made, with four levels "1", "2", "3" and "4"

**Source**

Examples

str(case1701)
attach(case1701)

## EXPLORATION
MagneticForces <- cbind(L1, L2, L3, L4, L5, L6, L7, L8, L9, L10, L11)
mfCor <- cor(MagneticForces)
round(mfCor, 2)  # Show correlations, rounded to two digits

mfPca <- prcomp(MagneticForces)  # principal components
summary(mfPca)  # Show the proportion of variance explained by each PC
plot(mfPca)  # Graph proportion of variances explained by each PC (Scree Plot)
mfCoefs <- mfPca$rotation  # Extract the coefficients
dim(mfCoefs)  # #11 rows and 11 columns
round(mfCoefs[, 1:3], 3)  # Show the first 3 columns of the score matrix, rounded

# Explore a possible meaningful linear combination suggested by first PC
round(mfCoefs[, 1], 1)  # Show coefficients of 1st pc, rounded to 1 digit
# Coefficients are all very similar, suggesting a constant coefficient; use 1/11
mfAve <- (L1 + L2 + L3 + L4 + L5 + L6 + L7 + L8 + L9 + L10 + L11)/11
mfScores <- mfPca$x
pc1 <- mfScores[, 1]  # Values for first principal component of MagneticForces
cor(mfAve, pc1)  # Correlation of the average and the first PC (>|0.999567|
plot(pc1 ~ mfAve)

# Explore a possible meaningful linear combination suggested by second PC
round(mfCoefs[, 2], 1)  # Show coefficients of 2nd pc, rounded to 1 digit
# Second set of coefficients are negative on the left end of the rod and
# positive on the right end. Try Ave(L9 + L1/zero.noslash + L11) - Ave(L1 + L2 + L3).
mfEnds <- (L9 + L1/zero.noslash + L11)/3 - (L1 + L2 + L3)/3
pc2 <- mfScores[, 2]
residualEnds <- lm(mfEnds ~ mfAve)$residual  # Ends with average effect removed
plot(pc2 ~ residualEnds)
cor(residualEnds, pc2)  # >0.973

# Explore a possible meaningful linear combination suggested by third PC
round(mfCoefs[, 3], 1)  # Show coefficients of 3rd pc, rounded to 1 digit
# Try a contrast between the first 4 positions and the 6th position
mfPeak <- (L1 + L2 + L3 + L4)/4 - L6
pc3 <- mfScores[, 3]
residualPeak <- lm(mfPeak ~ mfAve + mfEnds)$residual
plot(pc3 ~ residualPeak)
cor(residualPeak, pc3)  # >0.971
# Note: the variation explained by the third PC seems to be due almost entirely
# to one printer rod. (Keep this in mind.)

## INFERENCE: ANALYSIS OF EXPERIMENTAL FACTORS ON 3-DIMENSIONAL RESPONSE
myResponse <- cbind(mfAve, mfEnds, mfPeak)
cor(myResponse)
myLm1 <- lm(myResponse ~ Current + Config + Material)
anova(myLm1)  # Noticeable effect of Current but not Config or Material
plot(mfAve ~ Current)
myLm2 <- lm(mfAve ~ Current)
abline(myLm2)
summary(myLm2) # No evidence of an effect of current on average magnetic force

plot(mfEnds ~ Current)
myLm3 <- lm(mfEnds ~ Current)
abline(myLm3)
summary(myLm3) # Evidence that Current effects the difference in MF at the ends

plot(mfPeak ~ Current)
myLm4 <- lm(mfPeak ~ Current)
abline(myLm4)
summary(myLm4) # No evidence of an effect of Current on peak MF in the center

## GRAPHICAL DISPLAY FOR PRESENTATION
plot(mfEnds ~ jitter(Current,.1),
xlab="Electrical Current Used in Printer Rod Manufacture (milliamperes)",
ylab="Magnetic Force at Positions 9-11 Minus Magnetic Force at Positions 1-3",
main="Effect of Electrical Current on Magnetic Force Surrounding Printer Rod",
col="black", pch=21, lwd=2, bg="green", cex=2)
abline(myLm3, lwd=2)

detach(case1701)

description

Description

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

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

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

Usage

case1702

Format

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

**Couple** couple identification number

**HP** level of passionate love husband feels for spouse

**WP** level of passionate love wife feels for spouse

**HC** level of compassionate love husband feels for spouse
WC  level of compassionate love wife feels for spouse
PW  level of passionate love husband perceives spouse to have for him
PH  level of passionate love wife perceives spouse to have for her
CW  level of compassionate love husband perceives spouse to have for him
CH  level of compassionate love husband perceives spouse to have for her

Source

References

Examples
str(case1702)
attach(case1702)
## EXPLORATION
x <- cbind(HP,WP,HC,WC)  # 4 components of level of love you feel for spouse
y <- cbind(PW,PH,CW,CH)  # 4 components of level of love you perceive from spouse
if(require(CCA)){  # Use the CCA library
  myCCA <- cc(x,y)  # Store canonical correlation computations
  canCor <- myCCA$cor  # Extract the canonical correlations
  canCor  #[1] /zero.noslash.95/zero.noslash699/zero.noslash /zero.noslash.86656/zero.noslash1 /zero.noslash.5571876 /zero.noslash.11/zero.noslash6555
}
# Make a function to test the number of canonical correlations (advanced).
# Bartlett modification of likelihood ratio test
# Reference: Mardia, Kent, and Bibby, 1980, Multivariate Analysis,
myTest <- function(xMatrix,yMatrix) {
  if(require(CCA)){  # Use the CCA library
    myCCA <- cc(xMatrix,yMatrix)  # Store CCA computations
    canCor <- myCCA$cor  # extract the canonical correlations
    n <- dim(x)[1]  # number of rows of x, = sample size
    p <- dim(y)[2]  # number of component variables in y
    q <- dim(x)[2]  # number of component variables in x
    k <- min(p,q)  # the maximum number of canonical pairs
    testStat <- rep(0,k)  # store the test statistics; initially set to 0
    degFr <- rep(0,k)  # store the associated degrees of freedom
    canCor2 <- canCor[k:1]  # Reverse order for the following calculations
    productTerm <- 1
    for (i in 1:k) {
      productTerm <- productTerm*(1 - canCor2[i]^2)
      degFr[i] <- (p + 1 - i)*(q + 1 - i)
      testStat[i] <- -(n -(p+q+3)/2)*log(productTerm)
    }
    pair <- 1:k
    testStat <- round(testStat[1:k],2)  # Revert to original order; round
    pValue <- round(1 - pchisq(testStat,degFr),4)  # p-value to 4 digits
    canCor <- round(canCor,4)  # Round to 4 digits
    cbind(pair,canCor, testStat,degFr, pValue)  # Show the results;
myTest(x,y)
#
# Explore possible meaningful linear combination suggested by first pair of CCs
round(myCCA$xcoef,1)
round(myCCA$ycoef,1)
# The 1st column of xcoef is almost entirely HC; 1st column of ycoef is CW
ccX1 <- myCCA$scores$xscores[,1]
plot(ccX1 ~ jitter(HC)) # See if HC is a good substitute for 1st X canonical var
cor(ccX1,HC) # [1] 0.9719947

ccY1 <- myCCA$scores$yscores[,1]
plot(ccY1 ~ jitter(CW)) # See if CW is a good substitute for 1st y canonical var
cor(ccY1, CW) # [1] 0.9975468
#
# Analyze the correlation of the meaningful substitute variables

cor(HC,CW) # [1] 0.928323
myLm1 <- lm(HC ~ CW)
summary(myLm1) # p-value < 0.0001 (test for slope= equiv to test that corr = 0)

# Explore possible meaningful linear combination suggested by 2nd pair of CCs
# Suggested substitutes from 2nd columns above are WC and CH
ccX2 <- myCCA$scores$xscores[,2]
WCres <- lm(WC ~ HC)$res # WC with effect of HC removed
plot(ccX2 ~ WCres)
cor(ccX2,WCres) # [1] 0.9045225

ccY2 <- myCCA$scores$yscores[,2]
CHres <- lm(CH ~ CW)$res # CH with effect of CW removed
plot(ccY2 ~ CHres)
cor(ccY2,CHres) # [1] 0.9280248

cor(WC,CH) # [1] 0.8134213
myLm2 <- lm(WC ~ CH)
summary(myLm2) # p-value < 0.0001 for test that correlation = 0
#
# Explore cannonical correlations from other groupings of variables
x <- cbind(HP, HC, PW, CW) # husband's responses
y <- cbind(WP, WC, PH, CH) # wife's responses
myTest(x,y) # No evidence that husbands' responses are correlated with wifes'
x <- cbind(HP, PW, WP, PH) # passionate responses
y <- cbind(HC, CW, WC, CH) # compassionate responses
myTest(x,y) # No evidence that passionate anc compassionate responses are correlated

## GRAPHICAL DISPLAYS FOR PRESENTATION
jFactor <- 0.3 # Jittering factor (try different values to see what works)
jHC <- jitter(HC,factor=jFactor)
jCW <- jitter(CW,factor=jFactor)
jCH <- jitter(CH,factor=jFactor)
jWC <- jitter(WC,factor=jFactor)
opar <- par(no.readonly=TRUE) # Store current graphical parameter settings
par(mfrow=c(2,2)) # Prepare to make a 2x2 panel of graphs
par(mar=c(1.1,4.1,1.1,1.1)) # Adjust margins
plot(jHC ~ jCW, ylab="Husband's Compassionate Love For His Wife (Jittered)",
xlab="", ylim=c(3.5,1), xlim=c(3.5,1), col="black", pch=21, lwd=2,
Obesity and Heart Disease

Description

To better understand whether the relationship between heart disease and obesity could be due to the social stigma associated with obesity, researchers examined cardiovascular disease rates of obese and non-obese women in American Samoa, where obesity was considered socially desirable. 3,112 women were categorized according to whether they were obese or not and whether they died from cardiovascular disease (CVD).

Usage

case1801
Format

A data frame with 2 observations on the following 3 variables.

- **Obesity**: a factor with levels "NotObese" and "obese"
- **Deaths**: the number of women who died from CVD
- **NonDeaths**: the number that died from other causes

Source


References


Examples

```r
str(case18/zero.noslash1)
attach(case18/zero.noslash1)

## EXPLORATION
myTable <- cbind(Deaths,NonDeaths) # Form a 2 by 2 table of counts
row.names(myTable) <- Obesity # Assign the levels of Obesity as row names
myTable # Show the table

## INFERENCE (4 methods for getting p-values and confidence intervals)
prop.test(myTable, alternative="greater", correct=FALSE) # Compare 2 proportions
prop.test(myTable, alternative="greater", correct=TRUE) # ...with cont. corect.
prop.test(myTable,correct=TRUE) # 2-sided alternative (default) to get CI
chisq.test(myTable) # Pearson's Chi-Squared Test
fisher.test(myTable, alternative="greater") # Fisher's exact test
fisher.test(myTable) # 2-sided alternative to get CI for odds ratio
myGlm1 <- glm(myTable ~ Obesity, family=binomial) # Logistic reg (CH 21)
summary(myGlm1) # Get p-value-- /zero.noslash.734
beta <- myGlm1$coef
exp(beta[2]) #Odds of death are estimated to be 17% higher for obese women
exp(confint(myGlm1,2)) # 95% confidence interval

## GRAPHICAL DISPLAY FOR PRESENTATION
myTable
# Deaths NonDeaths
#Obese 16  2045
#NotObese  7   1044
prop.test(16,(16+2045)) #For one proportion, est: 0.0078 95% CI: 0.0046 to 0.013
prop.test(7,(7+1044)) #For one proportion, est: 0.0067 95% CI: 0.0029 to 0.014
phat <- c(0.007763222, 0.006660324)*1000 # Get estimated deaths per 1,000 women
lower95 <- c(0.00459943, 0.002921568)*1000
upper95 <- c(0.01287243, 0.014318321)*1000
if(require(Hmisc)) { # Use Hmisc library
  myObj <- Cbind(pHat,lower95,upper95)
  Dotplot(Obesity ~ myObj, # Draw a dot plot of estimates and CIs
           xlab="Estimated CVD Deaths Per 1,000 Women (and 95% Confidence Intervals)"
          )
```

Description

In a randomized experiment, researchers assigned 407 volunteers to receive 1,000 mg of Vitamin C daily throughout the cold season and 411 to receive a placebo. A physician interviewed the volunteers at the end of the study to determine whether or not they had suffered any colds during the study period.

Usage
case1802

Format

A data frame with 2 observations on the following 3 variables.

- **Treatment**: a factor with levels "Placebo" and "VitC"
- **Cold**: the number of who got colds
- **NoCold**: the number that did not get any colds

Source


References


Examples

```r
str(case1802)
attach(case1802)
library(MASS)
## INFERENCE (4 methods)
myTable <- cbind(Cold,NoCold)
row.names(myTable) <- c("Placebo","Vitamin C")
myTable
prop.test(myTable, alternative="greater") # Compare 2 binomial proportions
# Alternative: pop prop. of first column (cold) in larger in first row (placebo)
prop.test(myTable, alternative="greater", correct=TRUE)
prop.test(myTable,correct=TRUE) # Use 2-sided alternative to get CI
chisq.test(myTable) # Chi-square test
```
```r
fisher.test(myTable, alternative="greater")
fisher.test(myTable) # 2-sided alternative to get CI for odds ratio
myGlm1 <- glm(myTable ~ Treatment, family=binomial) # logistic reg (Ch 21)
summary(myGlm1)
beta <- myGlm1$coef
1 - exp(beta[2]) # 0.3479411
1 - exp(confint(myGlm1,2)) # 0.53365918 0.09042098
# Interpretation: The odds of getting a cold are 35% less on Vitamin C than 
# Placebo (95% confidence interval: 9% to 53% less).

detach(case182)
```

---

**Smoking and Lung Cancer**

**Description**

In a retrospective case-control study, researchers identified 86 lung cancer patients and 86 controls (without lung cancer), and categorized them according to whether they were smokers or non-smokers. The goal is to see whether the odds of lung cancer are greater for smokers than for non-smokers.

**Usage**

`case183`

**Format**

A data frame with 2 observations on the following 3 variables.

- **Smoking**: a factor with levels "NonSmokers" and "Smokers"
- **Cancer**: the number of who were lung cancer patients
- **Control**: the number who were controls

**Source**


**References**


**Examples**

```r
str(case183)
attach(case183)

## INFERENCE
myTable <- cbind(Cancer,Control) # Make a 2-by-2 table of counts
row.names(myTable) <- Smoking # Assign the levels of Smoking as row names
myTable
```
fisher.test(myTable, alternative="greater") # Alternative: that odds of Cancer # in first row are greater.
fisher.test(myTable) # 2-sided alternative to get CI for odds ratio
myGlm1 <- glm(myTable ~ Smoking, family=binomial) # logistic reg (Ch 21)
summary(myGlm1)
expt(myGlm1$coef[2]) # 5.37963 : Estimated odds ratio
expt(confint(myGlm1)[2,]) # 1.675169 24.009510: Approximate confidence interval
# Interpretation: The odds of cancer ar 5.4 times as large for smokers as for # non-smokers (95% confidence interval: 1.7 to 24.0 times as large).
detach(case18/zero.noslash3)

---

**Sex Role Sterotypes and Personnel Decisions**

**Description**

Researchers gave 48 male bank supervisors attending a management institute hypothetical personnel files and asked them whether they would promote the applicant based on the file. The personnel files were identical except that 24 of them listed a male and 24 listed a female applicant. The assignment of managers to receive either a male or female applicant file was carried out at random.

**Usage**

case1901

**Format**

A data frame with 2 observations on the following 3 variables.

- **Gender** a factor with levels "Female" and "Male"
- **Promoted** the number of managers who promoted the applicant
- **NotPromoted** the number of managers who did not promote the applicant

**Source**


**References**


**Examples**

```r
str(case1901)
attach(case1901)

## INFERENCE
myTable <- cbind(Promoted,NotPromoted)
row.names(myTable) <- Gender
```
myTable
fisher.test(myTable, alternative="greater")
# Alternative: that odds of Promotion in first row (Males) are greater.
fisher.test(myTable) # Use 2-sided to get confidence interval for odds ratio
prop.test(myTable) # Compare two binomial proportions

### GRAPHICAL DISPLAY FOR PRESENTATION
myTable
# Promoted NotPromoted
# Male 21 3
# Female 14 10
prop.test(21,(21+3)) # Est = .875; CI = .665 to .967
prop.test(14,(14+10))# Est = .583; CI = .369 to .772

pHat <- c(.875,.583)
lower95 <- c(.665, 0.369)
upper95 <- c(.967, .772)
if(require(Hmisc)) { # Use Hmisc library
  myObj<- Cbind(pHat,lower95,upper95) # Cbind: a form of cbind needed for Dotplot
  Dotplot(Gender ~ myObj,
    xlab="Probability of Promotion Based on Applicant File (and 95% Confidence Intervals)",
    ylab="Gender Listed in Applicant File", ylim=c(.5,2.5), cex=2)
}
detach(case1901)

descriptions:

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

Usage

case1902

Format

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

**Aggravation** the aggravation level of the crime, a numerical variable ranging from 1 to 6

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

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

**NoDeath** number in the aggravation and victim category who did not receive the death penalty
Source


References


Examples

```r
str(case19/zero.noslash2)
attach(case19/zero.noslash2)
## EXPLORATION
proportionDeath <- Death/(Death + NoDeath)
myPointCode <- ifelse(Victim=="White",22,24)
myPointColor <- ifelse(Victim=="White","white","black")
plot(proportionDeath ~ Aggravation, pch=myPointCode, bg=myPointColor)
oddsOfDeath <- Death/(NoDeath + .5) # Add .5 to the denominator to avoid 0's
plot(oddsOfDeath ~ Aggravation, pch=myPointCode, bg=myPointColor)
plot(oddsOfDeath ~ Aggravation, log="y", pch=myPointCode, bg=myPointColor)
# Use logistic regression (Ch 21) to see if the 6 odds ratios are constant
myGlm1 <- glm(cbind(Death,NoDeath) ~ Aggravation + Victim +
  Aggravation:Victim, family=binomial) # Logistic reg with interaction
myGlm2 <- update(myGlm1, ~ . - Aggravation:Victim) # without interaction
anova(myGlm2, myGlm1) # no evidence of interaction.

## INFERENCE
# Mantel Haenszel
myTable <- array(rbind(Death, NoDeath), dim=c(2,2,6),
  dimnames=list(Penalty=c("Death","No Death"), Victim=c("White","Black"),
  Aggravation=c("1","2","3","4","5","6")))
myTable # Show the 6 2x2 tables
mantelhaen.test(myTable, alternative="greater", correct=FALSE) # 1-sided p-value
mantelhaen.test(myTable, alternative="greater") # with continuity correction
mantelhaen.test(myTable) # two.sided (default) for confidence interval

# Logistic Regression (Ch 21) (treating aggravation level as numerical)
summary(myGlm2)
beta <- myGlm2$coef
exp(beta[3]) # 6.1144
exp(confint(myGlm2,3)) # 2.23/zero.noslash4/zero.noslash 18.72693
# Interpretation: The odds of death penalty for white victim murderers are
# estimated to be 6 times the odds of death penalty for black victim murderers
# with similar aggravation level (95% confidence interval: 2.2 to 18.7 times).

## GRAPHICAL DISPLAY FOR PRESENTATION
myPointColor <- ifelse(Victim=="White","green", "orange")
plot(jitter(proportionDeath,.1) ~ jitter(Aggravation,.1),
  xlab="Aggravation Level of the Murderer",
  ylab="Proportion of Murderers Who Received Death Penalty",
  pch=myPointCode, bg=myPointColor, cex=2, lwd=2)
legend(1,1, c("White Victim Murderers","Black Victim Murderers"), pch=c(21,22),
  pt.cex=c(2,2), pt.bg=c("green","orange"), pt.lw=c(2,2))
```

# Include logistic regression fit on plot
dummyAg <- seq(min(Aggravation),max(Aggravation),length=50)
etaB <- beta[1] + beta[2]*dummyAg
etaW <- etaB + beta[3]
pB <- exp(etaB)/(1 + exp(etaB)) # Estimated prob of DP; Black victim
pW <- exp(etaW)/(1 + exp(etaW)) # Estimated prob of DP; White victim
lines(pB ~ dummyAg,lty=1)
lines(pW ~ dummyAg,lty=2)
detach(case1902)

---

**Survival in the Donner Party**

### Description

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

### Usage

case2001

### Format

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

- **Age**  Age of person
- **Sex**  Sex of person
- **Status**  Whether the person survived or died

### Details

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

### Source


### References

See Also

\[\text{ex1918}\]

Examples

\begin{verbatim}
str(case2001)
attach(case2001)

## EXPLORATION AND MODEL BUILDING
myPointCode <- ifelse(Sex=="Female",22,24)
myPointColor <- ifelse(Sex=="Female","green","orange")
survivalIndicator <- ifelse(Status=="Survived",1,0)
jFactor <- 0.1  # jittering factor
plot(jitter(survivalIndicator,jFactor) ~ jitter(Age, jFactor),
pch=myPointCode, bg=myPointColor, cex=1.5)
# Logistic regression. Start with a rich model; use backward elimination
ageSquared <- Age^2
myGlm1 <- glm(Status ~ Age + ageSquared + Sex + Age:Sex + ageSquared:Sex, family=binomial)
# Use backward elimination, but remove interaction and squared terms 1st
summary(myGlm1)
myGlm2 <- update(myGlm1, ~ . - ageSquared:Sex)
summary(myGlm2)
myGlm3 <- update(myGlm2, ~ . - ageSquared)
summary(myGlm3)  # Wald test p-value for interaction of Age and Sex is: 0.0865
# More accurate likelihood ratio (drop in deviance) test:
myGlm4 <- update(myGlm3, ~ . - Age:Sex)
anova(myGlm4, myGlm3)  # Drop-in-deviance chi-square stat = 3.9909 on 1 d.f.
  1 - pchisq(3.9909,1)  # 2-sided p-value = 0.048

## INFERENCE AND INTERPRETATION
# Proceed by ignoring interaction (for a casual and approximate analysis)
myGlm5 <- update(myGlm4, ~ . - Sex)
anova(myGlm5, myGlm4)  # Drop-in-deviance chi-square statistic = 5.0344 on 1 d.f.
  1 - pchisq(5.0344,1)  # 2-sided p-value 0.02484869: Highly suggestive
  0.0248869/2  # 1-sided p-value = half the 2-sided p-value = 0.01244345
# Interpretation and confidence interval
Sex <- factor(Sex,levels=c("Male","Female"))  # Reorder levels so "Male" is ref
beta <- myGlm4b$coef
exp(beta[3])  # 4.939645
exp(confint(myGlm4b,3))  # 25.246069 1.215435
# Interpretation: The odds of survival for females are estimated to be 5 times
# the odds of survival of similarly-aged mean (95% CI: 1.2 times to 25.2 times).

## GRAPHICAL DISPLAY FOR PRESENTATION
myPointCode <- ifelse(Sex=="Female",22,24)
myPointColor <- ifelse(Sex=="Female","green","orange")
myLineColor <- ifelse(Sex=="Female","dark green","blue")
survivalIndicator <- ifelse(Status=="Survived",1,0)
jFactor <- 0.1
plot(jitter(survivalIndicator,jFactor) ~ jitter(Age, jFactor),
ylab="Estimated Survival Probability", xlab="Age (years)",
main=c("Donner Party Survival by Sex and Age"), xlim=c(15,75),
pch=myPointCode, bg=myPointColor, col=myLineColor, cex=2, lwd=3)
\end{verbatim}
beta <- myGlm4b$coef
dummyAge <- seq(15,65,length=50)
linearMale <- beta[1] + beta[2]*dummyAge  # log odds of survival for males
linearFemale <- linearMale + beta[3]  # log odds of survival for females
pCurveMale <- exp(linearMale)/(1 + exp(linearMale))  # survival prob; males
pCurveFemale <- exp(linearFemale)/(1 + exp(linearFemale))  # females
lines(pCurveMale ~ dummyAge,lty=2,col="blue",lwd=3)
lines(pCurveFemale[dummyAge <= 50] ~ dummyAge[dummyAge <= 50],lty=1,
     col="dark green",lwd=3)
legend(63,.5,legend=c("Females","Males"), pch=c(22,24),
     pt.bg = c("green","orange"), pt.cex=c(2,2), lty=c(1,2), lwd=c(3,3),
     col=c("dark green","blue"))
text(72,1,"Survived (20")
text(72,0,"Died (25")
detach(case2002)

---

Birdkeeping and Lung Cancer

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

Usage
case2002

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

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

Source
## References


## Examples

```r
str(case2002)
attach(case2002)

## EXPLORATION AND MODEL BUILDING
myCode <- ifelse(BK=="Bird" & LC=="LungCancer","Bird & Cancer",
    ifelse(BK=="Bird" & LC=="NoCancer","Bird & No Cancer",
        ifelse(BK=="NoBird" & LC=="LungCancer","No Bird & Cancer", "No Bird & No Cancer")))
table(myCode)
if(require(car)){ # Use the car library
    scatterplotMatrix(cbind(AG,YR,CD), groups=myCode, diagonal="none",reg.line=FALSE,
        pch=c(15,21,15,21), col=c("dark green","dark green","purple","purple"),
        var.labels=c("Age","Years Smoked","Cigarettes per Day"), cex=1.5)
}

# Reorder the levels so that the model is for log odds of cancer
LC <- factor(LC, levels=c("NoCancer","LungCancer"))
myGlm <- glm(LC ~ FM + SS + AG + YR + CD + BK, family=binomial)
if(require(car)){ # Use the car library
    crPlots(myGlm)  
}

# It appears that there's an effect of Years of Smoking and of Bird Keeping
# after accounting for other variables; no obvious effects of other variables

# Logistic regression model building using backward elimination (witholding BK)
myGlm1 <- glm(LC ~ FM + SS + AG + YR + CD, family=binomial)
summary(myGlm1)
myGlm2 <- update(myGlm1, ~ . - SS)
summary(myGlm2)
myGlm3 <- update(myGlm2, ~ . - CD)
summary(myGlm3)
myGlm4 <- update(myGlm3, ~ . - FM)
summary(myGlm4) # Everything left has a small p-value (retain the intercept)

## INFERENCE AND INTERPRETATION
myGlm5 <- update(myGlm4, ~ . + BK) # Now add bird keeping
summary(myGlm5)
myGlm6 <- update(myGlm5, ~ . + BK:YR + AG:YR) # Try interaction terms
anova(myGlm6,myGlm5) # Drop-in-deviance = 1.61 on 2 d.f.
1 - pchisq(1.61,2) # p-value = .45: no evidence of interaction
anova(myGlm4,myGlm5) # Test for bird keeping effect
(1 - pchisq(12.612,1))/2 # 1-sided p-value: 0.0001916391

BK <- factor(BK, levels=c("NoBird", "Bird")) # Make "no bird" the ref level
myGlm5b <- glm(LC ~ AG + YR + BK, family=binomial)
beta <- myGlm5b$coeff # Extract estimated coefficients
exp(beta[4]) # 3.961248
exp(confint(myGlm5b,4)) # 1.836764 8.900840

# Interpretation: The odds of lung cancer for people who kept birds were
# estimated to be 4 times the odds of lung cancer for people of similar age, sex,
# smoking history, and socio-economic status who didn't keep birds
```
# (95% confidence interval for this adjusted odds ratio: 1.8 times to 8.9 times).
# See bestglm library for an alternative variable selection technique.
detach(case2/zero.noslash/zero.noslash2)

case2101

*Island Size and Bird Extinctions*

**Description**

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

**Usage**

case2101

**Format**

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Island</td>
<td>Name of Island</td>
</tr>
<tr>
<td>Area</td>
<td>Area of Island</td>
</tr>
<tr>
<td>AtRisk</td>
<td>Number of species at risk</td>
</tr>
<tr>
<td>Extinct</td>
<td>Number of extinctions</td>
</tr>
</tbody>
</table>

**Details**

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

**Source**


**References**

Examples

```r
str(case2101)
attach(case2101)

## EXPLORATION AND MODEL BUILDING
proportionExtinct <- Extinct/AtRisk
oddsExtinct <- proportionExtinct/(1 - proportionExtinct)
logitExtinct <- log(oddsExtinct)  # Logit = Log Odds
logArea <- log(Area)
plot(logitExtinct ~ logArea)

binResponse <- cbind(Extinct, AtRisk - Extinct)
myGlm1 <- glm(binResponse ~ logArea, family=binomial)
summary(myGlm1)
logArea2 <- logArea^2
myGlm2 <- update(myGlm1, ~ . + logArea2)
summary(myGlm2) # p-value for quadratic term: 0.77

## INFERENCE AND INTERPRETATION
myGlm3 <- update(myGlm1, ~ . - logArea)
anova(myGlm3, myGlm1) # Drop in deviance statistic = 33.277 on 1 d.f.
1 - pchisq(33.277,1) # p-value = 7.992234e-09
beta <- myGlm1$coef
1 - 2^beta[2] # 0.1861153
1 - 2*confint(myGlm1,2) #0.2462041 0.1247743
# Interpretation: Associated with each doubling of island area is a 19%
# reduction in the odds of extinction (95% confidence interval: 12% to 25%
# reduction).

## GRAPHICAL DISPLAY FOR PRESENTATION
plot(oddsExtinct ~ Area, log="xy", ylab="Observed Odds of Extinction; log scale",
     xlab=expression(paste("Island Area (km"^{"2"},"); log scale")),
     main="Extinctions of Bird Species in the Krunnit Island Archipelago",
     pch=21, lwd=2, bg="green", cex=2) # Plot odds of extinction vs island area
dummyArea <- seq(min(Area),max(Area),length=50)
odds <- exp(lp)
lines(odds ~ dummyArea,lwd=2)

plot(proportionExtinct ~ Area, log="xy",
     ylab="Proportions of 1949 Species not Found in 1959",
     xlab=expression(paste("Island Area (km"^{"2"},"); log scale")),
     main="Proportions of 1949 Bird Species Extinct in 1959 on 18 Krunnit Archipelago Islands",
     pch=21, lwd=2, bg="green", cex=2) # Plot probability of extinction vs area
dummyArea <- seq(min(Area),max(Area),length=50)
myProbability <- exp(lp)/(1 + exp(lp))
lines(myProbability ~ dummyArea,lwd=2,col="blue")
legend(.08,.055,legend="Estimated Probability of Extinction",lty=1,lwd=2,col="blue")
detach(case2101)
```
Moth Coloration and Natural Selection

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

Usage
case2102

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

<table>
<thead>
<tr>
<th>Morph</th>
<th>Morph, a factor with levels &quot;light&quot; and &quot;dark&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>Distance from Liverpool (in km)</td>
</tr>
<tr>
<td>Placed</td>
<td>Number of moths placed</td>
</tr>
<tr>
<td>Removed</td>
<td>Number of moths removed</td>
</tr>
</tbody>
</table>

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

Source

References
Examples

```r
str(case2102)
attach(case2102)

## EXPLORATION AND MODEL BUILDING
proportionRemoved <- Removed/Placed
myPointCode <- ifelse(Morph=="dark",21,22)
myPointColor <- ifelse(Morph=="dark","blue","gray")
plot(proportionRemoved ~ Distance, pch=myPointCode, bg=myPointColor, cex=2, lwd=2)

binResponse <- cbind(Removed, Placed-Removed)
Morph <- factor(Morph, levels=c("light","dark"))  # Make "light" the ref level
myGlm1 <- glm(binResponse ~ Distance + Morph + Distance:Morph, family=binomial)
summary(myGlm1)  # Residual deviance: 13.230 on 10 degrees of freedom
1 - pchisq(13.230,10)  # No evidence of overdispersion
myGlm2 <- update(myGlm1, ~ . - Distance:Morph)
anova(myGlm2, myGlm1)  # Drop in deviance statistic = 11.931 on 1 d.f.
1 - pchisq(11.931,1)  # p-value = 0.00005520753 => strong evidence of interaction
# It appears that the intercepts are the same for both light and dark morphs,
# that there is no effect of Distance for light morphs, but there is an effect
# of Distance for dark morphs.

## INFERENCE AND INTERPRETATION
myTerm <- Distance*ifelse(Morph=="dark",1,0)  # Create indicator var for "dark"
myGlm3 <- glm(binResponse ~ myTerm, family=binomial)
summary(myGlm3)

## GRAPHICAL DISPLAY FOR PRESENTATION
myPointCode <- ifelse(Morph=="dark",22,24)
myPointColor <- ifelse(Morph=="dark","blue","orange")
plot(proportionRemoved ~ Distance, ylab="Proportion of Moths Taken",
main="Proportions of Moths Taken by Predators at Seven Locations",
xlab="Distance from Liverpool (km)", pch=myPointCode, bg=myPointColor, cex=2,
lwd=2)
beta <- myGlm3$coef
dummyDist <- seq(0,55,length=50)
propDark <- exp(lp)/(1 + exp(lp))
lines(propDark ~ dummyDist,lwd=2,col="blue")
propLight <- rep(exp(beta[1])/(1 + exp(beta[1])),length(dummyDist))
lines(propLight ~ dummyDist,lwd=2,col="orange")
legend(0,0.47,legend=c("Dark Morph","Light Morph"),
pch=c(22,24),pt.bg=c("blue","orange"),pt.cex=c(2,2),pt.lwd=c(2,2))
```

detach(case2102)

Description

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

**Usage**

```r
case2201
```

**Format**

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

- **Age**: Age of elephant at beginning of study
- **Matings**: Number of successful matings

**Source**


**References**


**Examples**

```r
str(case2201)
attach(case2201)

## EXPLORATION AND MODEL BUILDING
plot(Matings ~ Age, log="y")
ageSquared <- Age^2
myGlm1 <- glm(Matings ~ Age + ageSquared, family=poisson)
summary(myGlm1)  # No evidence of a need for ageSquared

## INFERENCE AND INTERPRETATION
myGlm2 <- update(myGlm1, ~ . - ageSquared)
summary(myGlm2)
beta <- myGlm2$coef
exp(beta[2])  # 1.071107
exp(confint(myGlm2,2))  # 1.042558 1.100360

# Interpretation: Associated with each 1 year increase in age is a 7% increase in the mean number of matings (95% confidence interval 4% to 10% increase).

## GRAPHICAL DISPLAY FOR PRESENTATION
plot(Matings ~ Age, ylab="Number of Successful Matings", xlab="Age of Male Elephant (Years)", main="Age and Number of Successful Matings for 41 African Elephants", pch=21, bg="green", cex=2, lwd=2)
dummyAge <- seq(min(Age),max(Age), length=50)
lp <- beta[1] + beta[2]*dummyAge
```
The Del Norte Salamander (*plethodon elongates*) is a small (5–7 cm) salamander found among rock rubble, rock outcrops and moss-covered talus in a narrow range of northwest California. To study the habitat characteristics of the species and particularly the tendency of these salamanders to reside in dwindling old-growth forests, researchers selected 47 sites from plausible salamander habitat in national forest and parkland. Randomly chosen grid points were searched for the presence of a site with suitable rocky habitat. At each suitable site, a 7 metre by 7 metre search area was examined for the number of salamanders it contained. This data frame contains the counts of salamanders at the sites, along with the percentage of forest canopy and age of the forest in years.

### Usage

```r
case2202
```

### Format

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

- **Site**: Investigated site
- **Salamanders**: Number of salamanders found in 49 m$^2$ area
- **PctCover**: Percentage of canopy cover
- **ForestAge**: Forest age

### Source


### References


### Examples

```r
str(case2202)
attach(case2202)

## EXPLORATION AND MODEL BUILDING
logSalamanders <- log(Salamanders + .5)
logForestAge <- log(ForestAge + .5)
myMatrix <- cbind(PctCover,logForestAge,logSalamanders)
if (require(car)) { # Use car library
  scatterplotMatrix(myMatrix, diagonal="histogram", reg.line=FALSE, spread=FALSE)
```
myGlm1 <- glm(Salamanders ~ PctCover + logForestAge + PctCover:logForestAge, family=poisson)
summary(myGlm1)  # Backward elimination...
myGlm2 <- update(myGlm1, ~ . - PctCover:logForestAge)
summary(myGlm2)
myGlm3 <- update(myGlm2, ~ . - logForestAge)
summary(myGlm3)  # PctCover is the only explanatory variable remaining
plot(Salamanders ~ PctCover)  # It appears that there are 2 distributions
# of Salamander counts; one for PctCover < 70 and one for PctCover > 70

# See if PctCover is associated Salamanders in each subset
myGlm4 <- glm(Salamanders ~ PctCover, family=poisson,subset=(PctCover > 70))
summary(myGlm4)  # No evidence of an effect for this subset
myGlm5 <- glm(Salamanders ~ PctCover, family=poisson,subset=(PctCover < 70))
summary(myGlm5)  # No evidence on this subset either

## INFERENCE (2 means)
Group <- ifelse(PctCover > 70,"High","Low")
Group <- factor(Group, levels=c("Low","High") )  # Make "Low Cover" the ref group
myGlm6 <- glm(Salamanders ~ Group, family=poisson)
summary(myGlm6)

## GRAPHICAL DISPLAY FOR PRESENTATION
plot(Salamanders ~ PctCover, ylab="Number of Salamanders",
     xlab="Percentage of Canopy Covered",
     main="Number of Salamanders versus Percent Canopy Cover",
     pch=21, bg="green", cex=2, lwd=2)
beta <- myGlm6$coef
lines(c(0,55),exp(c(beta[1],beta[1])),lwd=2)
text(56,exp(beta[1]),paste("mean= ",round(exp(beta[1]),1)),adj=0)
lines(c(76,93),exp(c(beta[1]+beta[2],beta[1]+beta[2])),lwd=2)
detach(case22/zero.noslash2)

---

**Fish Oil and Blood Pressure**

**Description**

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

**Usage**

ex0112
### Format

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

- **BP**: reduction in diastolic blood pressure (in mm of mercury)
- **Diet**: factor variable indicating the diet that the subject followed, with levels "FishOil" and "RegularOil"

### Source


### References


### Examples

```r
str(ex0112)
```

---

### Description

The data are the gross domestic product per capita for 228 countries in 2010.

### Usage

```r
ex0116
```

### Format

A data frame with 228 observations on the following 3 variables.

- **Rank**: rank order of country from highest to lowest GDP
- **Country**: name of country
- **PerCapitaGDP**: per capita GDP in $US

### Source


### References


### Examples

```r
str(ex0116)
```
Zinc concentrations for two groups of rats

Description
The data are the zinc concentrations (in mg/ml) in the blood of rats that received a dietary supplement and rats that did not receive the supplement.

Usage

ex0125

Format
A data frame with 39 observations on the following 2 variables.

Group  a factor representing the group, with levels "A" for the dietary supplement group and "B" for the control group
Zinc  measured zinc concentration in mg/ml

Source

Examples

str(ex0125)

Environmental Voting of Democrats and Republicans in the U.S. House of Representatives

Description
The data are the number of pro- and anti-environmental votes, according to the League of Conservation Voters, for each member of the U.S. House of Representatives in 2005, 2006, or 2007.

Usage

ex0126
Format

A data frame with 492 observations on the following 10 variables.

State  the state that the member represented
Representative  name of the representative
Party  a factor representing political party, with levels "R" for Republican, "D" for Democratic, and "I" for Independent
Pro05  the number of pro-environmental votes in 2005
Anti05  the number of anti-environmental votes in 2005
Pro06  the number of pro-environmental votes in 2006
Anti06  the number of anti-environmental votes in 2006
Pro07  the number of pro-environmental votes in 2007
Anti07  the number of anti-environmental votes in 2007
PctPro  the total percentage of a representative’s votes between 2005 and 2007 that were deemed to be pro-environmental

Source


See Also

ex0127

Examples

str(ex0126)

---

**Environmental Voting of Democrats and Republicans in the U.S. Senate**

Description

The data are the number of pro- and anti-environmental votes, according to the League of Conservation Voters, for each member of the U.S. Senate in 2005, 2006, or 2007.

Usage

ex0127
Format

A data frame with 112 observations on the following 10 variables.

**State**  the state that the member represented

**Senator**  name of the senator

**Party**  a factor representing political party, with levels "R" for Republican, "D" for Democratic, and "I" for Independent

**Pro2005**  the number of pro-environmental votes in 2005

**Anti2005**  the number of anti-environmental votes in 2005

**Pro2006**  the number of pro-environmental votes in 2006

**Anti2006**  the number of anti-environmental votes in 2006

**Pro2007**  the number of pro-environmental votes in 2007

**Anti2007**  the number of anti-environmental votes in 2007

**PctPro**  the total percentage of a representative’s votes between 2005 and 2007 that were deemed to be pro-environmental

Source


See Also

`ex0126`

Examples

```r
str(ex0127)
```

---

**Lifetimes of Guinea Pigs**

Description

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

Usage

`ex0211`

Format

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

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

**Group**  a factor with levels "Bacilli" and "Control", indicating the group to which the guinea pig was assigned
Source


References


Examples

```r
str(ex/zero.noslash211)
```

---

**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 751 finches caught the year before the drought (1976) and 89 finches captured the year after the drought (1978).

Usage

```r
ex/zero.noslash218
```

Format

A data frame with 840 observations on the following 2 variables.

- **Year** Year the finch was caught, 1976 or 1978
- **Depth** Beak depth of the finch (mm)

Source


References


See Also

```r
case/zero.noslash2/zero.noslash1
```
As evidence in support of natural selection, Bumpus presented measurements on house sparrows brought to the Anatomical Laboratory of Brown University after an uncommonly severe winter storm. Some of these birds had survived and some had perished. Bumpus asked whether those that perished did so because they lacked physical characteristics enabling them to withstand the intensity of that particular instance of selective elimination. The data are on the weights, in grams, for the 24 adult male sparrows that perished and for the 35 adult males that survived.

Usage

ex0221

Format

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

- **Humerus**: humerus length of adult male sparrows (inches)
- **Status**: factor variable indicating whether the sparrow perished or survived in a winter storm, with levels `Perished` and `Survived`

Source


See Also

ex0216

Examples

str(ex0221)
Description

These data are armed Forces Qualifying Test (AFQT) score percentiles and component test scores in arithmetic reasoning, word knowledge, paragraph comprehension, and mathematical knowledge for a sample of 1,278 U.S. women and 1,306 U.S. men in 1981.

Usage

ex0222

Format

A data frame with 2,584 observations on the following 6 variables.

Gender  a factor with levels "female" and "male"
Arith  score on the arithmetic reasoning component of the AFQT test
Word  score on the word knowledge component
Parag  score on the paragraph comprehension component
Math  score on the mathematical knowledge component
AFQT  percentile score on the AFQT test

Source


References


See Also

ex0330, ex0331, ex0524, ex0525, ex0828, ex0923, ex1033, ex1223

Examples

str(ex0222)
**Description**

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

**Usage**

ex0223

**Format**

A data frame with 51 observations on the following 5 variables.

- **State**: US state
- **Fatalities1995**: number of traffic fatalities in 1995
- **Fatalities1996**: number of traffic fatalities in 1996
- **PctChange**: percentage change in interstate traffic fatalities between 1995 and 1996
- **SpeedLimit**: a factor with levels "Inc" and "Ret", indicating whether the state increased or retained its speed limit

**Source**


**References**


**Examples**

`str(ex0223)`
Description
Researchers collected historical and current data on umpires to investigate their life expectancies following the collapse and death of a U.S. major league baseball umpire. They were investigating speculation that stress associated with the job posed a health risk. Data were found on 227 umpires who had died or had retired and were still living. The data set includes the dates of birth and death.

Usage
ex0321

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

- **Lifelength** observed lifetime for those umpires who had died by the time of the study or current age of those still living
- **Censored** 0 for those who had died by the time of the study or 1 for those who were still living
- **Expected** length from actuarial life tables for individuals who were alive at the time the person first became an umpire

Source

References

Examples
str(ex0321)

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

Usage
ex0323
Format

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

- **Year**  year
- **CancerRate** skin cancer rate per 100,000 people
- **SunspotActivity** a factor with levels "High" and "Low"

Source


References


Examples

```r
str(ex/zero.noslash323)
```

---

**ex0327**  
**Pollen Removal**

Description

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

Usage

```r
ex0327
```

Format

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

- **PollenRemoved** proportion of pollen removed
- **DurationOfVisit** duration of visit (in seconds)
- **BeeType** factor variable with levels "Queen" and "Worker"

Source


References

Examples

str(ex0327)

-------
ex0330  Education and Income
-------

Description

The data are incomes in U.S. dollars for 1,020 working Americans who had 12 years of education and 406 working Americans who had 16 years of education, in 2005.

Usage

ex0330

Format

A data frame with 1,426 observations on the following 3 variables.

Subject  a subject identification number
Educ    number of years of education—either 12 or 16
Income2005  income, in dollars

Source


References


See Also

ex0222, ex0331, ex0524, ex0525, ex0828, ex0923, ex1033, ex1223

Examples

str(ex0330)
**Description**

The data are incomes in U.S. dollars for 406 working Americans who had 16 years of education and 374 working Americans who had more than 16 years of education, in 2005.

**Usage**

ex0331

**Format**

A data frame with 780 observations on the following 3 variables.

- **Subject**  a subject identification number
- **Educ** factor with levels "16" and ">16"
- **Income2005** income, in dollars

**Source**


**References**


**See Also**

ex0222, ex0330, ex0524, ex0525, ex0828, ex0923, ex1033, ex1223

**Examples**

```r
str(ex0331)
```

---

**Description**


**Usage**

ex0332
Format
A data frame with 50 observations on the following 4 variables.

- **College** name of the college
- **Type** a factor with levels "Private" and "Public"
- **InState** in-state tuition in dollars
- **OutOfState** out-of-state tuition in dollars

Source

References

Examples
```r
str(ex0332)
```

---

**ex0333**

*Brain Size and Litter Size*

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

Usage
```r
ex0333
```

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

- **BrainSize** relative brain sizes (1000 * Brain weight/Body weight) for 96 species of mammals
- **LitterSize** factor variable with levels "Small" and "Large"

Source

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

### Usage
`ex0428`

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

- **Cross**  height (inches) of cross-fertilized plant
- **Self**  height (inches) of self-fertilized plant

### Source

### References

### Examples
`str(ex0428)`
Salvage Logging

Description
The data are the number of tree seedlings per transect in nine logged (L) and seven unlogged (U) plots affected by the Oregon Biscuit Fire, in 2004 and 2005, and the percentage of seedlings lost between 2004 and 2005. The goal is to see whether the distribution of seedlings lost differs in logged and unlogged plots.

Usage
ex0429

Format
A data frame with 16 observations on the following 5 variables.

- **Plot**: an identification code for plot
- **Action**: a factor with levels "L" for logged and "U" for unlogged
- **Seedlings2004**: the number of seedlings in the plot in 2004
- **Seedlings2005**: the number of seedlings in the plot in 2005
- **PercentLost**: the percentage of 2004 seedlings that were lost

Source

References

Examples
str(ex0429)

Sunlight Protection Factor

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

Usage
ex0430
Format
A data frame with 13 observations on the following 2 variables.

PreTreatment  tolerance to sunlight (minutes) prior to sunscreen application
Sunscreen  tolerance to sunlight (minutes) after sunscreen application

Source

References

Examples
```r
str(ex/zero.noslash431)
```

---

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

Usage
```r
ex0431
```

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

Survival  months of survival after beginning of study
Group a factor with levels "Control" and "Therapy"
Censor 0 if entire lifetime observed, 1 if patient known to have lived at least 122 months

Source

References

Examples
```r
str(ex0431)
```
Description
To investigate the capacity of marijuana to reduce the side effects of cancer chemotherapy, researchers performed a double-blind, randomized, crossover trial. Fifteen cancer patients on chemotherapy were randomly assigned to receive either a marijuana treatment or a placebo treatment after their first three sessions of chemotherapy. They were then crossed over to the opposite treatment for their next 3 sessions.

Usage
ex0432

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

Subject  subject number 1–15
Marijuana total number of vomiting and retching episodes under marijuana treatment
Placebo total number of vomiting and retching episodes under placebo treatment

Source

References

Examples
str(ex0432)

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

Usage
ex0518
Format
A data frame with 30 observations on the following 4 variables.

- **Protein**: levels of protein (x 10) found in rat livers
- **Treatment**: a factor with levels "Control", "CPFA50", "CPFA150", "CPFA300", "CPFA450" and "CPFA600"
- **Day**: a factor with levels "Day1", "Day2", "Day3", "Day4" and "Day5"
- **TrtDayGroup**: a factor with levels "Group1", "Group2", ..., "Group10"; the observed levels of the Treatment and Day interaction

Source

Examples
str(ex0518)

---

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

**Usage**
ex0523

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

- **Oxygen**: oxygen isotopic composition
- **Bone**: a factor with levels "Bone1", "Bone2", ..., "Bone12"

Source

References

See Also
ex1120
**Description**

These data are annual incomes in 2005 for 2,584 Americans who were selected in the National Longitudinal Study of Youth 1979, who were available for re-interview in 2006, and who had paying jobs in 2005, along with the quartile of their AFQT (IQ) test score taken in 1981. How strong is the evidence that the distributions of 2005 annual incomes differ in the four populations? By how many dollars or by what percent does the distribution of 2005 incomes for those within the highest (fourth) quartile of IQ test scores exceed the distribution for the lowest (first) quartile?

**Usage**

ex0524

**Format**

A data frame with 2,584 observations on the following 3 variables.

- **Subject**  subject identification number
- **IQquartile**  a factor with levels "1stQuartile","2ndQuartile","3rdQuartile" and "4thQuartile"
- **Income2005**  annual income in U.S. dollars, 2005

**Source**


**References**


**See Also**

ex0222, ex0330, ex0331, ex0525, ex0828, ex0923, ex1033, ex1223

**Examples**

str(ex0524)
Description

These data are annual incomes in 2005 of a random sample of 2,584 Americans who were selected for the National Longitudinal Survey of Youth in 1979 and who had paying jobs in 2005. The data set also includes a code for the number of years of education that each individual had completed by 2006: <12, 12, 13–15, 16, and >16. How strong is the evidence that at least one of the five population distributions (corresponding to the different years of education) is different from the others? By how many dollars or by what percent does the mean or median for each of the last four categories exceed that of the next lowest category?

Usage

ex0525

Format

A data frame with 2,584 observations on the following 3 variables.

- **Subject**: subject identification number
- **Educ**: a factor for years of education category with levels "<12", "12", "13–15", "16" and ">16"
- **Income2005**: Annual income in 2005, in U.S. dollars

Source


References


See Also

- ex0222, ex0330, ex0331, ex0524, ex0828, ex0923, ex1033, ex123

Examples

```r
str(ex0525)
```
**Diet Wars**

**Description**

These data are simulated to match the summary and conclusions of a real study of overweight employees who were randomly assigned to three diet groups: a low-fat diet, a low-carb diet (similar to the Atkins diet), and a Mediterranean diet. The study ran for two years, with 272 employees completing the entire protocol. Is there evidence of differences in average weight loss among these diets? If so, which diets appear to be better than which others?

**Usage**

`ex0623`

**Format**

A data frame with 272 observations on the following 3 variables.

- **Subject**: subject identification number
- **Group**: a factor with levels "Low-Carbohydrate", "Low-Fat", and "Mediterranean"
- **WtLoss24**: weight at the end of the 24 month study minus initial weight, in kg

**Source**


**See Also**

`ex1420, ex1921, ex1922`

**Examples**

`str(ex0623)`

---

**A Biological Basis for Homosexuality**

**Description**

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

**Usage**

`ex0624`
Format

A data frame with 41 observations on the following 5 variables.

**Volume** volumes of INAH3 (1000 × mm$^3$) cell clusters from 41 humans

**Group** a factor with levels

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

**Sex** a factor with levels "Female" and "Male"

**Orientation** a factor with levels "Heterosexual" and "Homosexual"

**Death** a factor with levels "AIDS" and "Non-AIDS"

Source


References


Examples

```r
str(ex0624)
```

---

**Planetary Distances and Order from the Sun**

Description

The first three columns are the names, orders of distance from the sun and distances from the sun (scaled so that earth is 1) of the 8 planets in our solar system and the dwarf planet, Pluto. The next three columns are the same, but also include the asteroid belt.

Usage

```r
ex0721
```

Format

A data frame with observations on the following 6 variables.

**Name** name of object in solar system, 9 objects

**Order** order of object’s distance from the sun

**Distance** distance of object from sun, with earth = 1
Name: name of object in solar system, including asteroid belt
Order: order of object’s distance from the sun
Distance: distance of object from sun, with earth = 1

Source

See Also
ex2226

Examples
str(ex/zero.noslash721)

---

### Crab Claw Size and Force

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

**Usage**

ex0722

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

- **Force**: closing strength of claw of the crab
- **Height**: propodus height of claw of the crab
- **Species**: species to which the crab belongs

**Source**

**References**

**Examples**
str(ex0722)
### Decline in Male Births

**Description**

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

**Usage**

`ex0724`

**Format**

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

- **Year**: year of observation
- **Denmark**: male birth rate of Denmark for given year
- **Netherlands**: male birth rate of The Netherlands for given year
- **Canada**: male birth rate of Canada for given year
- **USA**: male birth rate of the United States for given year

**Source**


**References**


**Examples**

```r
str(ex0724)
```

### The Big Bang II

**Description**

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

**Usage**

`ex0725`
Format

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

**Distance** distance from earth (in million parsec)

**Velocity** recession velocity (in kilometres per second)

Source


References


See Also

`case/zero.noslash7/zero.noslash1`

Examples

```r
str(ex/zero.noslash725)
```

---

### Orign of the Term Regression

Description

These data are heights of 933 adults and their parents, as measured by Karl Pearson in 1885.

Usage

`ex0726`

Format

A data frame with 933 observations on the following 5 variables.

**Gender** a factor with levels "female" and "male"

**Family** an identification number for family, 1, 2, . . . , 205

**Height** adult height of the child, inches

**Father** height of the child’s father, inches

**Mother** height of the child’s mother, inches

Source

References


Examples

```r
str(ex0725)
```

---

### Male Displays

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

#### Usage

```r
ex0727
```

#### Format

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

- **Mass**  average mass of stones carried by bird (in g)
- **Tcell**  T-cell response measurement (in mm)

#### Source


#### References


#### Examples

```r
str(ex0727)
```
**Brain Activity in Violin and String Players**

**Description**

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

**Usage**

ex0728

**Format**

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

- **Years** years that the individual has been playing
- **Activity** neuronal activity index

**Source**


**References**


**Examples**

str(ex0728)

**Sampling Bias in Exit Polls**

**Description**

These data are the number of percentage points by which exit polls over estimated the actual vote for candidate John Kerry in the 2004 U.S. presidential election, grouped according to the distance of the exit poll interviewer from the door of the polling location. How strong is the evidence that the mean Kerry overestimate increases with increasing distance of interviewer from the door (thus lending evidence to the theory that supporters of the other candidate, George W Bush, were more inclined to avoid exit pollsters)?
Usage

ex0729

Format

A data frame with 6 observations on the following 2 variables.

**OverEstimate** number of percentage points by which the exit poll estimate exceeded the actual percentage voting for Kerry (in all precincts with a similar distance of interviewer from the door)

**Distance** distance of the interviewer from the door of the polling location, in feet

Source


References


See Also

ex0730

Examples

str(ex0729)

---

**Sampling Bias in Exit Polls 2**

Description

These data are the average proportion of voters refusing to be interviewed by exit pollsters in the 2004 U.S. presidential election, grouped by age of the interviewer, and the approximate age of the interviewer. What evidence do the data provide that the mean refusal rate decreased with increasing age of interviewer?

Usage

ex0730

Format

A data frame with 6 observations on the following 2 variables.

**Age** age of the exit poll interviewer, years

**Refusal** average proportion of voters refusing to be interviewed
Source


References


See Also

ex0729

Examples

```r
str(ex0730)
```

---

**Meat Processing**

Description

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

Usage

```r
ex0816
```

Format

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

- **Time**  time after slaughter (hours)
- **pH**  pH level in postmortem muscle

Source


References


See Also

`case0702`

Examples

```r
str(ex0816)
```
Description

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

Usage

ex0817

Format

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

Load  flea beetle load (in beetles/gram of ragwort dry mass)
Mass  dry mass of ragwort weed

Source


References

Format

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

**Year**  Year of election

**District**  a factor with levels "D1", "D2", "D3", "D4", "D5", "D7", and "D8"

**DemAbsenteeVotes**  Number of absentee ballots indicating a vote for the Democratic candidate

**RepubAbsenteeVotes**  Number of absentee ballots indicating a vote for the Republican candidate

**DemMachineVotes**  Number of machine-counted ballots indicating a vote for the Democratic candidate

**RepubMachineVotes**  Number of machine-counted ballots indicating a vote for the Republican candidate

**DemPctOfAbsenteeVotes**  Percentage of absentee ballots indicating a vote for the Democratic candidate

**DemPctOfMachineVotes**  Percentage of machine-counted ballots indicating a vote for the Democratic candidate

**Disputed**  a factor taking on the value "yes" for the disputed election and "no" for all other elections

Details

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

Source


References


Examples

```r
str(ex0820)
```
Ecosystem Decay

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

Usage
ex0822

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

Area area (ha) of forest patch
Species number of butterfly species

Source

References

Examples
str(ex0822)

Wine Consumption and Heart Disease

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

Usage
data(ex0823)
Format

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

Country a character vector indicating the country
Wine consumption of wine (liters per person per year)
Mortality heart disease mortality rate (deaths per 1,000)

Source


References


Examples

str(ex0823)

---

Description

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

Usage

ex0824

Format

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

Age age in months of child
Rate respiratory rate (breaths per minute)

Source


References

**Examples**

```r
str(ex0824)
```

---

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

**Description**

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

**Usage**

```r
ex0825
```

**Format**

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

- **County** a character vector indicating the county
- **Buchanan2000** votes cast for P. Buchanan
- **Bush2000** votes cast for G.W. Bush

**Source**


**See Also**

```r
ex1222
```

**Examples**

```r
str(ex0825)
```

---

**Kleiber’s Law**

**Description**

The data are the average mass, metabolic rate, and lifespan for 95 species of mammals. Kleiber’s law states that the metabolic rate of an animal species, on average, is proportional to its mass raised to the power of 3/4.

**Usage**

```r
ex0826
```
Format

A data frame with 95 observations on the following 5 variables.

- **CommonName** the common name of the mammal species
- **Species** the scientific name of the mammal species
- **Mass** the average body mass in kg
- **Metab** the average metabolic rate in kJ per day
- **Life** the average lifespan in years

Source


Examples

```r
str(ex0826)
```

---

**IQ, Education, and Future Income**

Description

These data are Armed Forces Qualifying Test (AFQT) score percentiles, years of education, and annual income in 2005 for a subset of a random sample of 2,584 Americans selected in 1979 who were working in 2005 and re-interviewed in 2006.

Usage

```r
ex0828
```

Format

A data frame with 2,584 observations on the following 4 variables.

- **Subject** the subject identification number
- **AFQT** percentile score on the AFQT test
- **Educ** years of education achieved by 2005
- **Income2005** annual income in 2005

Source


References

See Also

ex0222, ex0330, ex0331, ex0524, ex0525, ex0923, ex1033, ex1223

Examples

str(ex0828)

<table>
<thead>
<tr>
<th>ex0829</th>
<th>Autism Rates</th>
</tr>
</thead>
</table>

Description


Usage

ex0829

Format

A data frame with 5 observations on the following 2 variables.

**Year** year

**Prevalence** the number of autism cases per 10,000 ten-year old children

Source


References


Examples

str(ex0829)
Pace of Life and Heart Disease

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

Usage
ex0914

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

- Bank  bank clerk speed
- Walk  pedestrian walking speed
- Talk  postal clerk talking speed
- Heart age adjusted death rate due to heart disease

Source

References

Examples
str(ex0914)

Rainfall and Corn Yield

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

Usage
ex0915
Format

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

- **Year**  year of observation (1890–1927)
- **Yield** average corn yield for the six states (in bu/acre)
- **Rainfall** average rainfall in the six states (in in/year)

Source


References


Examples

```r
str(ex0915)
```

---

**Speed of Evolution**

Description

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

Usage

```r
ex0918
```

Format

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

- **Continent**  a factor with levels "NA" and "EU"
- **Latitude**  latitude (degrees)
- **Females**  average wing size ($10^3 \times \log \text{mm}$) of female flies on log scale
- **SE_Females**  standard error of wing size ($10^3 \times \log \text{mm}$) of female flies on log scale
- **Males**  average wing size ($10^3 \times \log \text{mm}$) of male flies on log scale
- **SE_Males**  standard error of wing size ($10^3 \times \log \text{mm}$) of male flies on log scale
- **Ratio**  average basal length to wing size ratios of female flies
- **SE_Ratio**  standard error of average basal length to wing size ratio of female flies

Source

References


Examples

str(ex0918)

---

**ex0920**  
*Winning Speeds at the Kentucky Derby*

**Description**

Data set contains the year of the Kentucky Derby, the winning horse, the condition of the track and the average speed of the winner for years 1896–2011.

**Usage**

ex0920

**Format**

A data frame with 116 observations on the following 8 variables.

- **Year**  year of Kentucky Derby
- **Winner**  a character vector with the name of the winning horse
- **Starters**  number of horses that started the race
- **NetToWinner**  the net winnings of the winner, in U.S. dollars
- **Time**  the winning time in seconds
- **Speed**  the winning average speed, n miles per hour
- **Track**  a factor indicating track condition with levels "Fast", "Good", "Dusty", "Slow", "Heavy", "Muddy", and "Sloppy"
- **Conditions**  a factor with with 2 levels of track condition, with levels "Fast" and "Slow"

**Source**


**References**


**Examples**

str(ex0920)
Ingestion Rates of Deposit Feeders

Description

The data are the typical dry weight in mg, the typical ingestion rate (weight of food intake per day for one animal) in mg/day, and the percentage of the food that is composed of organic matter for 19 species of deposit feeders. The goal is to see whether the distribution of species’ ingestion rates depends on the percentage of organic matter in the food, after accounting for the effect of species weight and to describe the association.

Usage

ex0922

Format

A data frame with 19 observations on the following 4 variables.

Species  a character variable with the name of the species
Weight   the dry weight of the species, in mg
Ingestion ingestion rate in mg per day
Organic  percentage of organic matter in the food

Source


References


See Also

ex1125

Examples

str(ex0921)
Comparing Male and Female Incomes, Accounting for Education and IQ

Description

These data are a subset of the National Longitudinal Study of Youth data, with annual incomes in 2005, intelligence test scores (AFQT) measured in 1981, and years of education completed by 2006 for 1,306 males and 1,278 females who were between the ages of 14 and 22 when selected for the survey in 1979, who were available for re-interview in 2006, and who had paying jobs in 2005. Is there any evidence that the mean salary for males exceeds the mean salary for females with the same years of education and AFQT scores? By how many dollars or by what percent is the male mean larger?

Usage

ex0923

Format

A data frame with 2,584 observations on the following 5 variables.

Subject the subject identification number
Gender a factor with levels "female" and "male"
AFQT percentile score on the AFQT intelligence test
Educ years of education achieved by 2005
Income2005 annual income in 2005

Source


References


See Also

ex0222, ex0330, ex0331, ex0524, ex0525, ex0828, ex1033, ex1223

Examples

str(ex0923)
Toxic Effects of Copper and Zinc

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

Usage
ex1014

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

- **Copper** amount of copper received (in ppm)
- **Zinc** amount of zinc received (in ppm)
- **Protein** protein in minnow larvae exposed to copper and zinc (µg/larva)

Source

References

Examples
str(ex1014)

Thinning of Ozone Layer

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

Usage
ex1026
Format

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

- **Inhibit** percent inhibition of primary phytoplankton production in water
- **UVB** UVB exposure
- **Surface** a factor with levels "Deep" and "Surface"

Source


References


Examples

```r
str(ex1026)
```

---

Factors Affecting Extinction

Description

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

Usage

```r
ex1027
```

Format

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

- **Species** a character vector indicating the species
- **Time** average extinction time in years
- **Pairs** average number of nesting pairs
- **Size** a factor with levels "L" and "S"
- **Status** a factor with levels "M" and "R"

Source

References


Examples

```r
str(ex1027)
```

---

**El Nino and Hurricanes**

Description

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

Usage

```r
ex1028
```

Format

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

- **Year** year
- **ElNino** a factor with levels "cold", "neutral" and "warm"
- **Temperature** numeric variable with values -1 if ElNino is "cold", 0 if "neutral" and 1 if "warm"
- **WestAfrica** numeric variable indicating whether West Africa was wet (1) or dry (0)
- **Storms** number of storms
- **Hurricanes** number of hurricanes
- **StormIndex** index of overall intensity of hurricane season

Source


References

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

Examples

```r
str(ex1028)
```
Description

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

Usage

ex1029

Format

A data frame with 25,437 observations on the following 6 variables.

- **Region**: a factor with levels "Midwest", "Northeast", "South" and "West"
- **MetropolitanStatus**: a factor with levels "MetropolitanArea" and "NotMetropolitanArea"
- **Exper**: experience in years
- **Educ**: education in years
- **Race**: a factor with levels "Black" and "NotBlack"
- **WeeklyEarnings**: weekly wage in dollars

Source


References


Examples

```r
str(ex1029)
```
Description
A data set with weekly earnings for 4,952 males between the age of 18 and 70 sampled in the March 2011 Current Population Survey (CPS). These males are a subset who had reported earnings and who responded as having race as either “Only White” or “Only Black.” Also recorded are the region of the country (with four categories: Northeast, Midwest, South, and West), the metropolitan status of the men’s employment (with three categories: Metropolitan, Not Metropolitan, and Not Identified), age, education category (with 16 categories ranging from “Less than first grade” to “doctorate Degree”), and education code, which is a numerical value that corresponds roughly to increasing levels of education (and so may be useful for plotting). What evidence do the data provide that the distributions of weekly earnings differ in the populations of white and black workers after accounting for the other variables? By how many dollars or by what percent does the White population mean (or median) exceed the Black population mean (or median)?

Usage

ex1030

Format
A data frame with 4,952 observations on the following 7 variables.

Region a factor with levels "Midwest", "Northeast", "South" and "West"
MetropolitanStatus a factor with levels "Metropolitan", "Not Metropolitan " and "Not Identified"
Age age in years
EducationCategory a factor with 16 levels: "SomeCollegeButNoDegree", "AssocDegAcadem", "NinthGrade", "BachelorsDegree", "TenthGrade", "HighSchoolDiploma", "AssocDegOccupVocat", "DoctorateDegree", "TwelfthButNoDiploma", "LessThanFirstGrade", "EleventhGrade", "ProfSchoolDegree", "FifthorSixthGrade", "SeventhOrEighthGrade", "FirstSecondThirdOrFourthGrade"
EducationCode a numerical variable indicating the approximate ordering of EducationCategory, with higher numbers indicating more education
Race a factor with levels "Black" and "White"
WeeklyEarnings weekly wage in dollars

Source

References

Examples

str(ex1030)
Description

A data set with Clutch Volume (cubic millimeters) and adult Body Mass (kg) in six different groups of animals: modern maternal-care bird species (Mat), modern paternal-care bird species (Pat), modern biparental-care bird species (BiP), modern maternal-care crocodiles (Croc), non-avian maniraptoran dinosaurs thought to be ancestors of modern birds (Mani), and other non-avian dinosaurs (Othr). The question of interest was which group of modern creatures most closely matches the relationship in the maniraptoran dinosaurs.

Usage

ex1031

Format

A data frame with 443 observations on the following 6 variables.

- **CommonName**  the common name of the species
- **Genus** species genus
- **Species** species name
- **Group** a factor with 6 levels corresponding to the 6 groups of animals: "BiP", "Croc", "Mani", "Mat", "Othr", and "Pat"
- **BodyMass** the average body mass of individuals in the species (kg)
- **ClutchVolume** the total volume of all eggs in a clutch (average value for the species)

Source


References


See Also

ex1923

Examples

str(ex1031)
IQ Score and Income

Description

This is a subset of the National Longitudinal Study of Youth (NLSY79) data, with annual incomes in 2005 (in U.S. dollars, as Recorded in a 2006 interview); scores on the Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, and Mathematics Knowledge portions of the Armed Forces Vocational Aptitude Battery (ASVAB) of tests taken in 1981; and the percentile score of the Armed Forces Qualifying Test (AFQT), which is a linear combination of the four component tests mentioned above (but note that AFQT reported here is the percentile, which is not a linear combination of the four component scores). Which of the five test scores seem to be the most important predictors of 2005 income? Is the AFQT sufficient by itself?

Usage

ex1033

Format

A data frame with 2,584 observations on the following 7 variables.

Subject  the subject identification number
Arith  score on the Arithmetic Reasoning test in 1981
Word  score on the Word Knowledge Test in 1981
Parag  score on the Paragraph Comprehension test in 1981
Math  score on the Mathematics Knowledge test in 1981
AFQT  percentile score on the AFQT intelligence test in 1981
Income2005  annual income in 2005

Source


References


See Also

ex0222, ex0330, ex0331, ex0524, ex0525, ex0828, ex0923, ex1223

Examples

str(ex1033)
**Chernobyl Fallout**

**Description**

The data are the cesium concentrations (in Bq/kg) in soil and in mushrooms at 17 wooded locations in Umbria, Central Italy, from August 1986 to November 1989. Researchers wished to investigate the cesium transfer from contaminated soil to plants after the Chernobyl nuclear power plant accident in April 1986 by describing the distribution of the mushroom concentration as a function of soil concentration.

**Usage**

```r
ex1111
```

**Format**

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

- **Mushroom** the cesium concentration in mushrooms, Bq/kg
- **Soil** the cesium concentration in soil, Bq/kg

**Source**


**Examples**

```r
str(ex1111)
```

---

**Was Tyrannosaurus Rex Warm-Blooded?**

**Description**

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

**Usage**

```r
ex1120
```

**Format**

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

- **Carbonate** isotopic composition of bone carbonate
- **Calcite** isotopic composition of calcite cements
Source

References

See Also
ex0523

Examples
str(ex1122)

---

**Deforestation and Debt**

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

**Usage**
ex1122

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

- **Country**: a character vector indicating the country
- **Debt**: debt (millions of dollars)
- **Deforest**: deforestation (thousands of ha)
- **Pop**: population (thousands of people)

**Source**

**References**

**Examples**
str(ex1122)
**Description**


**Usage**

ex1123

**Format**

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

- **City**  a character vector indicating the city
- **Mort**  total age-adjusted mortality from all causes
- **Precip**  mean annual precipitation (inches)
- **Educ**  median number of school years completed for persons 25 years or older
- **NonWhite**  percentage of population that is nonwhite
- **NOX**  relative pollution potential of oxides of nitrogen
- **SO2**  relative pollution potential of sulfur dioxide

**Source**


**References**


**See Also**

ex1217

**Examples**

str(ex1123)
**Natal Dispersal Distances of Mammals**

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

**Usage**
```r
ex1124
```

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

- **Species** a character vector indicating the species
- **BodyMass** bodymass (kg)
- **MaxDist** maximum dispersal distance (km)
- **Type** a factor with levels "C", "H" and "O" indicating carnivore, herbivore, or omnivore

**Source**

**References**

**Examples**
```r
str(ex1124)
```

**Ingestion Rates of Deposit Feeders**

**Description**
The data are the typical dry weight in mg, the typical ingestion rate (weight of food intake per day for one animal) in mg/day, and the percentage of the food that is composed of organic matter for 22 species of deposit feeders. The goal is to see whether the distribution of species’ ingestion rates depends on the percentage of organic matter in the food, after accounting for the effect of species weight and to describe the association. The last three species happen to be Bivalves, so may behave differently from the others.
Usage

ex1125

Format

A data frame with 22 observations on the following 5 variables.

Species a character variable with the name of the species
Weight the dry weight of the species, in mg
Ingestion ingestion rate in mg per day
Organic percentage of organic matter in the food
Bivalve a factor with levels "no" and "yes" to indicate whether a species is a bivalve

Source


References


See Also

ex0921

Examples

str(ex1125)

---

ex1217  
*Pollution and Mortality*

Description

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

Usage

ex1217
Format

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

- **CITY**: a character vector indicating the city
- **Mortality**: total age-adjusted mortality from all causes
- **Precip**: mean annual precipitation (inches)
- **Humidity**: percent relative humidity (annual average at 1:00pm)
- **JanTemp**: mean January temperature (degrees F)
- **JulyTemp**: mean July temperature (degrees F)
- **Over65**: percentage of the population aged 65 years or over
- **House**: population per household
- **Educ**: median number of school years completed for persons 25 years or older
- **Sound**: percentage of the housing that is sound with all facilities
- **Density**: population density (in persons per square mile of urbanized area)
- **NonWhite**: percentage of population that is nonwhite
- **WhiteCol**: percentage of employment in white collar occupations
- **Poor**: percentage of households with annual income under $3,000 in 1960
- **HC**: relative pollution potential of hydrocarbons
- **NOX**: relative pollution potential of oxides of nitrogen
- **SO2**: relative pollution potential of sulfur dioxide

Source


References


See Also

- ex1123

Examples

```r
str(ex1217)
```
Description

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

Usage

ex1220

Format

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

- **Island**: a character vector indicating the island
- **Total**: total number of observed species
- **Native**: number of native species
- **Area**: area (km$^2$)
- **Elev**: elevation (m)
- **DistNear**: distance from nearest island (km)
- **DistSc**: distance from Santa Cruz (km)
- **AreaNear**: area of nearest island (km$^2$)

Source


References


Examples

\[ \text{str(ex1220)} \]
Predicting Desert Wildflower Blooms

Description
These data are monthly rainfalls from September to March and the subjectively rated quality of the following spring wildflower display for each of a number of years at each of four desert locations in the southwestern United States (Upland Sonoran Desert near Tucson, the lower Colorado River Valley section of the Sonoran Desert, the Baja California region of the Sonoran Desert, and the Mojave Desert). The quality of the display was judged subjectively with ordered rating categories of poor, fair, good, great, and spectacular. The variable Score is numerical variable corresponding to these ordered categories. A goal is to find an equation for predicting quality of wildflower blooms from the rainfall variables.

Usage
ex1221

Format
A data frame with 122 observations on the following 12 variables.

Year year of observed wildflower season
Region a factor variable with 4 levels: "baja", "colorado", "mojave", and "upland"
Sep the September rainfall, in inches
Oct the October rainfall, in inches
Nov the November rainfall, in inches
Dec the December rainfall, in inches
Jan the January rainfall, in inches
Feb the February rainfall, in inches
Mar the March rainfall, in inches
Total the total rainfall from September through March, in inches
Rating a factor with a subjective assessment of the quality of wildflower bloom with levels "FAIR", "GOOD", "GREAT", "POOR", and "SPECTACULAR"
Score a numerical variable corresponding to the order of rating categories, with Poor=0, Fair=1, Good=2, Great=3, and Spectacular=4

Source

References

Examples
str(ex1221)
Bush Gore Ballot Controversy

Description

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

Usage

ex1222

Format

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

- **County**: a character vector indicating the county
- **Buchanan2000**: votes cast for Buchanan in 2000 presidential election
- **Gore2000**: votes cast for Gore in 2000 presidential election
- **Bush2000**: votes cast for Bush in 2000 presidential election
- **Nader2000**: votes cast for Nader in 2000 presidential election
- **Browne2000**: votes cast for Browne in 2000 presidential election
- **Total2000**: total votes cast in 2000 presidential election
- **Clinton96**: votes cast for Clinton in 1996 presidential election
- **Dole96**: votes cast for Dole in 1996 presidential election
- **Perot96**: votes cast for Perot in 1996 presidential election
- **Buchanan96p**: votes cast for Buchanan in 1996 Republican primary
- **ReformReg**: the registration in Buchanan’s Reform party
- **TotalReg**: the total political party registration

Source


See Also

- ex0825

Examples

```r
str(ex1222)
```
Description

This is a subset of 2,584 individuals from the 1979 National Longitudinal Study of Youth (NLSY79) survey who were re-interviewed in 2006, who had paying jobs in 2005, and who had complete values for the variables listed below. A goal is to see whether intelligence (as measured by the ASVAB intelligence test score, AFQT, and its Components, Word, Parag, Math, and Arith) is a better predictor of 2005 income than education and socioeconomic status.

Usage

ex1223

Format

A data frame with 2,584 observations on the following 32 variables.

Subject  the subject identification number
Imagazine  a variable taking on the value 1 if anyone in the respondent’s household regularly read magazines in 1979, otherwise 0
Inewspaper  a variable taking on the value 1 if anyone in the respondent’s household regularly read newspapers in 1979, otherwise 0
Ilibrary  a variable taking on the value 1 if anyone in the respondent’s household had a library card in 1979, otherwise 0
MotherEd  mother’s years of education
FatherEd  father’s years of education
FamilyIncome78  family’s total net income in 1978
Race  1 = Hispanic, 2 = Black, 3 = Not Hispanic or Black
Gender  a factor with levels "female" and "male"
Educ  years of education completed by 2006
Science  score on the General Science test in 1981
Arith  score on the Arithmetic Reasoning test in 1981
Word  score on the Word Knowledge Test in 1981
Parag  score on the Paragraph Comprehension test in 1981
Numer  score on the Numerical Operations test in 1981
Coding  score on the Coding Speed test in 1981
Auto  score on the Automotive and Shop test in 1981
Math  score on the Mathematics Knowledge test in 1981
Mechanic  score on the Electronics Information test in 1981
Elec  score on the Paragraph Comprehension test in 1981
AFQT  percentile score on the AFQT intelligence test in 1981
Income2005  total annual income in 2005
Esteem1  self reported answer to 1st self esteem question, 2005
Esteem2  self reported answer to 2nd self esteem question, 2005
Esteem3  self reported answer to 3rd self esteem question, 2005
Esteem4  self reported answer to 4th self esteem question, 2005
Esteem5  self reported answer to 5th self esteem question, 2005
Esteem6  self reported answer to 6th self esteem question, 2005
Esteem7  self reported answer to 7th self esteem question, 2005
Esteem8  self reported answer to 8th self esteem question, 2005
Esteem9  self reported answer to 9th self esteem question, 2005
Esteem10 self reported answer to 10th self esteem question, 2005

Source

References

See Also
ex0222, ex0330, ex0331, ex0524, ex0525, ex0828, ex0923, ex1033

Examples
str(ex1223)
Dinosaur Extinctions—An Observational Study

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

Usage
ex1317

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

Iridium Iridium in samples (ppt)
Strata a factor with levels “Limestone” and “Shale”
DepthCat a factor with six levels: “1”, “2”, . . . , “6”
**Description**

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

**Usage**

```r
ex1319
```

**Format**

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

- **IQ**: IQ scores of adopted children
- **Adoptive**: a factor with levels "High" and "Low"; the socioeconomic status of the adoptive parents
- **Biological**: a factor with levels "High" and "Low"; the socioeconomic status of the biological parents

**Source**


**References**

Gender Differences in Performance on Mathematics Achievement Tests

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

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

Usage

ex1320

Format

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

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

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

Score ACT mathematics test score

Source


References


Examples

str(ex1320)
Description

A data set simulated to match the summary statistics and conclusions from Rosenthal and Jacobson’s Pygmalion study on elementary school students. The researchers assigned students at random to a pygmalion or control treatment group. They supplied information to the teachers of those in the pygmalion group with the false information that an intelligence test had indicated that the student was likely to excel. The researchers wished to see if the change in intelligence test scores for the students tended to be larger for those students labeled as likely to excel.

Usage

ex1321

Format

A data frame with 320 observations on the following 5 variables.

- **Student** a student identification number
- **Grade** the student’s grade, 1 through 6
- **Class** a factor with 17 levels "1a", "1b", and so on, to indicate the 17 distinct teacher/classrooms.
- **Treatment** a factor with levels "pygmalion" and "control" corresponding to whether the researchers had told the teacher that the student was “likely to succeed” or not
- **Gain** the intelligence test score taken at the end of the school year minus the intelligence test score taken at the begging of the school year

Source


References


Examples

str(ex1321)
Description

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

Usage

ex1416

Format

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

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

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

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

DaysPost days after inoculation

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

LH concentration of antibody in the unaffected part of the brain

Source


References


See Also

case1102, ex1417

Examples

str(ex1416)
Second Replicate of the Barrier Disruption Study

Description

Researchers designed an experiment to investigate how delivery of brain cancer antibody is influenced by tumor size, antibody molecular weight, blood-brain barrier disruption, and delivery route. The data for the first replicate of this study is in ex1416. This is the second replicate for the study.

Usage

ex1417

Format

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

- **Agent** a factor with levels "AIB", "DEX70" and "MTX"
- **Treatment** a factor with levels "BD" and "NS"
- **Route** a factor with levels "IA" and "IV"
- **DaysPost** days after inoculation
- **BAT** concentration of antibody in the part of the brain around the tumor
- **LH** concentration of antibody in the unaffected part of the brain

Source


References


See Also

- **case1102, ex1416**

Examples

```
str(ex1417)
```
Description

These data were simulated to match the summary statistics and conclusions of Rosenthal and Fode’s Clever Hans experiment. Each of 12 students trained rats to run a maze. The data set contains their number of successful runs out of 50 on each of 5 days. It also shows two summarizing statistics for each student: the overall success rate on all 5 days and the slope in the least squares regression of daily success rate (number of successes in a day divided by 50) on day. Also included are the student’s response to the prior expectation of success question and the student’s response to a post-experiment question about how relaxed they felt handling their rats (with higher values corresponding to more relaxed). The treatment variable shows whether or not the students were supplied with the fictitious information about whether their rats were bright or not.

Usage

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

- **Student**: a student identification number
- **PriorExp**: the student’s prior expectation of rat-training success, on a scale from -10 to 10
- **Block**: a numerical variable for pairs of students grouped according to their values of PriorExp
- **Treatment**: a factor with levels "bright" and "dull" corresponding to whether students were told (falsely) that their rats were bright or not
- **Day1**: the number of successful rat mazed runs on day 1
- **Day2**: the number of successful rat mazed runs on day 2
- **Day3**: the number of successful rat mazed runs on day 3
- **Day4**: the number of successful rat mazed runs on day 4
- **Day5**: the number of successful rat mazed runs on day 5
- **Relax**: degree of relaxation students felt in handling their rats, on a scale from 0 to 10
- **Success**: the total proportion of successful maze runs in 5 days
- **Slope**: the slope in the least squares regression of mean daily success as a function of day, estimated for each student individually

Source


References

Diet Wars

Description

These data are the weight losses of subjects randomly assigned to one of three diets, and these additional covariates sex, initial age, and body mass index. Is there any evidence from these data that the mean weight loss differs for the different diets, after accounting for the effect of the covariates? How big are the difference?

Usage

ex1420

Format

A data frame with 272 observations on the following 6 variables.

- **Subject**: a subject identification number
- **Diet**: a factor with levels "Low-Carbohydrate", "Low-Fat" and "Mediterranean"
- **Sex**: a factor with levels "F" and "M"
- **Age**: subject’s age in years
- **BMI**: body mass index in kg/squared meter
- **WtLoss24**: weight at the end of the 24 month study minus initial weight, in kg

Source


See Also

ex0623, ex1921, ex1922

Examples

str(ex1420)
Global Warming, Southern Hemisphere

Description
The data are the temperatures (in degrees Celsius) averaged for the southern hemisphere over a full year, for years 1850 to 2010. The 161-year average temperature has been subtracted, so each observation is the temperature difference from the series average.

Usage
ex1507

Format
A data frame with 161 observations on the following 2 variables.

- **Year**  year in which yearly average temperature was computed, from 1850 to 2010
- **Temperature** southern hemisphere temperature minus the 161-year average (degrees Celsius)

Source

References

Examples
str(ex1507)

---

Sunspots

Description
The data are the annual sunspot counts in each year from 1700 to 2010.

Usage
ex1507

Format
A data frame with 311 observations on the following 2 variables.

- **Year**  year in which sunspots were counted, from 1700 to 2010
- **Sunspots** the number of sunspots observed in a year
Source

References

Examples
str(ex15/zero.noslash9)

---

### Melanoma and Sunspot Activity—An Observational Study

Description
Several factors suggest that the incidence of melanoma is related to solar radiation. These data are the age-adjusted melanoma incidence among males in the Connecticut Tumor Registry and the sunspot activity, 1936–1972.

Usage
ex1514

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

- **Year** year
- **Melanoma** male melanoma incidence in number of cases per 100,000 population
- **Sunspot** sunspot relative number

Source

References

Examples
str(ex1514)
**Lynx Trappings and Sunspots**

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

**Usage**
ex1515

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

- **Year** year
- **Lynx** number of lynx trapped
- **Sunspots** number of sunspots

**Source**

**References**

**Examples**
str(ex1515)

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

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

**Usage**
ex1516
Format
A data frame with 26 observations on the following 3 variables.

**Year** year

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

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

Source

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

Examples
```
str(ex1516)
```

### Description
The Standard and Poor’s 500 stock index (S&P 500) is a benchmark of stock market performance, based on 400 industrial firms, 40 financial stocks, 40 utilities, and 20 transportation stocks. These data include the value of a $1 investment in 1871 at the end of each year from 1871 to 1999, according to the S&P 500, assuming all dividends are reinvested. Describe the distribution of the S&P value as a function of year.

Usage
```
ex1517
```

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

**Year** year

**S.P500Return** Value of Stock at the end of the year

Source

Examples
```
str(ex1517)
```
**Effectiveness of Measles Vaccine**

**Description**

The data are the number of measles cases reported in the United States for each year from 1950 to 2008. A goal is to explore the effect of the introduction of the measles vaccine in 1963 on the series mean.

**Usage**

ex1518

**Format**

A data frame with 59 observations on the following 3 variables.

- **Year**  
  year

- **Cases**  
  number of measles cases

- **Vaccine**  
  a factor with levels "no" and "yes" indicating whether the measles vaccine had been licensed or not (yes for every year starting with 1963)

**Source**


**References**

Center for Disease Control,  

**Examples**

str(ex1518)

---

**El Nino and the Southern Oscillation**

**Description**

The data are the Sea Surface Temperatures (SST) and Southern Oscillation Index (SOI) measurements from 1950 to 2010.

**Usage**

ex1519
Format

A data frame with 732 observations on the following 4 variables.

- **Year**  year
- **Month**  a numerical variable for month, with 1 = January
- **SOI**  the Southern Oscillation Index
- **SST**  the Sea Surface Temperature

Source


References

U.S. National Oceanographic and Atmospheric Administration (http://www.cpc.ncep.noaa.gov/data/indices/)

See Also

case1502

Examples

str(ex1519)

---

**ex1605**  *Nature—Nurture*

Description

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

Usage

ex1605

Format

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

- **FMED**  adoptive (foster) mother’s years of education
- **TMIQ**  biological mother’s score on IQ test
- **Age2IQ**  IQ of child at age 2
- **Age4IQ**  IQ of child at age 4
- **Age8IQ**  IQ of child at age 8
- **Age13IQ**  IQ of child at age 13
Source

References

See Also
ex1319

Examples
str(ex16/zero.noslash5)

---

**ex1611 Religious Competition**

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

Usage
ex1611

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

*Country* a character vector indicating the country

*PctCatholic* percent Catholics in the population

*PriestParishRatio* priest to parishioner ratio

*PctIndigenous* percent clergy indigenous

Source

References

Examples
str(ex1611)
**Wastewater**

**Description**

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

**Usage**

ex1612

**Format**

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

- **ComBOD** biochemical oxygen demand measurements from commercial laboratory
- **ComSS** suspended solids measurements from commercial laboratory
- **StaBOD** biochemical oxygen demand measurements from state laboratory
- **StaSS** suspended solids measurements from state laboratory

**Source**


**References**


**Examples**

```r
str(ex1612)
```

---

**Flea Beetle Distinction**

**Description**

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

**Usage**

ex1613
Format
A data frame with 36 observations on the following 4 variables.

- **Specimen**: specimen identification number
- **Jnt1**: measurement of first joint in micrometers
- **Jnt2**: measurement of second joint in micrometers
- **Species**: a factor with levels "conc" and "heik"

Source

References

Examples
```
str(ex1613)
```

---

**Psychoimmunology**

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

### Usage
```
ex1614
```

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

- **Subject**: subject identification number
- **PhaseA**: Interleukin-1 levels (counts per minute) from blood samples taken during activity A
- **PhaseB**: Interleukin-1 levels (counts per minute) from blood samples taken during activity B
- **PhaseC**: Interleukin-1 levels (counts per minute) from blood samples taken during activity C
Source

References

Examples

```r
str(ex1615)
```

---

**ex1615**  
*Trends in SAT Scores*

**Description**

**Usage**

```r
ex1615
```

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

- **State** a character vector indicating the state
- **M.V.89** average MATH SAT scores divided by average VERBAL SAT score in 1989
- **M.V.96** average MATH SAT scores divided by average VERBAL SAT score in 1996
- **M.V.97** average MATH SAT scores divided by average VERBAL SAT score in 1997
- **M.V.98** average MATH SAT scores divided by average VERBAL SAT score in 1998
- **M.V.99** average MATH SAT scores divided by average VERBAL SAT score in 1999

**Source**

**Examples**

```r
str(ex1615)
```
**Description**

In an experiment to identify genes of the plant Arabidopsis that react to a particular pathogen, researchers used RNA sequencing to produce gene profiles for a number of plants not subjected to the pathogen and several plants subjected to the pathogen. Tests comparing the distribution of gene expression in the two groups were performed for each gene individually. The data are the p-values from all these tests. The goal is to use a identify a set of genes that differentially express in the two groups, subject to some specified value for expected false discovery rate, such as 5%.

**Usage**

ex1620

**Format**

A data frame with 20,245 observations on the following 3 variables.

- **Gene**: an identification number for genes
- **GeneName**: a character variable with the name of the gene
- **pValue**: the p-value from a test that the mean expression level for the gene differs in the two groups (ignoring multiple testing)

**Source**


**References**

Chang, J, Department of Botany and Plant Pathology, Oregon State University, personnal communication.

**Examples**

str(ex1620)
Format

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

**Fat**  actual pig fat (in percent)
**M1**  magnetic resonance image at location 1
**M2**  magnetic resonance image at location 2
**M3**  magnetic resonance image at location 3
**M4**  magnetic resonance image at location 4
**M5**  magnetic resonance image at location 5
**M6**  magnetic resonance image at location 6
**M7**  magnetic resonance image at location 7
**M8**  magnetic resonance image at location 8
**M9**  magnetic resonance image at location 9
**M10** magnetic resonance image at location 10
**M11** magnetic resonance image at location 11
**M12** magnetic resonance image at location 12
**M13** magnetic resonance image at location 13

Source


References


Examples

```r
str(ex17/zero.noslash8)
```

---

ex1715  

*Church Distinctiveness*

Description

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

Usage

ex1715
Format

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

- **Denomination** a character vector indicating the church denomination
- **Distinct** distinctiveness (strictness of discipline on a seven point scale)
- **Attend** average percentage of weeks that individuals attended a church meeting (% weekly)
- **NonChurch** average number of secular organisations to which members belong
- **PctStrong** average percentage of members that describe themselves as being strong church members (%)
- **AnnInc** average income of members (US$)

Source


References


Examples

```r
str(ex1715)
```

---

**Description**

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

**Usage**

`ex1716`

**Format**

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

- **ZIP** last 2 digits of zip code
- **Fire** fires per 1000 housing units
- **Theft** thefts per 1000 population
- **Age** percentage of housing units built prior to 1940
- **Income** median family income
- **Race** percentage minority
- **Vol** number of new policies per 100 housing units
- **Invol** number of FAIR plan policies and renewals per 100 housing units
Source


References


Examples

```r
str(ex1716)
```

---

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

**Description**

Survival times for two groups of lymphoma patients.

**Usage**

```r
ex1914
```

**Format**

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

- **Months** months after diagnosis
- **Group** a factor with levels "no" and "radiation"
- **Survived** number of patients known to survive beyond this month
- **Died** number of patients known to die after this many months

**Source**


**References**


**Examples**

```r
str(ex1914)
```
Vitamin C and Colds

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

Usage
ex1916

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

- **Knew** a factor with levels "no" and "yes"
- **Treatment** a factor with levels "placebo" and "vitC"
- **Cold** number of people who got a cold
- **NoCold** number of people who did not get a cold

Source

Examples
str(ex1916)

Alcohol Consumption and Breast Cancer—A Retrospective Study

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

Usage
ex1917
**ex1918**

**Format**

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

- **BodyMass** a factor with levels "high", "low" and "medium"
- **Drinking** a factor with levels "high" and "low"
- **Cases** number of women with breast cancer
- **Controls** number of women without breast cancer

**Source**


**References**


**Examples**

```r
str(ex1917)
```

---

**ex1918  The Donner Party**

**Description**

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

**Usage**

```r
ex1918
```

**Format**

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

- **AgeCat** a numerical code corresponding to six age categories, with 1 = "15-19", 2 = "20-29", 3 = "30-39", 4 = "40-49", 5 = "50-59" and 6 = "60-69"
- **Sex** a factor with levels "female" and "male"
- **Lived** number that lived
- **Died** number that died

**Source**

References


See Also

case2001

Examples

`str(ex1918)`

<table>
<thead>
<tr>
<th>ex1919</th>
<th>Tire-Related Fatal Accidents and Ford Sports Utility Vehicles</th>
</tr>
</thead>
</table>

Description

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

Usage

`ex1919`

Format

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

- **SpeedCat** a numerical code corresponding to 4 categories of speed (in miles per hour), with 1 = "0-40", 2 = "41-55", 3 = "56-65" and 4 = ">65"
- **Make** a factor with levels "Ford" and "Other"
- **Other** cause of accident was other than tire-related
- **Tire** cause of accident was tire-related

Source


See Also

`ex2018`

Examples

`str(ex1919)`
Diet Wars II

Description
In the study of different diets for losing weight (ex0623, ex1420 and ex1922), there appear to have been many more experimental subjects that dropped out from the low carbohydrate diet group than from the other two diet groups. This data set contains the numbers who did and didn’t drop out in each diet group. Is there any evidence that the drop out rate differs in the three groups?

Usage
ex1921

Format
A data frame with 3 observations on the following 4 variables.

Diet a factor with levels "LowCarb", "LowFat", and "Medit"
DroppedOut the number of subjects who dropped out of the study
Completed the number of subjects who completed the study

Source

See Also
ex0623, ex1420, ex1922

Examples
str(ex1921)

Diet Wars III

Description
For the study of different diets for losing weight (ex0623, ex1420 and ex1921), it is desired to see whether women were more or less likely to drop out of the study than men (after accounting for the apparent differential drop out rates associated with diet). This data set includes the numbers that dropped out and completed the study for each combination of Sex and Diet.

Usage
ex1922
Format
A data frame with 6 observations on the following 4 variables.

**Diet** a factor with levels "LowCarb", "LowFat", and "Medit"

**Gender** a factor with levels "Men" and "Women"

**DroppedOut** the number of subjects who dropped out of the study

**Completed** the number of subjects who completed the study

Source

See Also
ex0623, ex1420, ex1921

Examples
str(ex1922)

---

**Description**
One issue concerning the validity of the clutch volume and parental care study of ex1031 is the selection of the bird species in the set of currently living animals. Was the selection just as good as a random sample of species from each of the groups? One way to study this for birds, at least, is to compare the numbers of species from each of the 29 orders of birds in the study with the known total number of species in each of the orders. If the selection of birds had been at random, the expected proportion of species in the study from one particular order, n, is the proportion of all species in that order (N=9,866) times the total number of species in the sample (414). That is, the expected number in each sample, if random sampling were used, is (N/9,866)x 414. Calculate the expected numbers and compare the observed numbers with them using Pearson’s chi-square statistic.

**Usage**
ex1923

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

**Order** a character variable with the name of the order

**N** the known number of species in the order

**n** the number of sampled species from the order

Source
Description

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

Usage

ex2011

Format

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

- **Temperature**: Launch temperature (in degrees Fahrenheit)
- **Failure**: Indicator of O-ring failure

Source


See Also

case0401, ex2223

Examples

str(ex2011)
Description

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

Usage

ex2012

Format

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

- **Group**  Indicator whether the woman has DMD ("Case") or not ("Control")
- **CK**  Creatine kinase reading
- **H**  Hemopexin reading

Source


References


Examples

str(ex2012)
Description

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

Usage

ex2015

Format

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

Site  Site, a factor with levels "Random" and "Nest"
PctRing1  Percentage of mature forest in ring with outer radius 0.91 km
PctRing2  Percentage of mature forest in ring with outer radius 1.18 km
PctRing3  Percentage of mature forest in ring with outer radius 1.40 km
PctRing4  Percentage of mature forest in ring with outer radius 1.60 km
PctRing5  Percentage of mature forest in ring with outer radius 1.77 km
PctRing6  Percentage of mature forest in ring with outer radius 2.41 km
PctRing7  Percentage of mature forest in ring with outer radius 3.38 km

Source


References


Examples

str(ex2015)
Hermon Bumpus analysed various characteristics of some house sparrows that were found on the ground after a severe winter storm in 1898. Some of the sparrows survived and some perished. This data set contains the survival status, age, the length from tip of beak to tip of tail (in mm), the alar extent (length from tip to tip of the extended wings, in mm), the weight in grams, the length of the head in mm, the length of the humerus (arm bone, in inches), the length of the femur (thigh bones, in inches), the length of the tibio–tarsus (leg bone, in inches), the breadth of the skull in inches and the length of the sternum in inches.

**Usage**

ex2016

**Format**

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

- **Status** Survival status, factor with levels "Perished" and "Survived"
- **AG** a numerical code corresponding to two categories of age, with 1 = "adult" and 2 = "juvenile"
- **TL** total length (in mm)
- **AE** alar extent (in mm)
- **WT** weight (in grams)
- **BH** length of beak and head (in mm)
- **HL** length of humerus (in inches)
- **FL** length of femur (in inches)
- **TT** length of tibio–tarsus (in inches)
- **SK** width of skull (in inches)
- **KL** length of keel of sternum (in inches)

**Source**


**See Also**

ex0221

**Examples**

str(ex2016)
The Catholic church has explicitly opposed authoritarian rule in some (but not all) Latin American countries. Although such action could be explained as a desire to counter repression or to increase the quality of life of its parishioners, A.J. Gill supplies evidence that the underlying reason may be competition from evangelical Protestant denominations. He compiled the data given in this data frame.

Usage

ex2017

Format

A data frame with 12 observations on the following 5 variables.

Stance Catholic church stance, factor with levels "Pro" and "Ant1"
Country Latin American country
PQLI Physical Quality of Life Index in the mid-1970s; Average of live expectancy at age 1, infant mortality and literacy at age 15+.
Repression Average civil rights score for the period of authoritarian rule until 1979
Competition Percentage increase of competitive religious groups during the period 1900–1970

Source


References


Examples

str(ex2017)
Description

This data frame contains data on 1995 and later model compact sports utility vehicles involved in fatal accidents in the United States between 1995 and 1999, excluding those that were struck by another car and excluding accidents that, according to police reports, involved alcohol.

Usage

ex2018

Format

A data frame with 2,321 observations on the following 4 variables.

- **Make** Type of sports utility vehicle, factor with levels "Other" and "Ford"
- **VehicleAge** Vehicle age (in years); surrogate for age of tires
- **Passengers** Number of passengers
- **Cause** Cause of fatal accident, factor with levels "NotTire" and "Tire"

Details

The Ford Explorer is a popular sports utility vehicle made in the United States and sold throughout the world. Early in its production concern arose over a potential accident risk associated with tires of the prescribed size when the vehicle was carrying heavy loads, but the risk was thought to be acceptable if a low tire pressure was recommended. The problem was apparently exacerbated by a particular type of Firestone tire that was overly prone to separation, especially in warm temperatures. This type of tire was a common one used on Explorers in model years 1995 and later. By the end of 1999 more than 30 lawsuits had been filed over accidents that were thought to be associated with this problem. U.S. federal data on fatal car accidents were analysed at that time, showing that the odds of a fatal accident being associated with tire failure were three times as great for Explorers as for other sports utility vehicles.

Additional data from 1999 and additional variables may be used to further explore the odds ratio. It is of interest to see whether the odds that a fatal accident is tire-related depend on whether the vehicle is a Ford, after accounting for age of the car and number of passengers. Since the Ford tire problem may be due to the load carried, there is some interest in seeing whether the odds associated with a Ford depend on the number of passengers.

Source


See Also

ex1919

Examples

str(ex2018)
Missile Defenses

Description
Following a successful test of an anti-ballistic missile (ABM) in July 1999, many prominent U.S. politicians called for the early deployment of a full ABM system. The scientific community was less enthusiastic about the efficacy of such a system. This data set contains the success or failure of ABM tests between March 1983 and May 1995. Do these data suggest any improvement in ABM test success probability over time?

Usage
ex2019

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

- **Date**  date of an ABM test
- **Months** number of months after March 1983
- **Result** a factor with levels "Failure" and "Success"

Source

References

Examples
str(ex2019)

Vitamin C and Colds

Description
These data are from a randomized experiment to assess the effect of large doses of vitamin C on the incidence of colds.

Usage
ex2113
Format

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

- **Dose**  the daily dose of vitamin C, in g
- **Number**  the number of subjects given that dose of vitamin C
- **WithoutIllness**  the number of subjects who did not become ill
- **ProportionWithout**  the proportion of subjects who did not become ill

Source


References


Examples

```
str(ex2113)
```

---

**Description**

The study the effect of context questions prior to target questions, researchers conducted a poll involving 1,054 subjects selected randomly from the Chicago phone directory. To include possibly unlisted phones, selected numbers were randomly altered in the last position. This data frame contains the responses to one of the questions asked concerning continuing U.S. aid to the Nicaraguan Contra rebels. Eight different versions of the interview were given, representing all possible combinations of three factors at each of two levels. The experimental factors were Context, Mode and Level.

Context refers to the type of context questions preceding the question about Nicaraguan aid. Some subjects received a context question about Vietnam, designed to elicit reticence about having the U.S. become involved in another foreign war in a third–world country. The other context question was about Cuba, designed to elicit anti–communist sentiments.

Mode refers to whether the target question immediately followed the context question or whether there were other questions scattered in between.

Level refers to two versions of the context question. In the "high" level the question was worded to elicit a higher level of agreement than in the "low" level wording.

**Usage**

ex2115
Format

A data frame with 8 observations on the following 7 variables.

- **Context** Factor referring to the context of the question preceding the target question about U.S. aid to the Nicaraguan Contra rebels
- **Mode** Factor with levels "not" and "scattered". "scattered" is used if the target question was not asked directly after the context question
- **Level** Factor with levels "low" and "high", refers to the wording of the question
- **Number** Number of people interviewed
- **InFavor** Number of people in favor of Contra Aid
- **NotInFavor** Number of people not in favor of Contra Aid
- **PercentInFavor** Percentage in favour of Contra aid

Details

Increasingly, politicians look to public opinion surveys to shape their public stances. Does this represent the ultimate in democracy? Or are seemingly scientific polls being rigged by the manner of questioning? Psychologists believe that opinions—expressed as answers to questions—are usually generated at the time the question is asked. Answers are based on a quick sampling of relevant beliefs held by the subject, rather than a systematic canvas of all such beliefs. Furthermore, this sampling of beliefs tends to overrepresent whatever beliefs happen to be most accessible at the time the question is asked. This aspect of delivering opinions can be abused by the pollster. Here, for example, is one sequence of questions:

1. “Do you believe the Bill of Rights protects personal freedom?”
2. “Are you in favor of a ban on handguns?”

Here is another:

1. “Do you think something should be done to reduce violent crime?”
2. “Are you in favor of a ban on handguns?”

The proportion of yes answers to question 2 may be quite different depending on which question 1 is asked first.

Source


References


Examples

```r
str(ex2115)
```
ex2116  Aflatoxicol and Liver Tumors in Trout

Description
An experiment at the Marine/Freshwater Biomedical Sciences Center at Oregon State University investigated the carcinogenic effects of aflatoxicol, a metabolite of Aflatoxin B1, which is a toxic by-product produced by a mold that infects cottonseed meal, peanuts and grains. Twenty tanks of rainbow trout embryos were exposed to one of five doses of Aflatoxicol for one hour. The data represent the numbers of fish in each tank and the numbers of these that had liver tumours after one year.

Usage
ex2116

Format
A data frame with 20 observations on the following 3 variables.
- **Dose**  Dose (in ppm)
- **Tumor**  Number of trout with liver tumours
- **Total**  Number of trout in tank

Source

Examples
str(ex2116)

---

ex2117  Effect of Stress During Conception on Odds of a Male Birth

Description
The probability of a male birth in humans is about .51. It has previously been noticed that lower proportions of male births are observed when offspring is conceived at times of exposure to smog, floods or earthquakes. Danish researchers hypothesised that sources of stress associated with severe life events may also have some bearing on the sex ratio. To investigate this theory they obtained the sexes of all 3,072 children who were born in Denmark between 1 January 1980 and 31 December 1992 to women who experienced the following kind of severe life events in the year of the birth or the year prior to the birth: death or admission to hospital for cancer or heart attack of their partner or of their other children. They also obtained sexes on a sample of 20,337 births to mothers who did not experience these life stress episodes. This data frame contains the data that were collected. Noticed that for one group the exposure is listed as taking place during the first trimester of pregnancy. The rationale for this is that the stress associated with the cancer or heart attack of a family member may well have started before the recorded time of death or hospital admission.
Usage

ex2117

Format

A data frame with 5 observations on the following 4 variables.

- **Group**: Indicator for groups to which mothers belong
- **Time**: Indicator for time at which severe life event occurred
- **Number**: Number of births
- **PctBoys**: Percentage of boys born

Source


References


Examples

```r
str(ex2117)
```

Description

Researchers in Kenya identified a cohort of more that 1,000 prostitutes who were known to be a major reservoir of sexually transmitted diseases in 1985. It was determined that more than 85% of them were infected with human immunodeficiency virus (HIV) in February, 1986. The researchers identified men who acquired a sexually-transmitted disease from this group of women after the men sought treatment at a free clinic. The data frame contains data on the subset of those men who did not test positive for HIV on their first visit and who agreed to participate in the study. The men are categorised according to whether they later tested positive for HIV during the study period, whether they had one or multiple sexual contacts with the prostitutes and whether they were circumcised.

Usage

ex2118
Format

A data frame with 4 observations on the following 5 variables.

- **Contact**: Whether men had single or multiple contact with prostitutes
- **Circumcised**: Whether the men are circumcised, factor with levels "No" and "Yes"
- **HIV**: Number of men that tested positive for HIV
- **Number**: Number of men
- **NoHIV**: Number of men that did not test positive for HIV (should be Number-HIV)

Source


References


Examples

```r
str(ex2118)
```

---

**ex2119**  
*Meta–Analysis of Breast Cancer and Lactation Studies*

Description

This data frame gives the results of 10 separate case–control studies on the association of breast cancer and whether a woman had breast–fed children.

Usage

```r
ex2119
```

Format

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

- **Study**: Factor indicating the study from which data was taken
- **Lactate**: Whether women had breast–fed children (lactated)
- **Cancer**: Number of women with breast cancer
- **NoCancer**: Number of women without breast cancer
Details

Meta-analysis refers to the analysis of analyses. When the main results of studies can be cast into 2×2 tables of counts, it is natural to combine individual odds ratios with a logistic regression model that includes a factor to account for different odds from the different studies. In addition, the odds ratio itself might differ slightly among studies because of different effects on different populations or different research techniques. One approach for dealing with this is to suppose an underlying common odds ratio and to model between-study variability as extra-binomial variation.

Source


References

Data gathered from various sources by Karolyn Kolassa as part of a Master’s project, Oregon State University.

Examples

str(ex2119)

---

<table>
<thead>
<tr>
<th>ex2120</th>
<th>Clever Hans Effect</th>
</tr>
</thead>
</table>

Description

These data were simulated to match the summary statistics and conclusions of Rosenthal and Fode’s Clever Hans experiment. Each of 12 students trained rats to run a maze. The data set contains their number of successful runs out of 50 on each of 5 days, the student’s prior expectation of success (on a scale from -10 to 10), and a variable indicating treatment—whether or not the students were supplied with the fictitious information that their rights were bright.

Usage

2120

Format

A data frame with 60 observations on the following 5 variables.

- **Student**: a student identification number
- **PriorExp**: the student’s prior expectation of rat-training success, on a scale from -10 to 10
- **Treatment**: a factor with levels “bright” and “dull” corresponding to whether students were told (falsely) that their rats were bright or not
- **Day**: day of the study, ranging from 1 to 5
- **Success**: the number of successful maze runs on a day, out of 50
Description
Some sociologist suspect that highly publicised suicides may trigger additional suicides. In one investigation of this hypothesis, D.P. Phillips collected information about 17 airplane crashes that were known (because of notes left behind) to be murder–suicides. For each of these crashes, Phillips reported an index of the news coverage (circulation of nine newspapers devoting space to the crash multiplied by length of coverage) and the number of multiple-fatality plane crashes during the week following the publicised crash. This data frame contains the collected data.

Usage
ex2216

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

Index  Index for the amount of newspaper coverage given the murder–suicide
Crashes Multiple-fatality crashes in the week following a murder–suicide by plane crash

Source

References

Examples
str(ex2216)
Description

The data are the number of cancer deaths among survivors of the atomic bombs dropped on Japan during World War II, categorized by time (years) after the bomb that death occurred and the amount of radiation exposure that the survivors received from the blast. Also listed in each cell is the person-years at risk, in 100's. This is the sum total of all years spent by all persons in the category. The data can be analyzed by supposing the number of cancer deaths in each cell is Poisson with mean = risk x rate, where risk is the person-years at risk and rate is the rate of cancer deaths per person per year. How does the rate depend on the radiation exposure, after accounting for years after exposure?

Usage

ex2220

Format

A data frame with 42 observations on the following 4 variables.

Exposure radiation exposure, in rads
YearsAfter years after the exposure
AtRisk number of survivors in the group
Deaths number of survivors in the group who died of Cancer

Source


References

Pierce, D.A., personal communication

Examples

str(ex2220)
Usage

ex2222

Format

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

- **Count**: Number of occurrences of a word in various chapters of books written by Jane Austen and the ghost writer
- **Book**: Title of books used
- **Word**: Words used

Source


References


Examples

```r
str(ex2222)
```

---

### ex2223  
*Space Shuttle O-Ring Failures*

Description

On January 27, 1986, the night before the space shuttle *Challenger* exploded, an engineer recommended to the National Aeronautics and Space Administration (NASA) that the shuttle not be launched in the cold weather. The forecasted temperature for the *Challenger* launch was 31 degrees Fahrenheit—the coldest launch ever. After an intense 3-hour telephone conference, officials decided to proceed with the launch. This data frame contains the launch temperatures and the number of O-ring problems in 24 shuttle launches prior to the *Challenger*.

Usage

ex2223

Format

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

- **Temp**: Launch temperatures (in degrees Fahrenheit)
- **Incidents**: Numbers of O-ring incidents
Source


See Also

case/zero.noslash4/zero.noslash1, ex2/zero.noslash11

Examples

`str(ex2223)`

---

### ex2224 Valve Failure in Nuclear Reactors

**Description**

This data frame contains data on characteristics and numbers of failures observed in valve types from one pressurised water reactor.

**Usage**

`ex2224`

**Format**

A data frame with 90 observations on the following 7 variables.

- **System** a numerical code corresponding to 5 categories of system (1 = containment, 2 = nuclear, 3 = power conversion, 4 = safety, 5 = process auxiliary)
- **Operator** a numerical code corresponding to 4 different operator types (1 = air, 2 = solenoid, 3 = motor-driven, 4 = manual)
- **Valve** a numerical code corresponding to 6 different valve types (1 = ball, 2 = butterfly, 3 = diaphragm, 4 = gate, 5 = globe, 6 = directional control)
- **Size** a numerical code corresponding to 3 head size categories (1 = less than 2 inches, 2 = 2–10 inches, 3 = 10–30 inches)
- **Mode** a numerical code corresponding to two categories of operation mode (1 = normally closed, 2 = normally open)
- **Failures** Number of failures observed
- **Time** Lengths of observation time

**Source**


**References**

Examples

str(ex2224)

---

**Body Size and Reproductive Success in a Population of Male Bullfrogs**

**Description**

As an example of field observation in evidence of theories of sexual selection, S.J. Arnold and M.J. Wade presented the following data set on size and number of mates observed in 38 bullfrogs.

**Usage**

ex2225

**Format**

A data frame with 38 observations on the following 2 variables.

- **BodySize**: Body size (in mm)
- **Mates**: Number of mates

**Source**


**References**


---

**Number of Moons**

**Description**

Apparently, larger planets have more moons, but is it the volume (as indicated by diameter) or mass that are more relevant, or is it both? These data include the diameter, mass, distance from the sun, and number of moons for 13 planets, gas giants, and dwarf planets in our solar system. Which size variable best explains mean number of moons (possible after accounting for distance from sun). (Consider negative binomial regression.)

**Usage**

ex2226
**Format**

A data frame with 13 observations on the following 5 variables.

- **Name**: a character variable with the name of the planet, gas giant, or dwarf planet.
- **Distance**: distance from sun, relative to earth’s.
- **Diameter**: diameter of the planet, relative to earth’s.
- **Mass**: mass, relative to earth’s.
- **Moons**: number of moons.

**Source**


**References**


**See Also**

ex0721

**Examples**

```r
str(ex2226)
```

## ex2414  Amphibian Crisis and UV-B

**Description**

Data frame contains the percentage of unsuccessful hatching from enclosures containing 150 eggs each in a study to investigate whether UV-B is responsible for low hatch rates.

**Usage**

ex2414

**Format**

A data frame with 71 observations on the following 4 variables.

- **Percent**: percentage of frog eggs failing to hatch.
- **Treatment**: factor variable with levels "NoFilter", "UV-BTransmitting" and "UV-BBlocking".
- **Location**: factor variable with levels "ThreeCreeks", "SparksLake", "SmallLake" and "LostLake".
- **Photolyase**: Photolyase activity.

**Source**

References


Examples

```r
str(ex2414)
```

---

**Description**

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

**Usage**

```r
Sleuth3Manual()
```

**Author(s)**

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

**References**


**Examples**

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

<table>
<thead>
<tr>
<th>Topic</th>
<th>datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>case0101, 5</td>
<td>case2102, 70</td>
</tr>
<tr>
<td>case0102, 6</td>
<td>case2201, 71</td>
</tr>
<tr>
<td>case0201, 7</td>
<td>case2202, 73</td>
</tr>
<tr>
<td>case0202, 8</td>
<td>ex0112, 74</td>
</tr>
<tr>
<td>case0301, 9</td>
<td>ex0116, 75</td>
</tr>
<tr>
<td>case0302, 11</td>
<td>ex0125, 76</td>
</tr>
<tr>
<td>case0401, 12</td>
<td>ex0126, 76</td>
</tr>
<tr>
<td>case0402, 13</td>
<td>ex0127, 77</td>
</tr>
<tr>
<td>case0501, 14</td>
<td>ex0211, 78</td>
</tr>
<tr>
<td>case0502, 16</td>
<td>ex0218, 79</td>
</tr>
<tr>
<td>case0601, 17</td>
<td>ex0221, 80</td>
</tr>
<tr>
<td>case0602, 19</td>
<td>ex0222, 81</td>
</tr>
<tr>
<td>case0701, 20</td>
<td>ex0223, 82</td>
</tr>
<tr>
<td>case0702, 21</td>
<td>ex0321, 83</td>
</tr>
<tr>
<td>case0801, 23</td>
<td>ex0323, 83</td>
</tr>
<tr>
<td>case0802, 24</td>
<td>ex0327, 84</td>
</tr>
<tr>
<td>case0901, 25</td>
<td>ex0330, 85</td>
</tr>
<tr>
<td>case0902, 26</td>
<td>ex0331, 86</td>
</tr>
<tr>
<td>case1001, 28</td>
<td>ex0332, 86</td>
</tr>
<tr>
<td>case1002, 29</td>
<td>ex0333, 87</td>
</tr>
<tr>
<td>case1101, 31</td>
<td>ex0428, 88</td>
</tr>
<tr>
<td>case1102, 33</td>
<td>ex0429, 89</td>
</tr>
<tr>
<td>case1201, 35</td>
<td>ex030, 90</td>
</tr>
<tr>
<td>case1202, 37</td>
<td>ex0432, 91</td>
</tr>
<tr>
<td>case1301, 39</td>
<td>ex0518, 91</td>
</tr>
<tr>
<td>case1302, 40</td>
<td>ex0523, 92</td>
</tr>
<tr>
<td>case1401, 42</td>
<td>ex0524, 93</td>
</tr>
<tr>
<td>case1402, 43</td>
<td>ex0525, 94</td>
</tr>
<tr>
<td>case1501, 45</td>
<td>ex0623, 95</td>
</tr>
<tr>
<td>case1502, 47</td>
<td>ex0624, 95</td>
</tr>
<tr>
<td>case1601, 48</td>
<td>ex0721, 96</td>
</tr>
<tr>
<td>case1602, 50</td>
<td>ex0722, 97</td>
</tr>
<tr>
<td>case1701, 52</td>
<td>ex0724, 98</td>
</tr>
<tr>
<td>case1702, 54</td>
<td>ex0725, 98</td>
</tr>
<tr>
<td>case1801, 57</td>
<td>ex0726, 99</td>
</tr>
<tr>
<td>case1802, 59</td>
<td>ex0727, 100</td>
</tr>
<tr>
<td>case1803, 60</td>
<td>ex0728, 101</td>
</tr>
<tr>
<td>case1901, 61</td>
<td>ex0729, 101</td>
</tr>
<tr>
<td>case1902, 62</td>
<td>ex0730, 102</td>
</tr>
<tr>
<td>case2001, 64</td>
<td>ex0816, 103</td>
</tr>
<tr>
<td>case2002, 66</td>
<td>ex0817, 104</td>
</tr>
<tr>
<td>case2101, 68</td>
<td>ex0820, 104</td>
</tr>
</tbody>
</table>
ex0822, 106
ex0823, 106
ex0824, 107
ex0825, 108
ex0826, 108
ex0828, 109
ex0829, 110
ex0914, 111
ex0915, 111
ex0918, 112
ex0920, 113
ex0921, 114
ex0923, 115
ex1014, 116
ex1026, 116
ex1027, 117
ex1028, 118
ex1029, 119
ex1030, 120
ex1031, 121
ex1033, 122
ex1111, 123
ex1120, 123
ex1122, 124
ex1123, 125
ex1124, 126
ex1125, 126
ex1217, 127
ex1220, 129
ex1221, 130
ex1222, 131
ex1223, 132
ex1225, 133
ex1317, 134
ex1319, 135
ex1320, 136
ex1321, 137
ex1416, 138
ex1417, 139
ex1419, 140
ex1420, 141
ex1507, 142
ex1509, 142
ex1514, 143
ex1515, 144
ex1516, 144
ex1517, 145
ex1518, 146
ex1519, 146
ex1605, 147
ex1611, 148
ex1612, 149
ex1613, 149
ex1614, 150
ex1615, 151
ex1620, 152
ex1708, 152
ex1715, 153
ex1716, 154
ex1914, 155
ex1916, 156
ex1917, 156
ex1918, 157
ex1919, 158
ex1921, 159
ex1923, 160
ex2011, 161
ex2012, 162
ex2015, 163
ex2016, 164
ex2017, 165
ex2018, 166
ex2019, 167
ex2113, 167
ex2115, 168
ex2116, 170
ex2117, 170
ex2118, 171
ex2119, 172
ex2120, 173
ex2216, 174
ex2220, 175
ex2222, 175
ex2223, 176
ex2224, 177
ex2225, 178
ex2226, 178
ex2414, 179
* Topic documentation
  Sleuth3Manual, 180
* Topic package
  Sleuth3-package, 5

case0101, 5
case0102, 6, 37
case0201, 7, 79
case0202, 8
case0301, 9
case0302, 11
case0401, 12, 161, 177
case0402, 13
case0501, 14
case0502, 16
case0601, 17
INDEX

<table>
<thead>
<tr>
<th>ex/zero.noslash</th>
<th>183</th>
</tr>
</thead>
<tbody>
<tr>
<td>case0602, 19</td>
<td></td>
</tr>
<tr>
<td>case0701, 20, 98, 99</td>
<td></td>
</tr>
<tr>
<td>case0702, 21, 103</td>
<td></td>
</tr>
<tr>
<td>case0801, 23</td>
<td></td>
</tr>
<tr>
<td>case0802, 24</td>
<td></td>
</tr>
<tr>
<td>case0901, 25</td>
<td></td>
</tr>
<tr>
<td>case0902, 26, 88</td>
<td></td>
</tr>
<tr>
<td>case1001, 28</td>
<td></td>
</tr>
<tr>
<td>case1002, 29</td>
<td></td>
</tr>
<tr>
<td>case1101, 31</td>
<td></td>
</tr>
<tr>
<td>case1102, 33, 138, 139</td>
<td></td>
</tr>
<tr>
<td>case1201, 35</td>
<td></td>
</tr>
<tr>
<td>case1202, 6, 37</td>
<td></td>
</tr>
<tr>
<td>case1301, 39</td>
<td></td>
</tr>
<tr>
<td>case1302, 40</td>
<td></td>
</tr>
<tr>
<td>case1401, 42</td>
<td></td>
</tr>
<tr>
<td>case1402, 43</td>
<td></td>
</tr>
<tr>
<td>case1501, 45</td>
<td></td>
</tr>
<tr>
<td>case1502, 47, 147</td>
<td></td>
</tr>
<tr>
<td>case1601, 48</td>
<td></td>
</tr>
<tr>
<td>case1602, 50</td>
<td></td>
</tr>
<tr>
<td>case1701, 52</td>
<td></td>
</tr>
<tr>
<td>case1702, 54</td>
<td></td>
</tr>
<tr>
<td>case1801, 57</td>
<td></td>
</tr>
<tr>
<td>case1802, 59</td>
<td></td>
</tr>
<tr>
<td>case1803, 60</td>
<td></td>
</tr>
<tr>
<td>case1901, 61</td>
<td></td>
</tr>
<tr>
<td>case1902, 62</td>
<td></td>
</tr>
<tr>
<td>case2001, 64, 158</td>
<td></td>
</tr>
<tr>
<td>case2002, 66</td>
<td></td>
</tr>
<tr>
<td>case2101, 68</td>
<td></td>
</tr>
<tr>
<td>case2102, 70</td>
<td></td>
</tr>
<tr>
<td>case2201, 71</td>
<td></td>
</tr>
<tr>
<td>case2202, 73</td>
<td></td>
</tr>
<tr>
<td>ex0112, 74</td>
<td></td>
</tr>
<tr>
<td>ex0116, 75</td>
<td></td>
</tr>
<tr>
<td>ex0125, 76</td>
<td></td>
</tr>
<tr>
<td>ex0126, 76, 78</td>
<td></td>
</tr>
<tr>
<td>ex0127, 77, 77</td>
<td></td>
</tr>
<tr>
<td>ex0211, 78</td>
<td></td>
</tr>
<tr>
<td>ex0218, 8, 79</td>
<td></td>
</tr>
<tr>
<td>ex0221, 80, 164</td>
<td></td>
</tr>
<tr>
<td>ex0222, 81, 85, 86, 93, 94, 110, 115, 122, 133</td>
<td></td>
</tr>
<tr>
<td>ex0223, 82</td>
<td></td>
</tr>
<tr>
<td>ex0321, 83</td>
<td></td>
</tr>
<tr>
<td>ex0323, 83</td>
<td></td>
</tr>
<tr>
<td>ex0327, 84</td>
<td></td>
</tr>
<tr>
<td>ex0330, 81, 85, 86, 93, 94, 110, 115, 122, 133</td>
<td></td>
</tr>
<tr>
<td>ex0331, 81, 85, 86, 93, 94, 110, 115, 122, 133</td>
<td></td>
</tr>
<tr>
<td>ex0332, 86</td>
<td></td>
</tr>
<tr>
<td>ex0333, 27, 87</td>
<td></td>
</tr>
<tr>
<td>ex0428, 88</td>
<td></td>
</tr>
</tbody>
</table>
ex1222, 108, 131
ex1223, 81, 85, 86, 93, 94, 110, 115, 122, 132
ex1225, 133
ex1317, 134
ex1319, 135, 148
ex1320, 136
ex1321, 137
ex1416, 33, 138, 139
ex1417, 33, 138, 139
ex1419, 140, 174
ex1420, 95, 141, 159, 160
ex1507, 142
ex1509, 142
ex1514, 143
ex1515, 144
ex1516, 144
ex1517, 145
ex1518, 146
ex1519, 47, 146
ex1605, 135, 147
ex1611, 148
ex1612, 149
ex1613, 149
ex1614, 150
ex1615, 151
ex1620, 152
ex1708, 152
ex1715, 153
ex1716, 154
ex1914, 155
ex1916, 156
ex1917, 156
ex1918, 65, 157
ex1919, 158, 166
ex1921, 95, 141, 159, 159, 160
ex1922, 95, 141, 159, 159
ex1923, 121, 160
ex2011, 12, 161, 177
ex2012, 162
ex2015, 163
ex2016, 80, 164
ex2017, 165
ex2018, 158, 166
ex2019, 167
ex2113, 167
ex2115, 168
ex2116, 170
ex2117, 170
ex2118, 171
ex2119, 172
ex2120, 141, 173
ex2216, 174

Sleuth3 (Sleuth3-package), 5
Sleuth3-package, 5
Sleuth3Manual, 180