# Package ‘BSDA’

October 12, 2022

**Type** Package  
**Title** Basic Statistics and Data Analysis  
**Version** 1.2.1  
**Date** 2021-09-04  
**LazyData** yes  
**Maintainer** Alan T. Arnholt <arnholt@appstate.edu>  
**Description** Data sets for book ”Basic Statistics and Data Analysis” by Larry J. Kitchens.  
**Depends** lattice, R (>= 2.10)  
**Imports** e1071  
**License** GPL-3  
**Suggests** ggplot2 (>= 2.1.0), dplyr, tidyr  
**URL** [https://github.com/alanarnholt/BSDA.](https://github.com/alanarnholt/BSDA)  
[https://alanarnholt.github.io/BSDA/](https://alanarnholt.github.io/BSDA/)  
**BugReports** [https://github.com/alanarnholt/BSDA/issues](https://github.com/alanarnholt/BSDA/issues)  
**RoxygenNote** 7.1.1  
**NeedsCompilation** no  
**Author** Alan T. Arnholt [aut, cre], Ben Evans [aut]  
**Repository** CRAN  
**Date/Publication** 2021-09-05 04:40:01 UTC

## R topics documented:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbey</td>
<td>8</td>
</tr>
<tr>
<td>Abc</td>
<td>9</td>
</tr>
<tr>
<td>Abilene</td>
<td>10</td>
</tr>
<tr>
<td>Ability</td>
<td>11</td>
</tr>
<tr>
<td>Abortion</td>
<td>12</td>
</tr>
<tr>
<td>Topics Documented</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Absent</td>
<td>13</td>
</tr>
<tr>
<td>Achieve</td>
<td>14</td>
</tr>
<tr>
<td>Adsales</td>
<td>14</td>
</tr>
<tr>
<td>Aggress</td>
<td>15</td>
</tr>
<tr>
<td>Aid</td>
<td>16</td>
</tr>
<tr>
<td>Aids</td>
<td>17</td>
</tr>
<tr>
<td>Airdisasters</td>
<td>18</td>
</tr>
<tr>
<td>Airline</td>
<td>19</td>
</tr>
<tr>
<td>Alcohol</td>
<td>20</td>
</tr>
<tr>
<td>Allergy</td>
<td>20</td>
</tr>
<tr>
<td>Anesthet</td>
<td>21</td>
</tr>
<tr>
<td>Anxiety</td>
<td>22</td>
</tr>
<tr>
<td>Apolipop</td>
<td>22</td>
</tr>
<tr>
<td>Append</td>
<td>23</td>
</tr>
<tr>
<td>Appendec</td>
<td>24</td>
</tr>
<tr>
<td>Aptitude</td>
<td>25</td>
</tr>
<tr>
<td>Archaeo</td>
<td>25</td>
</tr>
<tr>
<td>Arthriti</td>
<td>26</td>
</tr>
<tr>
<td>Artifici</td>
<td>27</td>
</tr>
<tr>
<td>Asprin</td>
<td>28</td>
</tr>
<tr>
<td>Asthamati</td>
<td>28</td>
</tr>
<tr>
<td>Attorney</td>
<td>29</td>
</tr>
<tr>
<td>Autogear</td>
<td>30</td>
</tr>
<tr>
<td>Backtoback</td>
<td>31</td>
</tr>
<tr>
<td>Bbsalaries</td>
<td>31</td>
</tr>
<tr>
<td>Bigten</td>
<td>32</td>
</tr>
<tr>
<td>Biology</td>
<td>33</td>
</tr>
<tr>
<td>Birth</td>
<td>34</td>
</tr>
<tr>
<td>Blackedu</td>
<td>35</td>
</tr>
<tr>
<td>Blood</td>
<td>35</td>
</tr>
<tr>
<td>Board</td>
<td>36</td>
</tr>
<tr>
<td>Bones</td>
<td>37</td>
</tr>
<tr>
<td>Books</td>
<td>38</td>
</tr>
<tr>
<td>Bookstor</td>
<td>38</td>
</tr>
<tr>
<td>Brain</td>
<td>39</td>
</tr>
<tr>
<td>Bumpers</td>
<td>40</td>
</tr>
<tr>
<td>Bus</td>
<td>41</td>
</tr>
<tr>
<td>Bypass</td>
<td>41</td>
</tr>
<tr>
<td>Cabinets</td>
<td>42</td>
</tr>
<tr>
<td>Cancer</td>
<td>43</td>
</tr>
<tr>
<td>Carbon</td>
<td>44</td>
</tr>
<tr>
<td>Cat</td>
<td>45</td>
</tr>
<tr>
<td>Censored</td>
<td>45</td>
</tr>
<tr>
<td>Challeng</td>
<td>46</td>
</tr>
<tr>
<td>Chemist</td>
<td>47</td>
</tr>
<tr>
<td>Chesapea</td>
<td>48</td>
</tr>
<tr>
<td>Chevy</td>
<td>49</td>
</tr>
<tr>
<td>Chicken</td>
<td>50</td>
</tr>
</tbody>
</table>
R topics documented:

Chipavg ................................................................. 50
Chips ................................................................. 51
Cigar ................................................................. 52
Cigaret ............................................................... 53
CIsim ................................................................. 54
Citrus ................................................................. 55
Clean ................................................................. 56
Coaxial ................................................................. 57
Coffee ............................................................... 58
Coins ................................................................. 58
Combinations ........................................................ 59
Commute ............................................................. 60
Concept ............................................................... 61
Concrete ............................................................. 61
Corn ................................................................. 62
Correlat ............................................................... 63
Counsel ............................................................... 64
Cpi ................................................................. 64
Crime ................................................................. 65
Darwin ............................................................... 66
Dealers ............................................................... 67
Defectiv ............................................................... 67
Degree ............................................................... 68
Delay ................................................................. 69
Depend ............................................................... 70
Detroit ............................................................... 70
Develop ............................................................. 71
Devmath ............................................................. 72
Dice ................................................................. 72
Diesel ................................................................. 73
Diplomat ............................................................. 74
Disposal ............................................................. 75
Dogs ................................................................. 76
Domestic ........................................................... 77
Dopamine ........................................................... 78
Dowjones ........................................................... 79
Drink ................................................................. 80
Drug ................................................................. 80
Dyslexia .............................................................. 81
Earthqk ............................................................... 82
EDA ................................................................. 83
Educat ............................................................... 84
Eggs ................................................................. 85
Elderly ............................................................... 85
Energy ............................................................... 86
Engineer ............................................................. 87
Entrance ............................................................ 88
Epaminicompact ................................................... 89
R topics documented:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epatwoseater</td>
<td>90</td>
</tr>
<tr>
<td>Executiv</td>
<td>91</td>
</tr>
<tr>
<td>Exercise</td>
<td>91</td>
</tr>
<tr>
<td>Fabric</td>
<td>92</td>
</tr>
<tr>
<td>Faithful</td>
<td>93</td>
</tr>
<tr>
<td>Family</td>
<td>94</td>
</tr>
<tr>
<td>Ferraro1</td>
<td>95</td>
</tr>
<tr>
<td>Ferraro2</td>
<td>95</td>
</tr>
<tr>
<td>Fertility</td>
<td>96</td>
</tr>
<tr>
<td>Firstchi</td>
<td>97</td>
</tr>
<tr>
<td>Fish</td>
<td>98</td>
</tr>
<tr>
<td>Fitness</td>
<td>99</td>
</tr>
<tr>
<td>Florida2000</td>
<td>100</td>
</tr>
<tr>
<td>Fluid</td>
<td>101</td>
</tr>
<tr>
<td>Food</td>
<td>102</td>
</tr>
<tr>
<td>Framingh</td>
<td>102</td>
</tr>
<tr>
<td>Freshman</td>
<td>103</td>
</tr>
<tr>
<td>Funeral</td>
<td>104</td>
</tr>
<tr>
<td>Galaxie</td>
<td>105</td>
</tr>
<tr>
<td>Gallup</td>
<td>105</td>
</tr>
<tr>
<td>Gasoline</td>
<td>106</td>
</tr>
<tr>
<td>German</td>
<td>107</td>
</tr>
<tr>
<td>Golf</td>
<td>108</td>
</tr>
<tr>
<td>Governor</td>
<td>109</td>
</tr>
<tr>
<td>Gpa</td>
<td>110</td>
</tr>
<tr>
<td>Grades</td>
<td>111</td>
</tr>
<tr>
<td>Graduate</td>
<td>112</td>
</tr>
<tr>
<td>Greenriv</td>
<td>112</td>
</tr>
<tr>
<td>Grnriv2</td>
<td>113</td>
</tr>
<tr>
<td>Groupabc</td>
<td>114</td>
</tr>
<tr>
<td>Groups</td>
<td>114</td>
</tr>
<tr>
<td>Gym</td>
<td>115</td>
</tr>
<tr>
<td>Habits</td>
<td>116</td>
</tr>
<tr>
<td>Haptoglo</td>
<td>117</td>
</tr>
<tr>
<td>Hardware</td>
<td>117</td>
</tr>
<tr>
<td>Hardwood</td>
<td>118</td>
</tr>
<tr>
<td>Heat</td>
<td>119</td>
</tr>
<tr>
<td>Heating</td>
<td>120</td>
</tr>
<tr>
<td>Hodgkin</td>
<td>121</td>
</tr>
<tr>
<td>Homes</td>
<td>122</td>
</tr>
<tr>
<td>Homework</td>
<td>123</td>
</tr>
<tr>
<td>Honda</td>
<td>124</td>
</tr>
<tr>
<td>Hostile</td>
<td>124</td>
</tr>
<tr>
<td>Housing</td>
<td>125</td>
</tr>
<tr>
<td>Hurrican</td>
<td>126</td>
</tr>
<tr>
<td>Iceberg</td>
<td>127</td>
</tr>
<tr>
<td>Income</td>
<td>128</td>
</tr>
<tr>
<td>Independent</td>
<td>129</td>
</tr>
<tr>
<td>R topics documented:</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>---</td>
</tr>
<tr>
<td>Indian</td>
<td>130</td>
</tr>
<tr>
<td>Indiapol</td>
<td>131</td>
</tr>
<tr>
<td>Indy500</td>
<td>131</td>
</tr>
<tr>
<td>Inflatio</td>
<td>132</td>
</tr>
<tr>
<td>Inletoil</td>
<td>133</td>
</tr>
<tr>
<td>Inmate</td>
<td>134</td>
</tr>
<tr>
<td>Inspect</td>
<td>135</td>
</tr>
<tr>
<td>Insulate</td>
<td>136</td>
</tr>
<tr>
<td>Iqgpa</td>
<td>137</td>
</tr>
<tr>
<td>Iris</td>
<td>137</td>
</tr>
<tr>
<td>Jdpower</td>
<td>138</td>
</tr>
<tr>
<td>Jobsat</td>
<td>139</td>
</tr>
<tr>
<td>Kidsmoke</td>
<td>140</td>
</tr>
<tr>
<td>Kilowatt</td>
<td>141</td>
</tr>
<tr>
<td>Kinder</td>
<td>141</td>
</tr>
<tr>
<td>Laminect</td>
<td>142</td>
</tr>
<tr>
<td>Lead</td>
<td>143</td>
</tr>
<tr>
<td>Leader</td>
<td>144</td>
</tr>
<tr>
<td>Lethal</td>
<td>144</td>
</tr>
<tr>
<td>Life</td>
<td>145</td>
</tr>
<tr>
<td>Lifespan</td>
<td>146</td>
</tr>
<tr>
<td>Lightmonth</td>
<td>146</td>
</tr>
<tr>
<td>Lodge</td>
<td>147</td>
</tr>
<tr>
<td>Longtail</td>
<td>148</td>
</tr>
<tr>
<td>Lowabil</td>
<td>149</td>
</tr>
<tr>
<td>Magnesiu</td>
<td>149</td>
</tr>
<tr>
<td>Malpract</td>
<td>150</td>
</tr>
<tr>
<td>Manager</td>
<td>151</td>
</tr>
<tr>
<td>Marked</td>
<td>151</td>
</tr>
<tr>
<td>Math</td>
<td>152</td>
</tr>
<tr>
<td>Mathcomp</td>
<td>153</td>
</tr>
<tr>
<td>Mathpro</td>
<td>154</td>
</tr>
<tr>
<td>Maze</td>
<td>155</td>
</tr>
<tr>
<td>Median</td>
<td>155</td>
</tr>
<tr>
<td>Mental</td>
<td>156</td>
</tr>
<tr>
<td>Mercury</td>
<td>157</td>
</tr>
<tr>
<td>Metrent</td>
<td>157</td>
</tr>
<tr>
<td>Miller</td>
<td>158</td>
</tr>
<tr>
<td>Miller1</td>
<td>159</td>
</tr>
<tr>
<td>Moisture</td>
<td>159</td>
</tr>
<tr>
<td>Monoxide</td>
<td>160</td>
</tr>
<tr>
<td>Movie</td>
<td>161</td>
</tr>
<tr>
<td>Music</td>
<td>162</td>
</tr>
<tr>
<td>Name</td>
<td>163</td>
</tr>
<tr>
<td>Nascar</td>
<td>164</td>
</tr>
<tr>
<td>Nervous</td>
<td>164</td>
</tr>
<tr>
<td>Newsstand</td>
<td>165</td>
</tr>
<tr>
<td>Nfldraf2</td>
<td>166</td>
</tr>
</tbody>
</table>
R topics documented:

Nfldraft .......................................................... 166
Nicotine ............................................................ 167
normarea .......................................................... 168
nsizer .............................................................. 169
ntester .............................................................. 170
Orange .............................................................. 171
Orioles .............................................................. 171
Oxytocin ............................................................ 172
Parented ............................................................ 173
Patrol ............................................................... 174
Pearson .............................................................. 175
Phone ............................................................... 175
Poison .............................................................. 176
Politic ............................................................... 177
Pollutio .............................................................. 178
Porosity ............................................................. 178
Poverty ............................................................. 179
Precinct ............................................................. 180
Prejudic ............................................................. 181
Presiden ............................................................ 181
Press ............................................................... 182
Prognost ........................................................... 183
Program ............................................................ 184
Psat ................................................................. 184
Psych ............................................................... 185
Puerto ............................................................... 186
Quail ............................................................... 186
Quality ............................................................. 187
Rainks .............................................................. 188
Randd .............................................................. 189
Rat ................................................................. 189
Ratings ............................................................. 190
Reaction ........................................................... 191
Reading ............................................................ 192
Readiq ............................................................. 192
Referend .......................................................... 193
Region ............................................................ 194
Register .......................................................... 195
Rehab ............................................................. 195
Remedial .......................................................... 196
Rentals ............................................................. 197
Repair ............................................................. 198
Retail .............................................................. 198
Ronbrown1 ....................................................... 199
Ronbrown2 ....................................................... 200
Rural .............................................................. 200
Salary ............................................................. 201
Salinity ............................................................ 202
R topics documented:

Sat ................................................................. 203
Saving .......................................................... 204
Scales ............................................................ 204
Schizop2 .......................................................... 205
Schizoph .......................................................... 206
Seatbelt ........................................................... 207
Selfdefe .......................................................... 208
Senior ............................................................. 208
Sentence .......................................................... 209
Shkdrug ........................................................... 210
Shock .............................................................. 211
Shoplift ........................................................... 211
Short .............................................................. 212
Shuttle ............................................................ 213
SIGN.test ......................................................... 214
Simpson .......................................................... 216
Ssitup ............................................................. 217
Skewed ........................................................... 218
Skin ............................................................... 218
Slc ................................................................. 219
Smoklyph ......................................................... 220
Snore ............................................................. 221
Snow .............................................................. 222
Soccer ............................................................ 223
Social ............................................................. 223
Sophomor ......................................................... 224
South ............................................................. 224
Speed ............................................................. 225
Spellers .......................................................... 226
Spelling ........................................................... 227
Sports ............................................................. 227
Spouse ........................................................... 228
SRS ............................................................... 229
Stable ............................................................ 230
Stamp ............................................................ 230
Statclas ............................................................ 231
Statelaw ........................................................... 232
Statisti ............................................................ 232
Step ............................................................... 233
Stress ............................................................. 234
Study ............................................................. 235
Submarin .......................................................... 235
Subway ............................................................ 236
Sunspot ........................................................... 237
Superbowl .......................................................... 238
Supercar .......................................................... 238
Tablrock ........................................................... 239
Teacher ............................................................ 241
Abbey

Daily price returns (in pence) of Abbey National shares between 7/31/91 and 10/8/91

Data used in problem 6.39
Usage

Abbey

Format

A data frame/tibble with 50 observations on one variable

price daily price returns (in pence) of Abbey National shares

Source


References


Examples

```r
qqnorm(Abbey$price)
qqline(Abbey$price)
t.test(Abbey$price, mu = 300)
hist(Abbey$price, main = "Exercise 6.39",
    xlab = "daily price returns (in pence)",
    col = "blue")
```

---

Abc

*Three samples to illustrate analysis of variance*

Description

Data used in Exercise 10.1

Usage

Abc

Format

A data frame/tibble with 54 observations on two variables

response a numeric vector
group a character vector A, B, and C
References


Examples

```r
boxplot(response ~ group, col=c("red", "blue", "green"), data = Abc )
anova(lm(response ~ group, data = Abc))
```

---

### Abilene

**Crimes reported in Abilene, Texas**

---

**Description**

Data used in Exercise 1.23 and 2.79

**Usage**

Abilene

**Format**

A data frame/tibble with 16 observations on three variables

- **crimetype**: a character variable with values Aggravated assault, Arson, Burglary, Forcible rape, Larceny theft, Murder, Robbery, and Vehicle theft.
- **year**: a factor with levels 1992 and 1999
- **number**: number of reported crimes

**Source**

*Uniform Crime Reports*, US Dept. of Justice.

**References**

### Examples

```r
tab <- mtcars[, c(4, 6)]
tab$Model <- factor(tab$Model)

par(mfrow = c(2, 1))
barplot(tab$mpg[tab$Model == "Airflow"],
        names.arg = tab$manufacturer[tab$Model == "Airflow"],
        main = "1992 Crime Stats", col = "red")
barplot(tab$mpg[tab$Model == "Airflow"],
        names.arg = tab$manufacturer[tab$Model == "Airflow"],
        main = "1999 Crime Stats", col = "blue")
par(mfrow = c(1, 1))
```

```r
## Not run: library(ggplot2)
ggplot(data = tab, aes(x = model, y = mpg, fill = year)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 30, hjust = 1))
## End(Not run)
```

---

### Ability

**Perceived math ability for 13-year olds by gender**

### Description

Data used in Exercise 8.57

### Usage

```r
Ability
```

### Format

A data frame/tibble with 400 observations on two variables

- **gender** a factor with levels girls and boys
- **ability** a factor with levels hopeless, belowavg, average, aboveavg, and superior

### References

Abortion

Examples

CT <- xtabs(~gender + ability, data = Ability)
CT
chisq.test(CT)

Abortion rate by region of country

Description

Data used in Exercise 8.51

Usage

Abortion

Format

A data frame/tibble with 51 observations on the following 10 variables:

state  a character variable with values alabama, alaska, arizona, arkansas, california, colorado, connecticut, delaware, dist of columbia, florida, georgia, hawaii, idaho, illinois, indiana, iowa, kansas, kentucky, louisiana, maine, maryland, massachusetts, michigan, minnesota, mississippi, missouri, montana, nebraska, nevada, new hampshire, new jersey, new mexico, new york, north carolina, north dakota, ohio, oklahoma, oregon, pennsylvania, rhode island, south carolina, south dakota, tennessee, texas, utah, vermont, virginia, washington, west virginia, wisconsin, and wyoming

region  a character variable with values midwest northeast south west

regcode  a numeric vector

rate1988  a numeric vector

rate1992  a numeric vector

rate1996  a numeric vector

provide1988  a numeric vector

provide1992  a numeric vector

lowhigh  a numeric vector

rate  a factor with levels Low and High

References

Examples

```r
T1 <- xtabs(~region + rate, data = Abortion)
T1
chisq.test(T1)
```

---

<table>
<thead>
<tr>
<th>Absent</th>
<th>Number of absent days for 20 employees</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Description

Data used in Exercise 1.28

Usage

Absent

Format

A data frame/tibble with 20 observations on one variable

- **days** days absent

References


Examples

```r
CT <- xtabs(~ days, data = Absent)
CT
barplot(CT, col = "pink", main = "Exercise 1.28")
plot(ecdf(Absent$days), main = "ECDF")
```
Achieve

Math achievement test scores by gender for 25 high school students

Description

Data used in Example 7.14 and Exercise 10.7

Usage

Achieve

Format

A data frame/tibble with 25 observations on two variables

- **score**: mathematics achievement score
- **gender**: a factor with 2 levels boys and girls

References


Examples

```r
anova(lm(score ~ gender, data = Achieve))
t.test(score ~ gender, var.equal = TRUE, data = Achieve)
```

Adsales

Number of ads versus number of sales for a retailer of satellite dishes

Description

Data used in Exercise 9.15

Usage

Adsales

Format

A data frame/tibble with six observations on three variables

- **month**: a character vector listing month
- **ads**: a numeric vector containing number of ads
- **sales**: a numeric vector containing number of sales
References


Examples

```r
plot(sales ~ ads, data = Adsales, main = "Exercise 9.15")
mod <- lm(sales ~ ads, data = Adsales)
abline(mod, col = "red")
summary(mod)
predict(mod, newdata = data.frame(ads = 6), interval = "conf", level = 0.99)
```

<table>
<thead>
<tr>
<th>Aggress</th>
<th>Agressive tendency scores for a group of teenage members of a street gang</th>
</tr>
</thead>
</table>

Description

Data used in Exercises 1.66 and 1.81

Usage

Aggress

Format

A data frame/tibble with 28 observations on one variable

aggres measure of aggresive tendency, ranging from 10-50

References


Examples

```r
with(data = Aggress,
    EDA(aggres))
# OR
IQR(Aggress$aggres)
diff(range(Aggress$aggres))
```
Aid

Monthly payments per person for families in the AFDC federal program

Description

Data used in Exercises 1.91 and 3.68

Usage

Aid

Format

A data frame/tibble with 51 observations on two variables


payment average monthly payment per person in a family

Source


References


Examples

hist(Aid$payment, xlab = "payment", main = "Average monthly payment per person in a family", col = "lightblue")
boxplot(Aid$payment, col = "lightblue")
dotplot(state ~ payment, data = Aid)
**Description**

Data used in Exercise 6.60

**Usage**

Aids

**Format**

A data frame/tibble with 295 observations on three variables

- **duration** time (in months) from HIV infection to the clinical manifestation of full-blown AIDS
- **age** age (in years) of patient
- **group** a numeric vector

**Source**


**References**


**Examples**

```r
with(data = Aids, 
    EDA(duration)
  )
with(data = Aids, 
    t.test(duration, mu = 30, alternative = "greater")
  )
with(data = Aids, 
    SIGN.test(duration, md = 24, alternative = "greater")
  )
```
Airdisasters

Aircraft disasters in five different decades

Description

Data used in Exercise 1.12

Usage

Airdisasters

Format

A data frame /tibble with 141 observations on the following seven variables

- **year** a numeric vector indicating the year of an aircraft accident
- **deaths** a numeric vector indicating the number of deaths of an aircraft accident
- **decade** a character vector indicating the decade of an aircraft accident

Source


References


Examples

```r
par(las = 1)
stripchart(deaths ~ decade, data = Airdisasters,
       subset = decade != "1930s" & decade != "1940s",
       method = "stack", pch = 19, cex = 0.5, col = "red",
       main = "Aircraft Disasters 1950 - 1990",
       xlab = "Number of fatalities")
par(las = 0)
```
Airline

<table>
<thead>
<tr>
<th>Airline</th>
<th>Percentage of on-time arrivals and number of complaints for 11 airlines</th>
</tr>
</thead>
</table>

Description

Data for Example 2.9

Usage

Airline

Format

A data frame/tibble with 11 observations on three variables

- **airline** a character variable with values Alaska, Amer West, American, Continental, Delta, Northwest, Pan Am, Southwest, TWA, United, and USAir
- **ontime** a numeric vector
- **complaints** complaints per 1000 passengers

Source

Transportation Department.

References


Examples

```r
with(data = Airline,
    barplot(complaints, names.arg = airline, col = "lightblue",
            las = 2)
)
plot(complaints ~ ontime, data = Airline, pch = 19, col = "red",
     xlab = "On time", ylab = "Complaints")
```
**Alcohol**

*Ages at which 14 female alcoholics began drinking*

**Description**

Data used in Exercise 5.79

**Usage**

Alcohol

**Format**

A data frame/tibble with 14 observations on one variable

- **age** age when individual started drinking

**References**


**Examples**

```r
qqnorm(Alcohol$age)
qqline(Alcohol$age)
SIGN.test(Alcohol$age, md = 20, conf.level = 0.99)
```

---

**Allergy**

*Allergy medicines by adverse events*

**Description**

Data used in Exercise 8.22

**Usage**

Allergy

**Format**

A data frame/tibble with 406 observations on two variables

- **event** a factor with levels insomnia, headache, and drowsiness
- **medication** a factor with levels seldane-d, pseudoephedrine, and placebo
Source
Marion Merrel Dow, Inc. Kansas City, Mo. 64114.

References

Examples

```r
T1 <- xtabs(~event + medication, data = Allergy)
T1
chisq.test(T1)
```

---

### Anesthet

*Recovery times for anesthetized patients*

**Description**
Data used in Exercise 5.58

**Usage**

```r
Anesthet
cwith(data = Anesthet,
t.test(recover, conf.level = 0.90)$conf
)
```
Anxiety

Math test scores versus anxiety scores before the test

Description

Data used in Exercise 2.96

Usage

Anxiety

Format

A data frame/tibble with 20 observations on two variables

anxiety  anxiety score before a major math test
math  math test score

References


Examples

```r
plot(math ~ anxiety, data = Anxiety, ylab = "score",
main = "Exercise 2.96")
with(data = Anxiety,
cor(math, anxiety)
)
linmod <- lm(math ~ anxiety, data = Anxiety)
abline(linmod, col = "purple")
summary(linmod)
```

Apolipop

Level of apolipoprotein B and number of cups of coffee consumed per day for 15 adult males

Description

Data used in Examples 9.2 and 9.9

Usage

Apolipop
Append

Format

A data frame/tibble with 15 observations on two variables

- **coffee**: number of cups of coffee per day
- **apolipB**: level of apolipoprotein B

References


Examples

```r
plot(apolipB ~ coffee, data = Apolipop)
linmod <- lm(apolipB ~ coffee, data = Apolipop)
summary(linmod)
summary(linmod)$sigma
anova(linmod)
anova(linmod)[2, 3]^:.5
par(mfrow = c(2, 2))
plot(linmod)
par(mfrow = c(1, 1))
```

---

**Median costs of an appendectomy at 20 hospitals in North Carolina**

Description

Data for Exercise 1.119

Usage

Append

Format

A data frame/tibble with 20 observations on one variable

- **fee**: fees for an appendectomy for a random sample of 20 hospitals in North Carolina

Source

North Carolina Medical Database Commission, August 1994.

References

Examples

```r
fee <- Appendec$fee
ll <- mean(fee) - 2*sd(fee)
ul <- mean(fee) + 2*sd(fee)
limits <- c(ll, ul)
limits
fee[fee < ll | fee > ul]
```

Appendec Median costs of appendectomies at three different types of North Carolina hospitals

Description

Data for Exercise 10.60

Usage

Appendec

Format

A data frame/tibble with 59 observations on two variables

- **cost**: median costs of appendectomies at hospitals across the state of North Carolina in 1992
- **region**: a vector classifying each hospital as rural, regional, or metropolitan

Source


References


Examples

```r
boxplot(cost ~ region, data = Appendec, col = c("red", "blue", "cyan"))
anova(lm(cost ~ region, data = Appendec))
```
**Aptitude**

*Aptitude test scores versus productivity in a factory*

**Description**

Data for Exercises 2.1, 2.26, 2.35 and 2.51

**Usage**

Aptitude

**Format**

A data frame/tibble with 8 observations on two variables

- **aptitude**: aptitude test scores
- **product**: productivity scores

**References**


**Examples**

```r
plot(product ~ aptitude, data = Aptitude, main = "Exercise 2.1")
model1 <- lm(product ~ aptitude, data = Aptitude)
abline(model1, col = "red", lwd=3)
resid(model1)
fitted(model1)
cor(Aptitude$product, Aptitude$aptitude)
```

---

**Archaeo**

*Radiocarbon ages of observations taken from an archaeological site*

**Description**

Data for Exercises 5.120, 10.20 and Example 1.16

**Usage**

Archaeo
### Arthriti

**Description**

Data for Exercise 10.58

**Usage**

`Arthriti`

**Format**

A data frame/tibble with 60 observations on two variables

- **age**  number of years before 1983 - the year the data were obtained
- **phase** Ceramic Phase numbers

**Source**


**References**


**Examples**

```r
boxplot(age ~ phase, data = Archaeo, col = "yellow", 
       main = "Example 1.16", xlab = "Ceramic Phase", ylab = "Age")
anova(lm(age ~ as.factor(phase), data = Archaeo))
```

---

### Arthriti

**Time of relief for three treatments of arthritis**

**Description**

Data for Exercise 10.58

**Usage**

`Arthriti`

**Format**

A data frame/tibble with 51 observations on two variables

- **time** time (measured in days) until an arthritis sufferer experienced relief
- **treatment** a factor with levels A, B, and C

**References**

**Examples**

```r
boxplot(time ~ treatment, data = Arthritis, col = c("lightblue", "lightgreen", "yellow"), ylab = "days")
anova(lm(time ~ treatment, data = Arthritis))
```

---

**Artifici**

*Durations of operation for 15 artificial heart transplants*

**Description**

Data for Exercise 1.107

**Usage**

`Artifici`

**Format**

A data frame/tibble with 15 observations on one variable

- **duration**  duration (in hours) for transplant

**References**


**Examples**

```r
stem(Artifici$duration, 2)
summary(Artifici$duration)
values <- Artifici$duration[Artifici$duration < 6.5]
summary(values)
```
Asprin

Dissolving time versus level of impurities in aspirin tablets

Description
Data for Exercise 10.51

Usage
Asprin

Format
A data frame/tibble with 15 observations on two variables

time  time (in seconds) for aspirin to dissolve
impurity  impurity of an ingredient with levels 1%, 5%, and 10%

References

Examples

```r
boxplot(time ~ impurity, data = Asprin,
        col = c("red", "blue", "green"))
```

Asthmati

Asthmatic relief index on nine subjects given a drug and a placebo

Description
Data for Exercise 7.52

Usage
Asthmati

Format
A data frame/tibble with nine observations on three variables

drug  asthmatic relief index for patients given a drug
placebo  asthmatic relief index for patients given a placebo
difference  difference between the placebo and drug
References


Examples

```r
qqnorm(Asthmati$difference)
qqline(Asthmati$difference)
shapiro.test(Asthmati$difference)
with(data = Asthmati,
    t.test(placebo, drug, paired = TRUE, mu = 0, alternative = "greater")
)
```

---

Attorney

Number of convictions reported by U.S. attorney's offices

<table>
<thead>
<tr>
<th>Attorney</th>
<th>Number of convictions reported by U.S. attorney's offices</th>
</tr>
</thead>
</table>

Description

Data for Example 2.2 and Exercises 2.43 and 2.57

Usage

Attorney

Format

A data frame/tibble with 88 observations on three variables

**staff**  U.S. attorneys' office staff per 1 million population

**convict**  U.S. attorneys' office convictions per 1 million population

**district**  a factor with levels Albuquerque, Alexandria, Va, Anchorage, Asheville, NC, Atlanta, Baltimore, Baton Rouge, Billings, Mt. Birmingham, Al, Boise, Id, Boston, Buffalo, Burlington, Vt, Cedar Rapids, Charleston, WVA, Cheyenne, Wy, Chicago, Cincinnati, Cleveland, Columbia, SC, Concord, NH, Denver, Des Moines, Detroit, East St. Louis, Fargo, ND, Fort Smith, Ark, Fort Worth, Grand Rapids, Mi, Greensboro, NC, Honolulu, Houston, Indianapolis, Jackson, Miss, Kansas City, Knoxville, Tn, Las Vegas, Lexington, Ky, Little Rock, Los Angeles, Louisville, Memphis, Miami, Milwaukee, Minneapolis, Mobile, Ala, Montgomery, Ala, Muskogee, Ok, Nashville, New Haven, Conn, New Orleans, New York (Brooklyn), New York (Manhattan), Newark, NJ, Oklahoma City, Omaha, Oxford, Miss, Pensacola, Fl, Philadelphia, Phoenix, Pittsburgh, Portland, Maine, Portland, Ore, Providence, RI, Raleigh, NC, Roanoke, Va, Sacramento, Salt Lake City, San Antonio, San Diego, San Francisco, Savannah, Ga, Scranton, Pa, Seattle, Shreveport, La, Sioux Falls, SD, South Bend, Ind, Spokane, Wash. Springfield, Ill, St. Louis, Syracuse, NY, Tampa, Topeka, Kan, Tulsa, Tyler, Tex, Washington, Wheeling, WVa, and Wilmington, Del
Autogear

References


Examples

par(mfrow=c(1, 2))
plot(convict ~ staff, data = Attorney, main = "With Washington, D.C.")
plot(convict[-86] ~ staff[-86], data = Attorney,
main = "Without Washington, D.C.")
par(mfrow=c(1, 1))

<table>
<thead>
<tr>
<th>Autogear</th>
<th>Number of defective auto gears produced by two manufacturers</th>
</tr>
</thead>
</table>

Description

Data for Exercise 7.46

Usage

Autogear

Format

A data frame/tibble with 20 observations on two variables

defectives number of defective gears in the production of 100 gears per day
manufacturer a factor with levels A and B

References


Examples

t.test(defectives ~ manufacturer, data = Autogear)
wilcox.test(defectives ~ manufacturer, data = Autogear)
t.test(defectives ~ manufacturer, var.equal = TRUE, data = Autogear)
Backtoback

*Illustrates inferences based on pooled t-test versus Wilcoxon rank sum test*

**Description**

Data for Exercise 7.40

**Usage**

Backtoback

**Format**

A data frame/tibble with 24 observations on two variables

- **score** a numeric vector
- **group** a numeric vector

**References**


**Examples**

```r
wilcox.test(score ~ group, data = Backtoback)
t.test(score ~ group, data = Backtoback)
```

Bbsalaries

*Baseball salaries for members of five major league teams*

**Description**

Data for Exercise 1.11

**Usage**

Bbsalaries

**Format**

A data frame/tibble with 142 observations on two variables

- **salary** 1999 salary for baseball player
- **team** a factor with levels Angels, Indians, Orioles, Redsoxs, and Whitesoxs
References


Examples

```r
stripchart(salary ~ team, data = Bbsalaries, method = "stack",
           pch = 19, col = "blue", cex = 0.75)
title(main = "Major League Salaries")
```

Bigten

*Graduation rates for student athletes and nonathletes in the Big Ten Conf.*

Description

Data for Exercises 1.124 and 2.94

Usage

```r
Bigten
```

Format

A data frame/tibble with 44 observations on the following four variables

- **school**: a factor with levels Illinois, Indiana, Iowa, Michigan, Michigan State, Minnesota, Northwestern, Ohio State, Penn State, Purdue, and Wisconsin
- **rate**: graduation rate
- **year**: factor with two levels 1984-1985 and 1993-1994
- **status**: factor with two levels athlete and student

Source


References

Examples

```r
boxplot(rate ~ status, data = subset(Bigten, year = "1993-1994"),
        horizontal = TRUE, main = "Graduation Rates 1993-1994")
with(data = Bigten,
    tapply(rate, list(year, status), mean)
)
```

<table>
<thead>
<tr>
<th>Biology</th>
<th>Test scores on first exam in biology class</th>
</tr>
</thead>
</table>

Description

Data for Exercise 1.49

Usage

Biology

Format

A data frame/tibble with 30 observations on one variable

`score` test scores on the first test in a beginning biology class

References


Examples

```r
hist(Biology$score, breaks = "scott", col = "brown", freq = FALSE,
    main = "Problem 1.49", xlab = "Test Score")
lines(density(Biology$score), lwd=3)
```
Birth

Live birth rates in 1990 and 1998 for all states

Description

Data for Example 1.10

Usage

Birth

Format

A data frame/tibble with 51 observations on three variables


- **rate**: live birth rates per 1000 population

- **year**: a factor with levels 1990 and 1998

Source


References


Examples

```r
rate1998 <- subset(Birth, year == "1998", select = rate)
stem(x = rate1998$rate, scale = 2)
hist(rate1998$rate, breaks = seq(10.9, 21.9, 1.0), xlab = "1998 Birth Rate",
    main = "Figure 1.14 in BSDA", col = "pink")
hist(rate1998$rate, breaks = seq(10.9, 21.9, 1.0), xlab = "1998 Birth Rate",
    main = "Figure 1.16 in BSDA", col = "pink", freq = FALSE)
lines(density(rate1998$rate), lwd = 3)
rm(rate1998)
```
### Blackedu

**Education level of blacks by gender**

#### Description

Data for Exercise 8.55

#### Usage

Blackedu

#### Format

A data frame/tibble with 3800 observations on two variables

- **gender** a factor with levels Female and Male
- **education** a factor with levels High school dropout, High school graduate, Some college, Bachelor’s degree, and Graduate degree

#### Source

Bureau of Census data.

#### References


#### Examples

```
T1 <- xtabs(~gender + education, data = Blackedu)
T1
chisq.test(T1)
```

---

### Blood

**Blood pressure of 15 adult males taken by machine and by an expert**

#### Description

Data for Exercise 7.84

#### Usage

Blood
Format

A data frame/tibble with 15 observations on the following two variables

- **machine** blood pressure recorded from an automated blood pressure machine
- **expert** blood pressure recorded by an expert using an at-home device

References


Examples

```r
DIFF <- Blood$machine - Blood$expert
shapiro.test(DIFF)
qqnorm(DIFF)
qqline(DIFF)
rm(DIFF)
t.test(Blood$machine, Blood$expert, paired = TRUE)
```

---

<table>
<thead>
<tr>
<th>Board</th>
<th>Incomes of board members from three different universities</th>
</tr>
</thead>
</table>

Description

Data for Exercise 10.14

Usage

Board

Format

A data frame/tibble with 7 observations on three variables

- **salary** 1999 salary (in $1000) for board directors
- **university** a factor with levels A, B, and C

References

**Bones**

**Examples**

```r
boxplot(salary ~ university, data = Board, col = c("red", "blue", "green"),
       ylab = "Income")
tapply(Board$salary, Board$university, summary)
anova(lm(salary ~ university, data = Board))
## Not run:
library(dplyr)
dplyr::group_by(Board, university) %>%
  summarize(Average = mean(salary))
## End(Not run)
```

**Bones**

**Bone density measurements of 35 physically active and 35 non-active women**

**Description**

Data for Example 7.22

**Usage**

Bones

**Format**

A data frame/tibble with 70 observations on two variables

- **density** bone density measurements
- **group** a factor with levels **active** and **nonactive**

**References**


**Examples**

```r
t.test(density ~ group, data = Bones, alternative = "greater")
t.test(rank(density) ~ group, data = Bones, alternative = "greater")
wilcox.test(density ~ group, data = Bones, alternative = "greater")
```
Number of books read and final spelling scores for 17 third graders

Description
Data for Exercise 9.53

Usage
Books

Format
A data frame/tibble with 17 observations on two variables

book  number of books read
spelling  spelling score

References

Examples

plot(spelling ~ book, data = Books)
mod <- lm(spelling ~ book, data = Books)
summary(mod)
abline(mod, col = "blue", lwd = 2)

Prices paid for used books at three different bookstores

Description
Data for Exercise 10.30 and 10.31

Usage
Bookstor
**Brain**

**Format**

A data frame/tibble with 72 observations on two variables

- **dollars** money obtained for selling textbooks
- **store** a factor with levels A, B, and C

**References**


**Examples**

```r
boxplot(dollars ~ store, data = Bookstor,
col = c("purple", "lightblue", "cyan"))
kruskal.test(dollars ~ store, data = Bookstor)
```

<table>
<thead>
<tr>
<th>Brain</th>
<th><strong>Brain weight versus body weight of 28 animals</strong></th>
</tr>
</thead>
</table>

**Description**

Data for Exercises 2.15, 2.44, 2.58 and Examples 2.3 and 2.20

**Usage**

Brain

**Format**

A data frame/tibble with 28 observations on three variables

- **species** a factor with levels African elephant, Asian Elephant, Brachiosaurus, Cat, Chimpanzee, Cow, Diplodocus, Donkey, Giraffe, Goat, Gorilla, Gray wolf, Guinea Pig, Hamster, Horse, Human, Jaguar, Kangaroo, Mole, Mouse, Mt Beaver, Pig, Potar monkey, Rabbit, Rat, Rhesus monkey, Sheep, and Triceratops
- **bodyweight** body weight (in kg)
- **brainweight** brain weight (in g)

**Source**


**References**

Examples

```r
plot(log(brainweight) ~ log(bodyweight), data = Brain,
     pch = 19, col = "blue", main = "Example 2.3")
mod <- lm(log(brainweight) ~ log(bodyweight), data = Brain)
abline(mod, lty = "dashed", col = "blue")
```

---

## Bumpers

*Repair costs of vehicles crashed into a barrier at 5 miles per hour*

### Description

Data for Exercise 1.73

### Usage

Bumpers

### Format

A data frame/tibble with 23 observations on two variables

- **car** a factor with levels Buick Century, Buick Skylark, Chevrolet Cavalier, Chevrolet Corsica, Chevrolet Lumina, Dodge Dynasty, Dodge Monaco, Ford Taurus, Ford Tempo, Honda Accord, Hyundai Sonata, Mazda 626, Mitsubishi Galant, Nissan Stanza, Oldsmobile Calais, Oldsmobile Ciera, Plymouth Acclaim, Pontiac 6000, Pontiac Grand Am, Pontiac Sunbird, Saturn SL2, Subaru Legacy, and Toyota Camry

- **repair** total repair cost (in dollars) after crashing a car into a barrier four times while the car was traveling at 5 miles per hour

### Source

Insurance Institute of Highway Safety.

### References


### Examples

```r
EDA(Bumpers$repair)
stripchart(Bumpers$repair, method = "stack", pch = 19, col = "blue")
library(lattice)
dotplot(car ~ repair, data = Bumpers)
```
Bus

Attendance of bus drivers versus shift

Description
Data for Exercise 8.25

Usage
Bus

Format
A data frame/tibble with 29363 observations on two variables

- **attendance** a factor with levels absent and present
- **shift** a factor with levels am, noon, pm, swing, and split

References

Examples

```r
T1 <- xtabs(~attendance + shift, data = Bus)
T1
chisq.test(T1)
```

Bypass
Median charges for coronary bypass at 17 hospitals in North Carolina

Description
Data for Exercises 5.104 and 6.43

Usage
Bypass
Cabinets

**Format**
A data frame/tibble with 17 observations on two variables

- **hospital**: a factor with levels Carolinas Med Ct, Duke Med Ct, Durham Regional, Forsyth Memorial, Frye Regional, High Point Regional, Memorial Mission, Mercy, Moore Regional, Moses Cone Memorial, NC Baptist, New Hanover Regional, Pitt Co. Memorial, Presbyterian, Rex, Univ of North Carolina, and Wake County

- **charge**: median charge for coronary bypass

**Source**

**References**

**Examples**

```r
EDA(Bypass$charge)
t.test(Bypass$charge, conf.level=.90)$conf
t.test(Bypass$charge, mu = 35000)
```

---

Cabinets

**Estimates of costs of kitchen cabinets by two suppliers on 20 prospective homes**

**Description**
Data for Exercise 7.83

**Usage**
Cabinets

**Format**
A data frame/tibble with 20 observations on three variables

- **home**: a numeric vector
- **supplA**: estimate for kitchen cabinets from supplier A (in dollars)
- **supplB**: estimate for kitchen cabinets from supplier A (in dollars)
Cancer Survival times of terminal cancer patients treated with vitamin C

Description

Data for Exercises 6.55 and 6.64

Usage

Cancer

Format

A data frame/tibble with 64 observations on two variables

<table>
<thead>
<tr>
<th>survival</th>
<th>survival time (in days) of terminal patients treated with vitamin C</th>
</tr>
</thead>
<tbody>
<tr>
<td>type</td>
<td>a factor indicating type of cancer with levels breast, bronchus, colon, ovary, and stomach</td>
</tr>
</tbody>
</table>

Source


References

Examples

```r
boxplot(survival ~ type, Cancer, col = "blue")
stomach <- Cancer$survival[Cancer$type == "stomach"]
bronchus <- Cancer$survival[Cancer$type == "bronchus"]
boxplot(stomach, ylab = "Days")
SIGN.test(stomach, md = 100, alternative = "greater")
SIGN.test(bronchus, md = 100, alternative = "greater")
rm(bronchus, stomach)
```

Carbon

Carbon monoxide level measured at three industrial sites

Description

Data for Exercise 10.28 and 10.29

Usage

Carbon

Format

A data frame/tibble with 24 observations on two variables

- **CO** carbon monoxide measured (in parts per million)
- **site** a factor with levels SiteA, SiteB, and SiteC

References


Examples

```r
boxplot(CO ~ site, data = Carbon, col = "lightgreen")
kruskal.test(CO ~ site, data = Carbon)
```
cat

Reading scores on the California achievement test for a group of 3rd graders

Description
Data for Exercise 1.116

Usage
Cat

Format
A data frame/tibble with 17 observations on one variable

score reading score on the California Achievement Test

References

Examples
stem(Cat$score)
fivenum(Cat$score)
boxplot(Cat$score, main = "Problem 1.116", col = "green")

Censored
Entry age and survival time of patients with small cell lung cancer under two different treatments

Description
Data for Exercises 7.34 and 7.48

Usage
Censored
Format

A data frame/tibble with 121 observations on three variables

**survival**  survival time (in days) of patients with small cell lung cancer

**treatment** a factor with levels armA and armB indicating the treatment a patient received

**age**  the age of the patient

Source


References


Examples

```r
boxplot(survival ~ treatment, data = Censored, col = "yellow")
wilcox.test(survival ~ treatment, data = Censored, alternative = "greater")
```

Challeng

Temperatures and O-ring failures for the launches of the space shuttle Challenger

Description

Data for Examples 1.11, 1.12, 1.13, 2.11 and 5.1

Usage

Challeng

Format

A data frame/tibble with 25 observations on four variables

**flight**  a character variable indicating the flight

**date**  date of the flight

**temp**  temperature (in fahrenheit)

**failures**  number of failures
Source


References


Examples

```r
stem(Challeng$temp)
summary(Challeng$temp)
IQR(Challeng$temp)
quantile(Challeng$temp)
fivenum(Challeng$temp)

stem(sort(Challeng$temp)[-1])
summary(sort(Challeng$temp)[-1])
IQR(sort(Challeng$temp)[-1])
quantile(sort(Challeng$temp)[-1])
fivenum(sort(Challeng$temp)[-1])

par(mfrow=c(1, 2))
qqnorm(Challeng$temp)
qqline(Challeng$temp)

par(mfrow=c(1, 1))
```

---

**Chemist**

*Starting salaries of 50 chemistry majors*

Description

Data for Example 5.3

Usage

Chemist

Format

A data frame/tibble with 50 observations on one variable

**salary** starting salary (in dollars) for chemistry major
References


Examples

```r
EDA(Chemist$salary)
```

---

Chesapeake

*Surface salinity measurements taken offshore from Annapolis, Maryland in 1927*

Description

Data for Exercise 6.41

Usage

Chesapeake

Format

A data frame/tibble with 16 observations on one variable

- **salinity**: surface salinity measurements (in parts per 1000) for station 11, offshore from Annapolis, Maryland, on July 3-4, 1927.

Source


References


Examples

```r
qqnorm(Chesapeake$salinity)
qqline(Chesapeake$salinity)
shapiro.test(Chesapeake$salinity)
t.test(Chesapeake$salinity, mu = 7)
```
Description

Data for Exercise 8.35

Usage

Chevy

Format

A data frame/tibble with 67 observations on two variables

- **year**: a factor with levels 1988-90 and 1991-93
- **frequency**: a factor with levels much better than average, above average, average, below average, and much worse than average

Source

Insurance Institute for Highway Safety and the Highway Loss Data Institute, 1995.

References


Examples

```r
T1 <- xtabs(~year + frequency, data = Chevy)
T1
chisq.test(T1)
rm(T1)
```
### Chicken

*Weight gain of chickens fed three different rations*

**Description**

Data for Exercise 10.15

**Usage**

Chicken

**Format**

A data frame/tibble with 13 observations on three variables

- **gain**: weight gain over a specified period
- **feed**: a factor with levels *ration1*, *ration2*, and *ration3*

**References**


**Examples**

```r
boxplot(gain ~ feed, col = c("red","blue","green"), data = Chicken)
anova(lm(gain ~ feed, data = Chicken))
```

### Chipavg

*Measurements of the thickness of the oxide layer of manufactured integrated circuits*

**Description**

Data for Exercises 6.49 and 7.47

**Usage**

Chipavg
Chips

**Format**

A data frame/tibble with 30 observations on three variables

- **wafer1** thickness of the oxide layer for wafer1
- **wafer2** thickness of the oxide layer for wafer2
- **thickness** average thickness of the oxide layer of the eight measurements obtained from each set of two wafers

**Source**


**References**


**Examples**

```r
EDA(Chipavg$thickness)
t.test(Chipavg$thickness, mu = 1000)
boxplot(Chipavg$wafer1, Chipavg$wafer2, name = c("Wafer 1", "Wafer 2"))
shapiro.test(Chipavg$wafer1)
shapiro.test(Chipavg$wafer2)
t.test(Chipavg$wafer1, Chipavg$wafer2, var.equal = TRUE)
```

---

**Description**

Data for Exercise 10.9

**Usage**

Chips
Format

A data frame/tibble with 30 observations on eight variables

- wafer11: first measurement of thickness of the oxide layer for wafer1
- wafer12: second measurement of thickness of the oxide layer for wafer1
- wafer13: third measurement of thickness of the oxide layer for wafer1
- wafer14: fourth measurement of thickness of the oxide layer for wafer1
- wafer21: first measurement of thickness of the oxide layer for wafer2
- wafer22: second measurement of thickness of the oxide layer for wafer2
- wafer23: third measurement of thickness of the oxide layer for wafer2
- wafer24: fourth measurement of thickness of the oxide layer for wafer2

Source


References


Examples

```r
with(data = Chips,
    boxplot(wafer11, wafer12, wafer13, wafer14, wafer21,
            wafer22, wafer23, wafer24, col = "pink")
)
```

<table>
<thead>
<tr>
<th>Cigar</th>
<th>Description</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milligrams of tar in 25 cigarettes selected randomly from 4 different brands</td>
<td>Data for Example 10.4</td>
<td>Cigar</td>
</tr>
</tbody>
</table>
Cigarettes

**Description**

Data for Exercise 2.27

**Usage**

Cigarettes

**Format**

A data frame/tibble with 100 observations on two variables

- **tar** amount of tar (measured in milligrams)
- **brand** a factor indicating cigarette brand with levels `brandA`, `brandB`, `brandC`, and `brandD`

**References**


**Examples**

```r
boxplot(tar ~ brand, data = Cigars, col = "cyan", ylab = "mg tar")
anova(lm(tar ~ brand, data = Cigars))
```

<table>
<thead>
<tr>
<th>Cigarettes</th>
<th>Effect of mother's smoking on birth weight of newborn</th>
</tr>
</thead>
</table>

**Cigarette**

**Description**

Data for Exercise 2.27

**Usage**

Cigarette

**Format**

A data frame/tibble with 16 observations on two variables

- **cigarettes** mothers’ estimated average number of cigarettes smoked per day
- **weight** children’s birth weights (in pounds)

**References**

Examples

```r
plot(weight ~ cigarettes, data = Cigarett)
model <- lm(weight ~ cigarettes, data = Cigarett)
abline(model, col = "red")
with(data = Cigarett,
    cor(weight, cigarettes)
)
rm(model)
```

CIsim

**Confidence Interval Simulation Program**

Description

This program simulates random samples from which it constructs confidence intervals for one of the parameters mean (Mu), variance (Sigma), or proportion of successes (Pi).

Usage

```r
CIsim(  
    samples = 100,
    n = 30,
    mu = 0,
    sigma = 1,
    conf.level = 0.95,
    type = "Mean"
)
```

Arguments

- `samples` the number of samples desired.
- `n` the size of each sample.
- `mu` if constructing confidence intervals for the population mean or the population variance, mu is the population mean (i.e., type is one of either "Mean", or "Var"). If constructing confidence intervals for the population proportion of successes, the value entered for mu represents the population proportion of successes (Pi), and as such, must be a number between 0 and 1.
- `sigma` the population standard deviation. sigma is not required if confidence intervals are of type "Pi".
- `conf.level` confidence level for the graphed confidence intervals, restricted to lie between zero and one.
- `type` character string, one of "Mean", "Var" or "Pi", or just the initial letter of each, indicating the type of confidence interval simulation to perform.
Details

Default is to construct confidence intervals for the population mean. Simulated confidence intervals for the population variance or population proportion of successes are possible by selecting the appropriate value in the type argument.

Value

Graph depicts simulated confidence intervals. The number of confidence intervals that do not contain the parameter of interest are counted and reported in the commands window.

Author(s)

Alan T. Arnholt

Examples

CIsim(100, 30, 100, 10)
    # Simulates 100 samples of size 30 from
    # a normal distribution with mean 100
    # and standard deviation 10. From the
    # 100 simulated samples, 95% confidence
    # intervals for the Mean are constructed
    # and depicted in the graph.

CIsim(100, 30, 100, 10, type="Var")
    # Simulates 100 samples of size 30 from
    # a normal distribution with mean 100
    # and standard deviation 10. From the
    # 100 simulated samples, 95% confidence
    # intervals for the variance are constructed
    # and depicted in the graph.

CIsim(100, 50, .5, type="Pi", conf.level=.90)
    # Simulates 100 samples of size 50 from
    # a binomial distribution where the population
    # proportion of successes is 0.5. From the
    # 100 simulated samples, 90% confidence
    # intervals for Pi are constructed
    # and depicted in the graph.

Citrus

Percent of peak bone density of different aged children

Description

Data for Exercise 9.7
Clean

Usage

Citrus

Format

A data frame/tibble with nine observations on two variables

age  age of children
percent  percent peak bone density

References


Examples

```r
model <- lm(percent ~ age, data = Citrus)
summary(model)
anova(model)
rm(model)
```

---

Clean  *Residual contaminant following the use of three different cleansing agents*

Description

Data for Exercise 10.16

Usage

Clean

Format

A data frame/tibble with 45 observations on two variables

clean  residual contaminants
agent  a factor with levels A, B, and C

References

Coaxial

**Examples**

```r
boxplot(clean ~ agent, col = c("red", "blue", "green"), data = Clean) 
anova(lm(clean ~ agent, data = Clean))
```

---

**Coaxial**  
*Signal loss from three types of coaxial cable*

**Description**

Data for Exercise 10.24 and 10.25

**Usage**

Coaxial

**Format**

A data frame/tibble with 45 observations on two variables

- **signal** signal loss per 1000 feet
- **cable** factor with three levels of coaxial cable typeA, typeB, and typeC

**References**


**Examples**

```r
boxplot(signal ~ cable, data = Coaxial, col = c("red", "green", "yellow"))
kruskal.test(signal ~ cable, data = Coaxial)
```
Coffee  

*Productivity of workers with and without a coffee break*

**Description**

Data for Exercise 7.55

**Usage**

Coffee

**Format**

A data frame/tibble with nine observations on three variables

- **without** workers’ productivity scores without a coffee break
- **with** workers’ productivity scores with a coffee break
- **differences** with minus without

**References**


**Examples**

```r
qqnorm(Coffee$differences)
normplot(Coffee$differences)
shapiro.test(Coffee$differences)
t.test(Coffee$with, Coffee$without, paired = TRUE, alternative = "greater")
wilcox.test(Coffee$with, Coffee$without, paired = TRUE, alternative = "greater")
```

---

Coins  

*Yearly returns on 12 investments*

**Description**

Data for Exercise 5.68

**Usage**

Coins
Combinations

Format

A data frame/tibble with 12 observations on one variable

**return** yearly returns on each of 12 possible investments

References


Examples

```r
qqnorm(Coins$return)
qqline(Coins$return)
```

---

**Description**

Computes all possible combinations of \( n \) objects taken \( k \) at a time.

**Usage**

`Combinations(n, k)`

**Arguments**

- \( n \) a number.
- \( k \) a number less than or equal to \( n \).

**Value**

Returns a matrix containing the possible combinations of \( n \) objects taken \( k \) at a time.

**See Also**

`SRS`

**Examples**

```r
Combinations(5, 2)
# The columns in the matrix list the values of the 10 possible
# combinations of 5 things taken 2 at a time.
```
**Commute**

*Commuting times for selected cities in 1980 and 1990*

**Description**

Data for Exercises 1.13, and 7.85

**Usage**

`Commute`

**Format**

A data frame/tibble with 39 observations on three variables

- **city** a factor with levels Atlanta, Baltimore, Boston, Buffalo, Charlotte, Chicago, Cincinnati, Cleveland, Columbus, Dallas, Denver, Detroit, Hartford, Houston, Indianapolis, Kansas City, Los Angeles, Miami, Milwaukee, Minneapolis, New Orleans, New York, Norfolk, Orlando, Philadelphia, Phoenix, Pittsburgh, Portland, Providence, Rochester, Sacramento, Salt Lake City, San Antonio, San Diego, San Francisco, Seattle, St. Louis, Tampa, and Washington

- **year** year

- **time** commute times

**Source**

Federal Highway Administration.

**References**


**Examples**

```r
stripplot(year ~ time, data = Commute, jitter = TRUE)
dotplot(year ~ time, data = Commute)
bwplot(year ~ time, data = Commute)
stripchart(time ~ year, data = Commute, method = "stack", pch = 1, cex = 2, col = c("red", "blue"),
group.names = c("1980", "1990"),
main = "", xlab = "minutes")
title(main = "Commute Time")
boxplot(time ~ year, data = Commute, names=c("1980", "1990"), horizontal = TRUE, las = 1)
```
Concept

Tennessee self concept scale scores for a group of teenage boys

Description

Data for Exercise 1.68 and 1.82

Usage

Concept

Format

A data frame/tibble with 28 observations on one variable

self Tennessee self concept scores

References


Examples

```r
summary(Concept$self)
sd(Concept$self)
diff(range(Concept$self))
IQR(Concept$self)
summary(Concept$self/10)
IQR(Concept$self/10)
sd(Concept$self/10)
diff(range(Concept$self/10))
```

Concrete

Compressive strength of concrete blocks made by two different methods

Description

Data for Example 7.17

Usage

Concrete
Format

A data frame/tibble with 20 observations on two variables

strength comprehensive strength (in pounds per square inch)
method factor with levels new and old indicating the method used to construct a concrete block

References


Examples

```r
wilcox.test(strength ~ method, data = Concrete, alternative = "greater")
```

---

**Corn**

*Comparison of the yields of a new variety and a standard variety of corn planted on 12 plots of land*

Description

Data for Exercise 7.77

Usage

Corn

Format

A data frame/tibble with 12 observations on three variables

new corn yield with new method
standard corn yield with standard method
differences new minus standard

References

**Examples**

```r
boxplot(Corn$differences)
qunorm(Corn$differences)
qqline(Corn$differences)
shapiro.test(Corn$differences)
t.test(Corn$new, Corn$standard, paired = TRUE, alternative = "greater")
```

**Correlat**

*Exercise to illustrate correlation*

**Description**

Data for Exercise 2.23

**Usage**

```r
Correlat
```

**Format**

A data frame/tibble with 13 observations on two variables

- `x` a numeric vector
- `y` a numeric vector

**References**


**Examples**

```r
plot(y ~ x, data = Correlat)
model <- lm(y ~ x, data = Correlat)
abline(model)
rm(model)
```
### Counsel

**Scores of 18 volunteers who participated in a counseling process**

**Description**

Data for Exercise 6.96

**Usage**

Counsel

**Format**

A data frame/tibble with 18 observations on one variable

- **score**: standardized psychology scores after a counseling process

**References**


**Examples**

```r
EDA(Counsel$score)
t.test(Counsel$score, mu = 70)
```

### Cpi

**Consumer price index from 1979 to 1998**

**Description**

Data for Exercise 1.34

**Usage**

Cpi

**Format**

A data frame/tibble with 20 observations on two variables

- **year**: year
- **cpi**: consumer price index
Crime

Source


References


Examples

```r
plot(cpi ~ year, data = Cpi, type = "l", lty = 2, lwd = 2, col = "red")
barplot(Cpi$cpi, col = "pink", las = 2, main = "Problem 1.34")
```

<table>
<thead>
<tr>
<th>Crime</th>
<th>Violent crime rates for the states in 1983 and 1993</th>
</tr>
</thead>
</table>

Description

Data for Exercises 1.90, 2.32, 3.64, and 5.113

Usage

Crime

Format

A data frame/tibble with 102 observations on three variables

- **year** a factor with levels 1983 and 1993
- **rate** crime rate per 100,000 inhabitants

Source

References


Examples

```r
boxplot(rate ~ year, data = Crime, col = "red")
```

---

**Darwin**

*Charles Darwin’s study of cross-fertilized and self-fertilized plants*

Description

Data for Exercise 7.62

Usage

Darwin

Format

A data frame/tibble with 15 observations on three variables

- **pot** number of pot
- **cross** height of plant (in inches) after a fixed period of time when cross-fertilized
- **self** height of plant (in inches) after a fixed period of time when self-fertilized

Source


References


Examples

```r
differ <- Darwin$cross - Darwin$self
dqnorm(differ)
dqline(differ)
shapiro.test(differ)
wilcox.test(Darwin$cross, Darwin$self, paired = TRUE)
rm(differ)
```
Dealers

Automobile dealers classified according to type dealership and service rendered to customers

Description
Data for Example 2.22

Usage
Dealers

Format
A data frame/tibble with 122 observations on two variables

- **type** a factor with levels Honda, Toyota, Mazda, Ford, Dodge, and Saturn
- **service** a factor with levels Replaces unnecessarily and Follows manufacturer guidelines

References

Examples

```r
xtabs(~type + service, data = Dealers)
T1 <- xtabs(~type + service, data = Dealers)
T1
addmargins(T1)
pt <- prop.table(T1, margin = 1)
pt
barplot(t(pt), col = c("red", "skyblue"), legend = colnames(T1))
rm(T1, pt)
```

Defectiv

Number of defective items produced by 20 employees

Description
Data for Exercise 1.27

Usage
Defectiv
Format

A data frame/tibble with 20 observations on one variable

number  number of defective items produced by the employees in a small business firm

References


Examples

T1 <- xtabs(~ number, data = Defectiv)
T1
barplot(T1, col = "pink", ylab = "Frequency",
xlab = "Defective Items Produced by Employees", main = "Problem 1.27")
rm(T1)

---

<table>
<thead>
<tr>
<th>Degree</th>
<th>Percent of bachelor's degrees awarded women in 1970 versus 1990</th>
</tr>
</thead>
</table>

Description

Data for Exercise 2.75

Usage

Degree

Format

A data frame/tibble with 1064 observations on two variables

field  a factor with levels Health, Education, Foreign Language, Psychology, Fine Arts, Life Sciences, Business, Social Science, Physical Sciences, Engineering, and All Fields

awarded  a factor with levels 1970 and 1990

Source

U.S. Department of Health and Human Services, National Center for Education Statistics.

References

**Examples**

```r
t1 <- xtabs(~field + awarded, data = Degree)
t1
barplot(t(T1), beside = TRUE, col = c("red", "skyblue"), legend = colnames(T1))
rm(T1)
```

---

**Delay**

*Delay times on 20 flights from four major air carriers*

---

**Description**

Data for Exercise 10.55

**Usage**

`Delay`

**Format**

A data frame/tibble with 80 observations on two variables

- **delay** the delay time (in minutes) for 80 randomly selected flights
- **carrier** a factor with levels A, B, C, and D

**References**


**Examples**

```r
boxplot(delay ~ carrier, data = Delay,
       main = "Exercise 10.55", ylab = "minutes",
       col = "pink")
kruskal.test(delay ~ carrier, data = Delay)
```
### Deargent

**Number of dependent children for 50 families**

**Description**

Data for Exercise 1.26

**Usage**

Depend

**Format**

A data frame/tibble with 50 observations on one variable

- **number**: number of dependent children in a family

**References**


**Examples**

```r
t1 <- xtabs(~ number, data = Depend)
t1
t1
barplot(t1, col = "lightblue", main = "Problem 1.26",
       xlab = "Number of Dependent Children", ylab = "Frequency")
rm(t1)
```

---

### Detroit

**Educational levels of a sample of 40 auto workers in Detroit**

**Description**

Data for Exercise 5.21

**Usage**

Detroit

**Format**

A data frame/tibble with 40 observations on one variable

- **educ**: the educational level (in years) of a sample of 40 auto workers in a plant in Detroit
Develop

References


Examples

EDA(Detroit$educ)

---

Demographic characteristics of developmental students at 2-year colleges and 4-year colleges

Description

Data used for Exercise 8.50

Usage

Develop

Format

A data frame/tibble with 5656 observations on two variables

- **race**: a factor with levels African American, American Indian, Asian, Latino, and White
- **college**: a factor with levels Two-year and Four-year

Source


References


Examples

```r
T1 <- xtabs(~race + college, data = Develop)
T1
chisq.test(T1)
rm(T1)
```
Devmath

*Test scores for students who failed developmental mathematics in the fall semester 1995*

Description

Data for Exercise 6.47

Usage

Devmath

Format

A data frame/tibble with 40 observations on one variable

- **score**: first exam score

Source

Data provided by Dr. Anita Kitchens.

References


Examples

```r
EDA(Devmath$score)
t.test(Devmath$score, mu = 80, alternative = "less")
```

---

Dice

*Outcomes and probabilities of the roll of a pair of fair dice*

Description

Data for Exercise 3.109

Usage

Dice
**Format**

A data frame/tibble with 11 observations on two variables

- **x** possible outcomes for the sum of two dice
- **px** probability for outcome **x**

**References**


**Examples**

```r
roll1 <- sample(1:6, 20000, replace = TRUE)
roll2 <- sample(1:6, 20000, replace = TRUE)
outcome <- roll1 + roll2
T1 <- table(outcome)/length(outcome)
remove(roll1, roll2, outcome)
T1
round(t(Dice), 5)
rm(roll1, roll2, T1)
```

---

**Diesel**

Diesel fuel prices in 1999-2000 in nine regions of the country

**Description**

Data for Exercise 2.8

**Usage**

Diesel

**Format**

A data frame/tibble with 650 observations on three variables

- **date** date when price was recorded
- **pricepergallon** price per gallon (in dollars)
- **location** a factor with levels California, CentralAtlantic, Coast, EastCoast, Gulf, LowerAtlantic, NatAvg, NorthEast, Rocky, and WesternMountain

**Source**

References


Examples

```r
par(las = 2)
boxplot(pricepergallon ~ location, data = Diesel)
boxplot(pricepergallon ~ location,
       data = droplevels(Diesel[Diesel$location == "EastCoast" |
                                Diesel$location == "Gulf" | Diesel$location == "NatAvg" |
                                Diesel$location == "Rocky" | Diesel$location == "California", ]),
       col = "pink", main = "Exercise 2.8")
par(las = 1)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Diesel, aes(x = date, y = pricepergallon,
                                   color = location)) +
    geom_point() +
    geom_smooth(se = FALSE) +
    theme_bw() +
    labs(y = "Price per Gallon (in dollars)"

## End(Not run)
```

---

Diplomat

Parking tickets issued to diplomats

Description

Data for Exercises 1.14 and 1.37

Usage

Diplomat

Format

A data frame/tibble with 10 observations on three variables

- **country**: a factor with levels Brazil, Bulgaria, Egypt, Indonesia, Israel, Nigeria, Russia, S. Korea, Ukraine, and Venezuela
- **number**: total number of tickets
- **rate**: number of tickets per vehicle per month

Source

Time, November 8, 1993. Figures are from January to June 1993.
References


Examples

```r
par(las = 2, mfrow = c(2, 2))
stripchart(number ~ country, data = Diplomat, pch = 19,
          col = "red", vertical = TRUE)
stripchart(rate ~ country, data = Diplomat, pch = 19,
          col = "blue", vertical = TRUE)
with(data = Diplomat,
     barplot(number, names.arg = country, col = "red"))
with(data = Diplomat,
     barplot(rate, names.arg = country, col = "blue"))
par(las = 1, mfrow = c(1, 1))
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Diplomat, aes(x = reorder(country, number),
     y = number)) +
  geom_bar(stat = "identity", fill = "pink", color = "black") +
  theme_bw() + labs(x = ",
     y = "Total Number of Tickets")
ggplot2::ggplot(data = Diplomat, aes(x = reorder(country, rate),
     y = rate)) +
  geom_bar(stat = "identity", fill = "pink", color = "black") +
  theme_bw() + labs(x = ",
     y = "Tickets per vehicle per month")
## End(Not run)
```

<table>
<thead>
<tr>
<th>Disposal</th>
<th>Toxic intensity for manufacturing plants producing herbicidal preparations</th>
</tr>
</thead>
</table>

Description

Data for Exercise 1.127

Usage

Disposal

Format

A data frame/tibble with 29 observations on one variable

- **pounds**: pounds of toxic waste per $1000 of shipments of its products
Source

References

Examples

```r
stem(Disposal$pounds)
fivenum(Disposal$pounds)
EDA(Disposal$pounds)
```

---

### Dogs

*Rankings of the favorite breeds of dogs*

#### Description
Data for Exercise 2.88

#### Usage
Dogs

#### Format
A data frame/tibble with 20 observations on three variables

- **breed** a factor with levels Beagle, Boxer, Chihuahua, Chow, Dachshund, Dalmatian, Doberman, Huskie, Labrador, Pomeranian, Poodle, Retriever, Rotweiler, Schnauzer, Shepherd, Shetland, ShihTzu, Spaniel, Springer, and Yorkshire
- **ranking** numeric ranking

#### Source

#### References
Examples

cor(Dogs$ranking[Dogs$year == "1992"], Dogs$ranking[Dogs$year == "1993"])
cor(Dogs$ranking[Dogs$year == "1997"], Dogs$ranking[Dogs$year == "1998"])
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Dogs, aes(x = reorder(breed, ranking), y = ranking)) +
  geom_bar(stat = "identity") +
  facet_grid(year ~. ) +
  theme(axis.text.x = element_text(angle = 85, vjust = 0.5))
## End(Not run)

### Domestic

Rates of domestic violence per 1,000 women by age groups

#### Description

Data for Exercise 1.20

#### Usage

Domestic

#### Format

A data frame/tibble with five observations on two variables

- **age**  a factor with levels 12–19, 20–24, 25–34, 35–49, and 50–64
- **rate** rate of domestic violence per 1000 women

#### Source

U.S. Department of Justice.

#### References


#### Examples

barplot(Domestic$rate, names.arg = Domestic$age)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Domestic, aes(x = age, y = rate)) +
  geom_bar(stat = "identity", fill = "purple", color = "black") +
  labs(x = "", y = "Domestic violence per 1000 women") +
Dopamine b-hydroxylase activity of schizophrenic patients treated with an antipsychotic drug

Description

Data for Exercises 5.14 and 7.49

Usage

Dopamine

Format

A data frame/tibble with 25 observations on two variables

- **dbh** dopamine b-hydroxylase activity (units are nmol/(ml)(h)/(mg) of protein)
- **group** a factor with levels nonpsychotic and psychotic

Source


References


Examples

```r
boxplot(dbh ~ group, data = Dopamine, col = "orange")
t.test(dbh ~ group, data = Dopamine, var.equal = TRUE)
```
Closing yearend Dow Jones Industrial averages from 1896 through 2000

Description

Data for Exercise 1.35

Usage

Dowjones

Format

A data frame/tibble with 105 observations on three variables

year date

close Dow Jones closing price

change percent change from previous year

References


Examples

```r
plot(close ~ year, data = Dowjones, type = "l", main = "Exercise 1.35")
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Dowjones, aes(x = year, y = close)) +
geom_point(size = 0.5) +
geom_line(color = "red") +
theme_bw() +
labs(y = "Dow Jones Closing Price")
## End(Not run)
```
Drug

Number of trials to master a task for a group of 28 subjects assigned to a control and an experimental group

Description

Data for Example 7.15

Usage

Drug

Drink

Opinion on referendum by view on moral issue of selling alcoholic beverages

Description

Data for Exercise 8.53

Usage

Drink

Format

A data frame/tibble with 472 observations on two variables

- **drinking**: a factor with levels ok, tolerated, and immoral
- **referendum**: a factor with levels for, against, and undecided

References


Examples

```r
T1 <- xtabs(~drinking + referendum, data = Drink)
T1
chisq.test(T1)
rm(T1)
```
Dyslexia

Format
A data frame/tibble with 28 observations on two variables

- **trials**: number of trials to master a task
- **group**: a factor with levels control and experimental

References

Examples

```r
boxplot(trials ~ group, data = Drug,
        main = "Example 7.15", col = c("yellow", "red"))
wilcox.test(trials ~ group, data = Drug)
t.test(rank(trials) ~ group, data = Drug, var.equal = TRUE)
```

---

Dyslexia

*Data on a group of college students diagnosed with dyslexia*

Description
Data for Exercise 2.90

Usage
Dyslexia

Format
A data frame/tibble with eight observations on seven variables

- **words**: number of words read per minute
- **age**: age of participant
- **gender**: a factor with levels female and male
- **handed**: a factor with levels left and right
- **weight**: weight of participant (in pounds)
- **height**: height of participant (in inches)
- **children**: number of children in family

References
Examples

plot(height ~ weight, data = Dyslexia)
plot(words ~ factor(handed), data = Dyslexia,
     xlab = "hand", col = "lightblue")

Earthqk

One hundred year record of worldwide seismic activity (1770-1869)

Description

Data for Exercise 6.97

Usage

Earthqk

Format

A data frame/tibble with 100 observations on two variables

year  year seismic activity recorded
severity annual incidence of severe earthquakes

Source


References


Examples

EDA(Earthqk$severity)
t.test(Earthqk$severity, mu = 100, alternative = "greater")
EDA Exploratory Data Analysis

Description
Function that produces a histogram, density plot, boxplot, and Q-Q plot.

Usage
EDA(x, trim = 0.05)

Arguments
x numeric vector. NAs and Infs are allowed but will be removed.
trim fraction (between 0 and 0.5, inclusive) of values to be trimmed from each end of the ordered data. If trim = 0.5, the result is the median.

Details
Will not return command window information on data sets containing more than 5000 observations. It will however still produce graphical output for data sets containing more than 5000 observations.

Value
Function returns various measures of center and location. The values returned for the Quartiles are based on the definitions provided in BSDA. The boxplot is based on the Quartiles returned in the commands window.

Note
Requires package e1071.

Author(s)
Alan T. Arnholt

Examples
EDA(rnorm(100))
# Produces four graphs for the 100 randomly
# generated standard normal variates.
Description

Data for Exercise 2.41

Usage

Educat

Format

A data frame/tibble with 51 observations on three variables


nodegree  percent of the population without a high school degree

crime  violent crimes per 100,000 population

References


Examples

plot(crime ~ nodegree, data = Educat,
     xlab = "Percent of population without high school degree",
     ylab = "Violent Crime Rate per 100,000")
Eggs

Number of eggs versus amounts of feed supplement

Description

Data for Exercise 9.22

Usage

Eggs

Format

A data frame/tibble with 12 observations on two variables

- **feed**: amount of feed supplement
- **eggs**: number of eggs per day for 100 chickens

References


Examples

```r
plot(eggs ~ feed, data = Eggs)
model <- lm(eggs ~ feed, data = Eggs)
abline(model, col = "red")
summary(model)
rm(model)
```

Elderly

Percent of the population over the age of 65

Description

Data for Exercise 1.92 and 2.61

Usage

Elderly
Format

A data frame/tibble with 51 observations on three variables


percent1985  percent of the population over the age of 65 in 1985
percent1998  percent of the population over the age of 65 in 1998

Source

U.S. Census Bureau Internet site, February 2000.

References


Examples

```r
with(data = Elderly,
    stripchart(x = list(percent1998, percent1985), method = "stack", pch = 19,
               col = c("red","blue"), group.names = c("1998", "1985"))
)
with(data = Elderly, cor(percent1998, percent1985))
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Elderly, aes(x = percent1985, y = percent1998)) +
  geom_point() +
  theme_bw()
## End(Not run)
```

---

**Energy**

*Amount of energy consumed by homes versus their sizes*

Description

Data for Exercises 2.5, 2.24, and 2.55

Usage

Energy
Engineer

Format

A data frame/tibble with 12 observations on two variables

size size of home (in square feet)
kilowatt kilowatt-hours per month

References


Examples

plot(kilowatt ~ size, data = Energy)
with(data = Energy, cor(size, kilowatt))
model <- lm(kilowatt ~ size, data = Energy)
plot(Energy$size, resid(model), xlab = "size")

Description

Data for Example 10.7

Usage

Engineer

Format

A data frame/tibble with 51 observations on two variables

salary salary (in $1000) 10 years after graduation
university a factor with levels A, B, and C

References

Examples

```r
boxplot(salary ~ university, data = Engineer,
        main = "Example 10.7", col = "yellow")
kruskal.test(salary ~ university, data = Engineer)
anova(lm(salary ~ university, data = Engineer))
anova(lm(rank(salary) ~ university, data = Engineer))
```

Entrance

College entrance exam scores for 24 high school seniors

Description

Data for Example 1.8

Usage

Entrance

Format

A data frame/tibble with 24 observations on one variable

score  college entrance exam score

References


Examples

```r
stem(Entrance$score)
stem(Entrance$score, scale = 2)
```
**Description**

Data for Exercise 1.65

**Usage**

Epaminicompact

**Format**

A data frame/tibble with 22 observations on ten variables

- **class**: a character variable with value MINICOMPACT CARS
- **manufacturer**: a character variable with values AUDI, BMW, JAGUAR, MERCEDES-BENZ, MITSUBISHI, and PORSCHE
- **carline**: a character variable with values 325CI CONVERTIBLE, 330CI CONVERTIBLE, 911 CARRERA 2/4, 911 TURBO, CLK320 (CABRIOLET), CLK430 (CABRIOLET), ECLIPSE SPYDER, JAGUAR XK8 CONVERTIBLE, JAGUAR XKR CONVERTIBLE, M3 CONVERTIBLE, TT COUPE, and TT COUPE QUATTRO
- **displ**: engine displacement (in liters)
- **cyl**: number of cylinders
- **trans**: a factor with levels Auto(L5), Auto(S4), Auto(S5), Manual(M5), and Manual(M6)
- **drv**: a factor with levels 4 (four wheel drive), F (front wheel drive), and R (rear wheel drive)
- **cty**: city mpg
- **hwy**: highway mpg
- **cmb**: combined city and highway mpg

**Source**

EPA data.

**References**


**Examples**

```r
summary(Epaminicompact$cty)
plot(hwy ~ cty, data = Epaminicompact)
```
Description

Data for Exercise 5.8

Usage

Epatwoseater

Format

A data frame/tibble with 36 observations on ten variables

class a character variable with value TWO SEATERS
manufacturer a character variable with values ACURA, AUDI, BMW, CHEVROLET, DODGE, FERRARI, HONDA, LAMBORGHINI, MAZDA, MERCEDES-BENZ, PLYMOUTH, PORSCHE, and TOYOTA
carline a character variable with values BOXSTER, BOXSTER S, CORVETTE, DB132/144 DIABLO, FERRARI 360 MODENA/SPIDER, FERRARI 550 MARANELLO/BARCHETTA, INSIGHT, MR2, MZ5 MIATA, NSX, PROWLER, S2000, SL500, SL600, SLK230 KOMPRESSOR, SLK320, TT ROADSTER, TT ROADSTER QUATTRO, VIPER CONVERTIBLE, VIPER COUPE, Z3 COUPE, Z3 ROADSTER, and Z8
displ engine displacement (in liters)
cyl number of cylinders
class a factor with levels Auto(L4), Auto(L5), Auto(S4), Auto(S5), Auto(S6), Manual(M5), and Manual(M6)
driv a factor with levels 4(four wheel drive) F(front wheel drive) R(rear wheel drive)
city city mpg
highway mpg
cmb combined city and highway mpg
@source Environmental Protection Agency.

References


Examples

summary(Epatwoseater$city)
plot(hwy ~ city, data = Epatwoseater)
boxplot(city ~ driv, data = Epatwoseater, col = "lightgreen")
### Executiv

**Ages of 25 executives**

**Description**

Data for Exercise 1.104

**Usage**

```
Executiv
```

**Format**

A data frame/tibble with 25 observations on one variable

- **age** a numeric vector

**References**


**Examples**

```
hist(Executiv$age, xlab = "Age of banking executives", breaks = 5, main = "", col = "gray")
```

### Exercise

**Weight loss for 30 members of an exercise program**

**Description**

Data for Exercise 1.44

**Usage**

```
Exercise
```

**Format**

A data frame/tibble with 30 observations on one variable

- **loss** a numeric vector
References

Examples

stem(Exercise$loss)

<table>
<thead>
<tr>
<th>Fabric</th>
<th>Measures of softness of ten different clothing garments washed with and without a softener</th>
</tr>
</thead>
</table>

Description
Data for Example 7.21

Usage
Fabric

Format
A data frame/tibble with 20 observations on three variables
garment a numeric vector
softner a character variable with values with and without
softness a numeric vector

References

Examples

wilcox.test(softness ~ softner, data = Fabric,
            paired = TRUE, alternative = "greater")

## Not run:
library(tidyrr)
T7 <- tidyr::spread(Fabric, softner, softness) %>%
  mutate(di = with - without, adi = abs(di), rk = rank(adi),
         srk = sign(di)*rk)
T7
t.test(T7$srk, alternative = "greater")

## End(Not run)
Faithful

Waiting times between successive eruptions of the Old Faithful geyser

Description

Data for Exercise 5.12 and 5.111

Usage

Faithful

Format

A data frame/tibble with 299 observations on two variables

- **time**  a numeric vector
- **eruption**  a factor with levels 1 and 2

Source


References


Examples

t.test(time ~ eruption, data = Faithful)
hist(Faithful$time, xlab = "wait time", main = "", freq = FALSE)
lines(density(Faithful$time))

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Faithful, aes(x = time, y = ..density..)) +
geom_histogram(binwidth = 5, fill = "pink", col = "black") +
geom_density() +
theme_bw() +
labs(x = "wait time")

## End(Not run)
Size of family versus cost per person per week for groceries

Description

Data for Exercise 2.89

Usage

Family

Format

A data frame/tibble with 20 observations on two variables

- **number**  number in family
- **cost**  cost per person (in dollars)

References


Examples

```r
plot(cost ~ number, data = Family)
abline(lm(cost ~ number, data = Family), col = "red")
cor(Family$cost, Family$number)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Family, aes(x = number, y = cost)) +
  geom_point() +
  geom_smooth(method = "lm") +
  theme_bw()

## End(Not run)
```
Ferraro1

Choice of presidential ticket in 1984 by gender

Description

Data for Exercise 8.23

Usage

Ferraro1

Format

A data frame/tibble with 1000 observations on two variables

gender  a factor with levels Men and Women

candidate a character vector of 1984 president and vice-president candidates

References


Examples

T1 <- xtabs(~gender + candidate, data = Ferraro1)
T1
chisq.test(T1)
rm(T1)

Ferraro2

Choice of vice presidential candidate in 1984 by gender

Description

Data for Exercise 8.23

Usage

Ferraro2
Fertility

Format
A data frame/tibble with 1000 observations on two variables

- **gender**: a factor with levels Men and Women
- **candidate**: a character vector of 1984 president and vice-president candidates

References

Examples

```
T1 <- xtabs(~gender + candidate, data = Ferraro2)
T1
chisq.test(T1)
rm(T1)
```

<table>
<thead>
<tr>
<th>Fertility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fertility rates of all 50 states and DC</td>
</tr>
</tbody>
</table>

Description
Data for Exercise 1.125

Usage
Fertility

Format
A data frame/tibble with 51 observations on two variables

- **rate**: fertility rate (expected number of births during childbearing years)

Source
Population Reference Bureau.
References


Examples

```r
stem(Fertility$rate)
fivenum(Fertility$rate)
EDA(Fertility$rate)
```

<table>
<thead>
<tr>
<th>Firstchi</th>
<th>Ages of women at the birth of their first child</th>
</tr>
</thead>
</table>

Description

Data for Exercise 5.11

Usage

Firstchi

Format

A data frame/tibble with 87 observations on one variable

| age | age of woman at birth of her first child |

References


Examples

```r
EDA(Firstchi$age)
```
Fish

Length and number of fish caught with small and large mesh codend

Description

Data for Exercises 5.83, 5.119, and 7.29

Usage

Fish

Format

A data frame/tibble with 1534 observations on two variables

- **codend** a character variable with values smallmesh and largemesh
- **length** length of the fish measured in centimeters

Source


References


Examples

```r
tapply(Fish$length, Fish$codend, median, na.rm = TRUE)
SIGN.test(Fish$length[Fish$codend == "smallmesh"], conf.level = 0.99)
## Not run:
dplyr::group_by(Fish, codend) %>%
  summarize(MEDIAN = median(length, na.rm = TRUE))
## End(Not run)
```
Fitness

Number of sit-ups before and after a physical fitness course

Description

Data for Exercise 7.71

Usage

Fitness

Format

A data frame/tibble with 18 observations on the three variables

subject a character variable indicating subject number

test a character variable with values After and Before

number a numeric vector recording the number of sit-ups performed in one minute

References


Examples

```r
t.test(number ~ test, data = Fitness, alternative = "greater", paired = TRUE)

## Not run:
Wide <- tidyr::spread(Fitness, test, number) %>%
  mutate(diff = After - Before)
Wide
qqnorm(Wide$diff)
qqline(Wide$diff)
t.test(Wide$diff, alternative = "greater")

## End(Not run)
```
Florida voter results in the 2000 presidential election

Description

Data for Statistical Insight Chapter 2

Usage

Florida2000

Format

A data frame/tibble with 67 observations on 12 variables

- **county** a character variable with values ALACHUA, BAKER, BAY, BRADFORD, BREVARD, BROWARD, CALHOUN, CHARLOTTE, CITRUS, CLAY, COLLIER, COLUMBIA, DADE, DE SOTO, DIXIE, DUVAL, ESCAMBIA, FLAGLER, FRANKLIN, GADSDEN, GILCHRIST, GLADES, GULF, HAMILTON, HARDEE, HENDRY, HERNANDO, HIGHLANDS, HILLSBOROUGH, HOLMES, INDIAN RIVER, JACKSON, JEFFERSON, LAFAYETTE, LAKE, LEE, LEON, LEVY, LIBERTY, MADISON, MANATEE, MARION, MARTIN, MONROE, NASSAU, OKALOOSA, OKEECHOBEE, ORANGE, OSCEOLA, PALM BEACH, PASCOS, PINELLAS, POLK, PUTNAM, SANTA ROSA, SARASOTA, SEMINOLE, ST. JOHNS, ST. LUCIE, SUMTER, SUWANNEE, TAYLOR, UNION, VOLUSIA, WAKULLA, WALTON, and WASHINGTON

- **gore** number of votes
- **bush** number of votes
- **buchanan** number of votes
- **nader** number of votes
- **browne** number of votes
- **hagelin** number of votes
- **harris** number of votes
- **mcreynolds** number of votes
- **moorehead** number of votes
- **phillips** number of votes
- **total** number of votes

References


Examples

```r
plot(buchanan ~ total, data = Florida2000,
     xlab = "Total votes cast (in thousands)",
     ylab = "Votes for Buchanan")
```
Fluid  

Breakdown times of an insulating fluid under various levels of voltage stress

Description
Data for Exercise 5.76

Usage
Fluid

Format
A data frame/tibble with 76 observations on two variables

kilovolts  a character variable showing kilowatts

time  breakdown time (in minutes)

Source

References

Examples

```
DF1 <- Fluid[Fluid$kilovolts == "34kV", ]
DF1
# OR
DF2 <- subset(Fluid, subset = kilovolts == "34kV")
DF2
stem(DF2$time)
SIGN.test(DF2$time)
## Not run:
library(dplyr)
DF3 <- dplyr::filter(Fluid, kilovolts == "34kV")
DF3
## End(Not run)
```
Food

Annual food expenditures for 40 single households in Ohio

Description

Data for Exercise 5.106

Usage

Food

Format

A data frame/tibble with 40 observations on one variable

- **expenditure** a numeric vector recording annual food expenditure (in dollars) in the state of Ohio.

Source


References


Examples

EDA(Food$expenditure)

Framingham

Cholesterol values of 62 subjects in the Framingham Heart Study

Description

Data for Exercises 1.56, 1.75, 3.69, and 5.60

Usage

Framingham

Format

A data frame/tibble with 62 observations on one variable

- **cholest** a numeric vector with cholesterol values
Source

References

Examples

```r
stem(Framingh$cholest)
boxplot(Framingh$cholest, horizontal = TRUE)
hist(Framingh$cholest, freq = FALSE)
lines(density(Framingh$cholest))
mean(Framingh$cholest > 200 & Framingh$cholest < 240)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Framingh, aes(x = factor(1), y = cholest)) +
  geom_boxplot() + # boxplot
  labs(x = "") + # no x label
  theme_bw() + # black and white theme
  geom_jitter(width = 0.2) + # jitter points
  coord_flip() # Create horizontal plot

ggplot2::ggplot(data = Framingh, aes(x = cholest, y = ..density..)) +
  geom_histogram(fill = "pink", binwidth = 15, color = "black") +
  geom_density() +
  theme_bw()

## End(Not run)
```

---

<table>
<thead>
<tr>
<th>Freshman</th>
<th>Ages of a random sample of 30 college freshmen</th>
</tr>
</thead>
</table>

Description
Data for Exercise 6.53

Usage
Freshman

Format
A data frame/tibble with 30 observations on one variable

- **age** a numeric vector of ages
References


Examples

```
SIGN.test(Freshman$age, md = 19)
```

---

**Funeral**

*Cost of funeral by region of country*

<table>
<thead>
<tr>
<th>Region</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central</td>
<td>less than expected</td>
</tr>
<tr>
<td>East</td>
<td>about what expected</td>
</tr>
<tr>
<td>South</td>
<td>more than expected</td>
</tr>
<tr>
<td>West</td>
<td></td>
</tr>
</tbody>
</table>

Description

Data for Exercise 8.54

Usage

```
Funeral
```

Format

A data frame/tibble with 400 observations on two variables

- **region**: a factor with levels Central, East, South, and West
- **cost**: a factor with levels less than expected, about what expected, and more than expected

References


Examples

```
T1 <- xtabs(~region + cost, data = Funeral)
T1
chisq.test(T1)
rm(T1)
```
Galaxie

Velocities of 82 galaxies in the Corona Borealis region

Description

Data for Example 5.2

Usage

Galaxie

Format

A data frame/tibble with 82 observations on one variable

velocity velocity measured in kilometers per second

Source


References


Examples

`EDA(Galaxie$velocity)`

Gallup

Results of a Gallup poll on possession of marijuana as a criminal offense conducted in 1980

Description

Data for Exercise 2.76

Usage

Gallup
Gasoline

Price of regular unleaded gasoline obtained from 25 service stations

Description
Data for Exercise 1.45

Usage
Gasoline

Format
A data frame/tibble with 1,200 observations on two variables

demographics  a factor with levels National, Gender: Male Gender: Female, Education: College, Education: High School, Education: Grade School, Age: 18-24, Age: 25-29, Age: 30-49, Age: 50-older, Religion: Protestant, and Religion: Catholic

opinion  a factor with levels Criminal, Not Criminal, and No Opinion

Source

References

Examples

```r
T1 <- xtabs(~demographics + opinion, data = Gallup)
T1
T1[c(2, 3), ]
barplot(t(T1[c(2, 3), ]))
barplot(t(T1[c(2, 3), ]), beside = TRUE)

## Not run:
library(dplyr)
library(ggplot2)
dplyr::filter(Gallup, demographics == "Gender: Male" | demographics == "Gender: Female") %>%
ggplot2::ggplot(aes(x = demographics, fill = opinion)) +
  geom_bar() +
  theme_bw() +
  labs(y = "Fraction")

## End(Not run)
```
German

Format

A data frame/tibble with 25 observations on one variable

price  price for one gallon of gasoline

References


Examples

stem(Gasoline$price)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Gasoline, aes(x = factor(1), y = price)) +
  geom_violin() +
  geom_jitter() +
  theme_bw()

## End(Not run)

German

Number of errors in copying a German passage before and after an experimental course in German

Description

Data for Exercise 7.60

Usage

German

Format

A data frame/tibble with ten observations on three variables

student  a character variable indicating student number

when  a character variable with values Before and After to indicate when the student received experimental instruction in German

errors  the number of errors in copying a German passage
References


Examples

t.test(errors ~ when, data = German, paired = TRUE)
wilcox.test(errors ~ when, data = German)

## Not run:
T8 <- tidyr::spread(German, when, errors) %>%
mutate(di = After - Before, adi = abs(di), rk = rank(adi), srk = sign(di)*rk)
T8
qqnorm(T8$di)
qqline(T8$di)
t.test(T8$srk)

## End(Not run)

Golf

Distances a golf ball can be driven by 20 professional golfers

Description

Data for Exercise 5.24

Usage

Golf

Format

A data frame/tibble with 20 observations on one variable

yards distance a golf ball is driven in yards

References

Examples

stem(Golf$yards)
qqnorm(Golf$yards)
qqline(Golf$yards)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Golf, aes(sample = yards)) +
  geom_qq() +
  theme_bw()

## End(Not run)

### Governor

<table>
<thead>
<tr>
<th>Governor</th>
<th>Annual salaries for state governors in 1994 and 1999</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Description

Data for Exercise 5.112

Usage

Governor

Format

A data frame/tibble with 50 observations on three variables


- **year**: a factor indicating year

- **salary**: a numeric vector with the governor’s salary (in dollars)

Source

*The 2000 World Almanac and Book of Facts.*

References

Examples

```r
boxplot(salary ~ year, data = Governor)
```

## Not run:
```
library(ggplot2)

ggplot2::ggplot(data = Governor, aes(x = salary)) +
  geom_density(fill = "pink") +
  facet_grid(year ~ .) +
  theme_bw()
```

## End(Not run)

---

### Gpa

**High school GPA versus college GPA**

#### Description

Data for Example 2.13

#### Usage

```r
Gpa
```

#### Format

A data frame/tibble with 10 observations on two variables

- **hsgpa** high school gpa
- **collgpa** college gpa

#### References


#### Examples

```r
plot(collgpa ~ hsgpa, data = Gpa)
mod <- lm(collgpa ~ hsgpa, data = Gpa)
abline(mod)  # add line
yhat <- predict(mod)  # fitted values
e <- resid(mod)  # residuals
cbind(Gpa, yhat, e)  # Table 2.1
cor(Gpa$hsgpa, Gpa$collgpa)
```

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Gpa, aes(x = hsgpa, y = collgpa)) +
  geom_point() +
  geom_smooth(method = "lm") +
  theme_bw()

## End(Not run)

---

**Grades**

Test grades in a beginning statistics class

Description

Data for Exercise 1.120

Usage

Grades

Format

A data frame with 29 observations on one variable

grades a numeric vector containing test grades

References


Examples

```r
hist(Grades$grades, main = "", xlab = "Test grades", right = FALSE)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Grades, aes(x = grades, y = ..density..)) +
  geom_histogram(fill = "pink", binwidth = 5, color = "black") +
  geom_density(lwd = 2, color = "red") +
  theme_bw()

## End(Not run)
```
### Graduate

*Graduation rates for student athletes in the Southeastern Conf.*

<table>
<thead>
<tr>
<th>School</th>
<th>Code</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>Al</td>
<td>75.6</td>
</tr>
<tr>
<td>Arkansas</td>
<td>Ar</td>
<td>68.3</td>
</tr>
<tr>
<td>Auburn</td>
<td>Au Fl</td>
<td>78.2</td>
</tr>
<tr>
<td>Florida</td>
<td>Fl</td>
<td>80.1</td>
</tr>
<tr>
<td>Georgia</td>
<td>Ge</td>
<td>79.5</td>
</tr>
<tr>
<td>Kentucky</td>
<td>Ke</td>
<td>77.9</td>
</tr>
<tr>
<td>Louisiana St</td>
<td>LSt</td>
<td>79.1</td>
</tr>
<tr>
<td>Mississippi</td>
<td>Mi</td>
<td>82.1</td>
</tr>
<tr>
<td>Mississippi St</td>
<td>MSt</td>
<td>81.7</td>
</tr>
<tr>
<td>South Carolina</td>
<td>SC</td>
<td>80.3</td>
</tr>
<tr>
<td>Tennessee</td>
<td>Te</td>
<td>78.7</td>
</tr>
<tr>
<td>Vanderbilt</td>
<td>Va</td>
<td>77.9</td>
</tr>
</tbody>
</table>

**Description**

Data for Exercise 1.118

**Usage**

Graduate

**Format**

A data frame/tibble with 12 observations on three variables

- **school**: a character variable with values Alabama, Arkansas, Auburn, Florida, Georgia, Kentucky, Louisiana St, Mississippi, Mississippi St, South Carolina, Tennessee, and Vanderbilt
- **code**: a character variable with values Al, Ar, Au Fl, Ge, Ke, LSt, Mi, MSt, SC, Te, and Va
- **percent**: graduation rate

**References**


**Examples**

```r
barplot(Graduate$percent, names.arg = Graduate$school,
       las = 2, cex.names = 0.7, col = "tomato")
```

---

### Greenriv

*Varve thickness from a sequence through an Eocene lake deposit in the Rocky Mountains*

<table>
<thead>
<tr>
<th>Greenriv</th>
</tr>
</thead>
</table>

**Description**

Data for Exercise 6.57

**Usage**

Greenriv
**Format**

A data frame/tibble with 37 observations on one variable

**thick** varve thickness in millimeters

**References**


**Examples**

```r
stem(Greenriv$thick)
SIGN.test(Greenriv$thick, md = 7.3, alternative = "greater")
```

<table>
<thead>
<tr>
<th>Greenriv2</th>
<th>Thickness of a varved section of the Green river oil shale deposit near a major lake in the Rocky Mountains</th>
</tr>
</thead>
</table>

**Description**

Data for Exercises 6.45 and 6.98

**Usage**

Greenriv2

**Format**

A data frame/tibble with 101 observations on one variable

**thick** varve thickness (in millimeters)

**Source**


**References**


**Examples**

```r
stem(Grnriv2$thick)
t.test(Grnriv2$thick, mu = 8, alternative = "less")
```
Description
Data for Exercise 10.42

Usage
Groupabc

Format
A data frame/tibble with 45 observations on two variables

group a factor with levels A, B, and C
response a numeric vector

References

Examples

```r
boxplot(response ~ group, data = Groupabc,
        col = c("red", "blue", "green"))
anova(lm(response ~ group, data = Groupabc))
```

Description
An illustration of analysis of variance

Usage
Groups

Format
A data frame/tibble with 78 observations on two variables

group a factor with levels A, B, and C
response a numeric vector
Gym

References


Examples

```r
boxplot(response ~ group, data = Groups, col = c("red", "blue", "green"))
anova(lm(response ~ group, data = Groups))
```

---

Gym

*Children’s age versus number of completed gymnastic activities*

Description

Data for Exercises 2.21 and 9.14

Usage

Gym

Format

A data frame/tibble with eight observations on three variables

- **age** age of child
- **number** number of gymnastic activities successfully completed

References


Examples

```r
plot(number ~ age, data = Gym)
model <- lm(number ~ age, data = Gym)
abline(model, col = "red")
summary(model)
```
Habits

Study habits of students in two matched school districts

Description

Data for Exercise 7.57

Usage

Habits

Format

A data frame/tibble with 11 observations on four variables

- **A**: study habit score
- **B**: study habit score
- **differ**: B minus A
- **signrks**: the signed-ranked-differences

References


Examples

```r
shapiro.test(Habits$differ)
qqnorm(Habits$differ)
qqline(Habits$differ)
wilcox.test(Habits$B, Habits$A, paired = TRUE, alternative = "less")
t.test(Habits$signrks, alternative = "less")
```

```r
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Habits, aes(x = differ)) +
  geom_dotplot(fill = "blue") +
  theme_bw()
## End(Not run)
```
**Haptoglo**

*Haptoglobin concentration in blood serum of 8 healthy adults*

**Description**

Data for Example 6.9

**Usage**

Haptoglo

**Format**

A data frame/tibble with eight observations on one variable

- **concent** haptoglobin concentration (in grams per liter)

**References**


**Examples**

```r
shapiro.test(Haptoglo$concent)
t.test(Haptoglo$concent, mu = 2, alternative = "less")
```

---

**Hardware**

*Daily receipts for a small hardware store for 31 working days*

**Description**

Daily receipts for a small hardware store for 31 working days

**Usage**

Hardware

**Format**

A data frame with 31 observations on one variable

- **receipt** a numeric vector of daily receipts (in dollars)
Source


References


Examples

```r
stem(Hardware$receipt)
```

<table>
<thead>
<tr>
<th>Hardwood</th>
<th>Tensile strength of Kraft paper for different percentages of hardwood in the batches of pulp</th>
</tr>
</thead>
</table>

Description

Data for Example 2.18 and Exercise 9.34

Usage

`Hardwood`

Format

A data frame/tibble with 19 observations on two variables

- `tensile`: tensile strength of kraft paper (in pounds per square inch)
- `hardwood`: percent of hardwood in the batch of pulp that was used to produce the paper

Source


References

Examples

```r
plot(tensile ~ hardwood, data = Hardwood)
model <- lm(tensile ~ hardwood, data = Hardwood)
abline(model, col = "red")
plot(model, which = 1)
```

---

**Heat**

*Primary heating sources of homes on Indian reservations versus all households*

---

**Description**

Data for Exercise 1.29

**Usage**

Heat

**Format**

A data frame/tibble with 301 observations on two variables

- **fuel** a factor with levels Utility gas, LP bottled gas, Electricity, Fuel oil, Wood, and Other
- **location** a factor with levels American Indians on reservation, All U.S. households, and American Indians not on reservations

**Source**


**References**


**Examples**

```r
T1 <- xtabs(~ fuel + location, data = Heat)
T1
barplot(t(T1), beside = TRUE, legend = TRUE)
```

```r
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Heat, aes(x = fuel, fill = location)) +
```
Heating

Fuel efficiency ratings for three types of oil heaters

Description

Data for Exercise 10.32

Usage

Heating

Format

A data frame/tibble with 90 observations on the two variables

type a factor with levels A, B, and C denoting the type of oil heater
efficiency heater efficiency rating

References


Examples

```r
boxplot(efficiency ~ type, data = Heating,
       col = c("red", "blue", "green"))
kruskal.test(efficiency ~ type, data = Heating)
```
**Hodgkin**

Results of treatments for Hodgkin's disease

---

**Description**

Data for Exercise 2.77

**Usage**

Hodgkin

**Format**

A data frame/tibble with 538 observations on two variables

- **type** a factor with levels LD, LP, MC, and NS
- **response** a factor with levels Positive, Partial, and None

**Source**

I. Dunsmore, F. Daly, *Statistical Methods, Unit 9, Categorical Data*, Milton Keynes, The Open University, 18.

**References**


**Examples**

```r
T1 <- xtabs(~type + response, data = Hodgkin)
T1
barplot(t(T1), legend = TRUE, beside = TRUE)

## Not run:
library(ggplot2)
ggplot(data = Hodgkin, aes(x = type, fill = response)) +
  geom_bar(position = "dodge") +
  theme_bw()

## End(Not run)
```
## Homes

**Median prices of single-family homes in 65 metropolitan statistical areas**

### Description

Data for Statistical Insight Chapter 5

### Usage

Homes

### Format

A data frame/tibble with 65 observations on the four variables

- **city**: a character variable with values Akron OH, Albuquerque NM, Anaheim CA, Atlanta GA, Baltimore MD, Baton Rouge LA, Birmingham AL, Boston MA, Bradenton FL, Buffalo NY, Charleston SC, Chicago IL, Cincinnati OH, Cleveland OH, Columbia SC, Columbus OH, Corpus Christi TX, Dallas TX, Daytona Beach FL, Denver CO, Des Moines IA, Detroit MI, El Paso TX, Grand Rapids MI, Hartford CT, Honolulu HI, Houston TX, Indianapolis IN, Jacksonville FL, Kansas City MO, Knoxville TN, Las Vegas NV, Los Angeles CA, Louisville KY, Madison WI, Memphis TN, Miami FL, Milwaukee WI, Minneapolis MN, Mobile AL, Nashville TN, New Haven CT, New Orleans LA, New York NY, Oklahoma City OK, Omaha NE, Orlando FL, Philadelphia PA, Phoenix AZ, Pittsburgh PA, Portland OR, Providence RI, Sacramento CA, Salt Lake City UT, San Antonio TX, San Diego CA, San Francisco CA, Seattle WA, Spokane WA, St Louis MO, Syracuse NY, Tampa FL, Toledo OH, Tulsa OK, and Washington DC

- **region**: a character variable with values Midwest, Northeast, South, and West

- **year**: a factor with levels 1994 and 2000

- **price**: median house price (in dollars)

### Source

National Association of Realtors.

### References


### Examples

```r
taxapply(Homes$price, Homes$year, mean)
taxapply(Homes$price, Homes$region, mean)
p2000 <- subset(Homes, year == "2000")
p1994 <- subset(Homes, year == "1994")
## Not run:
```
library(dplyr)
library(ggplot2)
dplyr::group_by(Homes, year, region) %>%
  summarize(AvgPrice = mean(price))
ggplot2::ggplot(data = Homes, aes(x = region, y = price)) +
  geom_boxplot() +
  theme_bw() +
  facet_grid(year ~ .)

## End(Not run)

---

**Homework**

**Number of hours per week spent on homework for private and public high school students**

---

**Description**

Data for Exercise 7.78

**Usage**

Homework

**Format**

A data frame with 30 observations on two variables

- **school** type of school either private or public
- **time** number of hours per week spent on homework

**References**


**Examples**

```r
boxplot(time ~ school, data = Homework,
  ylab = "Hours per week spent on homework")
#
t.test(time ~ school, data = Homework)
```
### Honda

**Miles per gallon for a Honda Civic on 35 different occasions**

**Description**

Data for Statistical Insight Chapter 6

**Usage**

Honda

**Format**

A data frame/tibble with 35 observations on one variable

- **mileage** miles per gallon for a Honda Civic

**References**


**Examples**

```r
t.test(Honda$mileage, mu = 40, alternative = "less")
```

### Hostile

**Hostility levels of high school students from rural, suburban, and urban areas**

**Description**

Data for Example 10.6

**Usage**

Hostile

**Format**

A data frame/tibble with 135 observations on two variables

- **location** a factor with the location of the high school student (Rural, Suburban, or Urban)
- **hostility** the score from the Hostility Level Test
Housing

References


Examples

```r
boxplot(hostility ~ location, data = Hostile,
       col = c("red", "blue", "green"))
kruskal.test(hostility ~ location, data = Hostile)
```

---

### Housing

*Median home prices for 1984 and 1993 in 37 markets across the U.S.*

---

**Description**

Data for Exercise 5.82

**Usage**

Housing

**Format**

A data frame/tibble with 74 observations on three variables

- **city** a character variable with values Albany, Anaheim, Atlanta, Baltimore, Birmingham, Boston, Chicago, Cincinnati, Cleveland, Columbus, Dallas, Denver, Detroit, Ft Lauderdale, Houston, Indianapolis, Kansas City, Los Angeles, Louisville, Memphis, Miami, Milwaukee, Minneapolis, Nashville, New York, Oklahoma City, Philadelphia, Providence, Rochester, Salt Lake City, San Antonio, San Diego, San Francisco, San Jose, St Louis, Tampa, and Washington

- **year** a factor with levels 1984 and 1993

- **price** median house price (in dollars)

**Source**

National Association of Realtors.

**References**

Examples

```r
stripchart(price ~ year, data = Housing, method = "stack",
           pch = 1, col = c("red", "blue"))
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Housing, aes(x = price, fill = year)) +
               geom_dotplot() +
               facet_grid(year ~ .) +
               theme_bw()
## End(Not run)
```

Hurrican

Number of storms, hurricanes and El Nino effects from 1950 through 1995

Description

Data for Exercises 1.38, 10.19, and Example 1.6

Usage

Hurrican

Format

A data frame/tibble with 46 observations on four variables

- **year**: a numeric vector indicating year
- **storms**: a numeric vector recording number of storms
- **hurrican**: a numeric vector recording number of hurricanes
- **elnino**: a factor with levels cold, neutral, and warm

Source

National Hurricane Center.

References

Iceberg

Examples

```r
T1 <- xtabs(~hurrican, data = Hurrican)
T1
barplot(T1, col = "blue", main = "Problem 1.38",
       xlab = "Number of hurricanes",
       ylab = "Number of seasons")
boxplot(storms ~ elnino, data = Hurrican,
        col = c("blue", "yellow", "red"))
anova(lm(storms ~ elnino, data = Hurrican))
rm(T1)
```

<table>
<thead>
<tr>
<th>Iceberg</th>
<th>Number of icebergs sighted each month south of Newfoundland and south of the Grand Banks in 1920</th>
</tr>
</thead>
</table>

Description

Data for Exercise 2.46 and 2.60

Usage

Iceberg

Format

A data frame with 12 observations on three variables

- **month** a character variable with abbreviated months of the year
- **Newfoundland** number of icebergs sighted south of Newfoundland
- **Grand Banks** number of icebergs sighted south of Grand Banks

Source


References


Examples

```r
plot(Newfoundland ~ `Grand Banks`, data = Iceberg)
abline(lm(Newfoundland ~ `Grand Banks`, data = Iceberg), col = "blue")
```
Income

Percent change in personal income from 1st to 2nd quarter in 2000

Description

Data for Exercise 1.33

Usage

Income

Format

A data frame/tibble with 51 observations on two variables


- **percent_change**: percent change in income from first quarter to the second quarter of 2000

Source

US Department of Commerce.

References


Examples

```r
Income$class <- cut(Income$percent_change, 
    breaks = c(-Inf, 0.5, 1.0, 1.5, 2.0, Inf))
T1 <- xtabs(~class, data = Income)
T1
barplot(T1, col = "pink")
## Not run:
library(ggplot2)
DF <- as.data.frame(T1)
DF
ggplot2::ggplot(data = DF, aes(x = class, y = Freq)) + 
  geom_bar(stat = "identity", fill = "purple") + 
  theme_bw()
```
## Illustrates a comparison problem for long-tailed distributions

### Description

Data for Exercise 7.41

### Usage

Independent

### Format

A data frame/tibble with 46 observations on two variables

- **score** a numeric vector
- **group** a factor with levels A and B

### References


### Examples

```r
qqnorm(Independent$score[Independent$group=="A"])
qqline(Independent$score[Independent$group=="A"])
qqnorm(Independent$score[Independent$group=="B"])
qqline(Independent$score[Independent$group=="B"])
boxplot(score ~ group, data = Independent, col = "blue")
wilcox.test(score ~ group, data = Independent)
```
Indian

*Educational attainment versus per capita income and poverty rate for American Indians living on reservations*

**Description**

Data for Exercise 2.95

**Usage**

Indian

**Format**

A data frame/tibble with ten observations on four variables

- **reservation**  a character variable with values Blackfeet, Fort Apache, Gila River, Hopi, Navajo, Papago, Pine Ridge, Rosebud, San Carlos, and Zuni Pueblo
- **percent high school**  percent who have graduated from high school
- **per capita income**  per capita income (in dollars)
- **poverty rate**  percent poverty

**References**


**Examples**

```r
par(mfrow = c(1, 2))
plot('per capita income' ~ 'percent high school', data = Indian,
    xlab = "Percent high school graduates", ylab = "Per capita income")
plot('poverty rate' ~ 'percent high school', data = Indian,
    xlab = "Percent high school graduates", ylab = "Percent poverty")
par(mfrow = c(1, 1))
```
**Indiapol**

*Average miles per hour for the winners of the Indianapolis 500 race*

**Description**

Data for Exercise 1.128

**Usage**

Indiapol

**Format**

A data frame/tibble with 39 observations on two variables

- **year** the year of the race
- **speed** the winners average speed (in mph)

**Source**


**References**


**Examples**

```r
plot(speed ~ year, data = Indiapol, type = "b")
```

---

**Indy500**

*Qualifying miles per hour and number of previous starts for drivers in 79th Indianapolis 500 race*

**Description**

Data for Exercises 7.11 and 7.36

**Usage**

Indy500
Inflation

Private pay increase of salaried employees versus inflation rate

Description

Data for Exercises 2.12 and 2.29

Usage

Inflation

Format

A data frame/tibble with 33 observations on four variables

**driver** a character variable with values andretti, bachelart, boesel, brayton, c.guerrero, cheever, fabi, Fernandez, ferran, fittipaldi, fox, goodyear, gordon, gugelmin, herta, james, johansson, jones, lazier, luyendyk, matsuda, matsushita, pruett, r.guerrero, rahal, ribeiro, salazar, sharp, sullivan, tracy, vasser, villeneuve, and zampedri

**qualif** qualifying speed (in mph)

**starts** number of Indianapolis 500 starts

**group** a numeric vector where 1 indicates the driver has 4 or fewer Indianapolis 500 starts and a 2 for drivers with 5 or more Indianapolis 500 starts

References


Examples

```r
stripchart(qualif ~ group, data = Indy500, method = "stack", pch = 19, col = c("red", "blue"))
boxplot(qualif ~ group, data = Indy500)
t.test(qualif ~ group, data = Indy500)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Indy500, aes(sample = qualif)) + geom_qq() + facet_grid(group ~ .) + theme_bw()
## End(Not run)
```
**Format**

A data frame/tibble with 24 observations on four variables

- **year** a numeric vector of years
- **pay** average hourly wage for salaried employees (in dollars)
- **increase** percent increase in hourly wage over previous year
- **inflation** percent inflation rate

**Source**


**References**


**Examples**

```r
plot(increase ~ inflation, data = Inflation)
cor(Inflation$increase, Inflation$inflation, use = "complete.obs")
```

---

**Inletoil**  
*Inlet oil temperature through a valve*

**Description**

Data for Exercises 5.91 and 6.48

**Usage**

Inletoil

**Format**

A data frame/tibble with 12 observations on one variable

- **temp** inlet oil temperature (Fahrenheit)

**References**

Inmate

Examples

```r
hist(Inletoil$temp, breaks = 3)
norm(Inletoil$temp)
qline(Inletoil$temp)
t.test(Inletoil$temp)
t.test(Inletoil$temp, mu = 98, alternative = "less")
```

Inmate  Type of drug offense by race

Description

Data for Statistical Insight Chapter 8

Usage

Inmate

Format

A data frame/tibble with 28,047 observations on two variables

- **race**: a factor with levels white, black, and hispanic
- **drug**: a factor with levels heroin, crack, cocaine, and marijuana

Source


References


Examples

```r
T1 <- xtabs(~race + drug, data = Inmate)
T1
chisq.test(T1)
rm(T1)
```
Inspect

<table>
<thead>
<tr>
<th>Percent of vehicles passing inspection by type inspection station</th>
</tr>
</thead>
</table>

**Description**

Data for Exercise 8.59

**Usage**

Inspect

**Format**

A data frame/tibble with 174 observations on two variables

- `station` a factor with levels auto inspection, auto repair, car care center, gas station, new car dealer, and tire store
- `passed` a factor with levels less than 70%, between 70% and 84%, and more than 85%

**Source**


**References**


**Examples**

```r
T1 <- xtabs(~ station + passed, data = Inspect)
T1
barplot(T1, beside = TRUE, legend = TRUE)
chisq.test(T1)
rm(T1)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Inspect, aes(x = passed, fill = station)) +
  geom_bar(position = "dodge") +
  theme_bw()

## End(Not run)
```
Insulate  

*Heat loss through a new insulating medium*

**Description**

Data for Exercise 9.50

**Usage**

`Insulate`

**Format**

A data frame/tibble with ten observations on two variables

- **temp** outside temperature (in degrees Celsius)
- **loss** heat loss (in BTUs)

**References**


**Examples**

```r
plot(loss ~ temp, data = Insulate)
model <- lm(loss ~ temp, data = Insulate)
abline(model, col = "blue")
summary(model)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Insulate, aes(x = temp, y = loss)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  theme_bw()

## End(Not run)
```
Iqgpa

GPA versus IQ for 12 individuals

Description

Data for Exercises 9.51 and 9.52

Usage

Iqgpa

Format

A data frame/tibble with 12 observations on two variables

iq  IQ scores
gpa  Grade point average

References


Examples

```r
plot(gpa ~ iq, data = Iqgpa, col = "blue", pch = 19)
model <- lm(gpa ~ iq, data = Iqgpa)
summary(model)
rm(model)
```

Irises

R.A. Fishers famous data on Irises

Description

Data for Examples 1.15 and 5.19

Usage

Irises
Format

A data frame/tibble with 150 observations on five variables

- **sepal_length** sepal length (in cm)
- **sepal_width** sepal width (in cm)
- **petal_length** petal length (in cm)
- **petal_width** petal width (in cm)
- **species** a factor with levels setosa, versicolor, and virginica

Source


References


Examples

```r
tapply(Irises$sepal_length, Irises$species, mean)
t.test(Irises$sepal_length[Irises$species == "setosa"], conf.level = 0.99)
hist(Irises$sepal_length[Irises$species == "setosa"],
    main = "Sepal length for\nIris Setosa",
    xlab = "Length (in cm)"
)boxplot(sepal_length ~ species, data = Irises)
```

**Jdpower**

*Number of problems reported per 100 cars in 1994 versus 1995s*

Description

Data for Exercise 2.14, 2.17, 2.31, 2.33, and 2.40

Usage

Jdpower
Format

A data frame/tibble with 29 observations on three variables

car a factor with levels Acura, BMW, Buick, Cadillac, Chevrolet, Dodge Eagle, Ford, Geo, Honda, Hyundai, Infiniti, Jaguar, Lexus, Lincoln, Mazda, Mercedes-Benz, Mercury, Mitsubishi, Nissan, Oldsmobile, Plymouth, Pontiac, Saab, Saturn, and Subaru, Toyota Volkswagen, Volvo

1994 number of problems per 100 cars in 1994

1995 number of problems per 100 cars in 1995

Source


References


Examples

```r
model <- lm(`1995` ~ `1994`, data = Jdpower)
summary(model)
plot(`1995` ~ `1994`, data = Jdpower)
abline(model, col = "red")
rm(model)
```

---

**Jobsat**

*Job satisfaction and stress level for 9 school teachers*

Description

Data for Exercise 9.60

Usage

Jobsat

Format

A data frame/tibble with nine observations on two variables

wspt Wilson Stress Profile score for teachers

satisfaction job satisfaction score
References


Examples

plot(satisfaction ~ wspt, data = Jobsat)
model <- lm(satisfaction ~ wspt, data = Jobsat)
abline(model, col = "blue")
summary(model)
rm(model)

kidsmoke

Smoking habits of boys and girls ages 12 to 18

Description

Data for Exercise 4.85

Usage

kidsmoke

Format

A data frame/tibble with 1000 observations on two variables

gender character vector with values female and male

smoke a character vector with values no and yes

References


Examples

T1 <- xtabs(~smoke + gender, data = Kidsmoke)
T1
prop.table(T1)
prop.table(T1, 1)
prop.table(T1, 2)
Kilowatt

Rates per kilowatt-hour for each of the 50 states and DC

Description
Data for Example 5.9

Usage
Kilowatt

Format
A data frame/tibble with 51 observations on two variables

<table>
<thead>
<tr>
<th>state</th>
<th>rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td></td>
</tr>
<tr>
<td>Alaska</td>
<td></td>
</tr>
<tr>
<td>Arizona</td>
<td></td>
</tr>
<tr>
<td>Arkansas</td>
<td></td>
</tr>
<tr>
<td>California</td>
<td></td>
</tr>
<tr>
<td>Colorado</td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td></td>
</tr>
<tr>
<td>Delaware</td>
<td></td>
</tr>
<tr>
<td>District of Columbia</td>
<td></td>
</tr>
<tr>
<td>Florida</td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td></td>
</tr>
<tr>
<td>Hawaii</td>
<td></td>
</tr>
<tr>
<td>Idaho</td>
<td></td>
</tr>
<tr>
<td>Illinois</td>
<td></td>
</tr>
<tr>
<td>Indiana</td>
<td></td>
</tr>
<tr>
<td>Iowa</td>
<td></td>
</tr>
<tr>
<td>Kansas</td>
<td></td>
</tr>
<tr>
<td>Kentucky</td>
<td></td>
</tr>
<tr>
<td>Louisiana</td>
<td></td>
</tr>
<tr>
<td>Maine</td>
<td></td>
</tr>
<tr>
<td>Maryland</td>
<td></td>
</tr>
<tr>
<td>Massachusetts</td>
<td></td>
</tr>
<tr>
<td>Michigan</td>
<td></td>
</tr>
<tr>
<td>Minnesota</td>
<td></td>
</tr>
<tr>
<td>Mississippi</td>
<td></td>
</tr>
<tr>
<td>Missouri</td>
<td></td>
</tr>
<tr>
<td>Montana</td>
<td></td>
</tr>
<tr>
<td>Nebraska</td>
<td></td>
</tr>
<tr>
<td>Nevada</td>
<td></td>
</tr>
<tr>
<td>New Hampshire</td>
<td></td>
</tr>
<tr>
<td>New Jersey</td>
<td></td>
</tr>
<tr>
<td>New Mexico</td>
<td></td>
</tr>
<tr>
<td>New York</td>
<td></td>
</tr>
<tr>
<td>North Carolina</td>
<td></td>
</tr>
<tr>
<td>North Dakota</td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td></td>
</tr>
<tr>
<td>Oklahoma</td>
<td></td>
</tr>
<tr>
<td>Oregon</td>
<td></td>
</tr>
<tr>
<td>Pennsylvania</td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td></td>
</tr>
<tr>
<td>South Carolina</td>
<td></td>
</tr>
<tr>
<td>South Dakota</td>
<td></td>
</tr>
<tr>
<td>Tennessee</td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td></td>
</tr>
<tr>
<td>Utah</td>
<td></td>
</tr>
<tr>
<td>Vermont</td>
<td></td>
</tr>
<tr>
<td>Virginia</td>
<td></td>
</tr>
<tr>
<td>Washington</td>
<td></td>
</tr>
<tr>
<td>West Virginia</td>
<td></td>
</tr>
<tr>
<td>Wisconsin</td>
<td></td>
</tr>
<tr>
<td>Wyoming</td>
<td></td>
</tr>
</tbody>
</table>

rate a numeric vector indicating rates for kilowatt per hour

References

Examples
EDA(Kilowatt$rate)

Kinder

Reading scores for first grade children who attended kindergarten versus those who did not

Description
Data for Exercise 7.68

Usage
Kinder
Laminect

Format
A data frame/tibble with eight observations on three variables

- **pair** a numeric indicator of pair
- **kinder** reading score of kids who went to kindergarten
- **nokinder** reading score of kids who did not go to kindergarten

References

Examples

```r
boxplot(Kinder$kinder, Kinder$nokinder)
diff <- Kinder$kinder - Kinder$nokinder
qnorm(diff)
qqline(diff)
shapiro.test(diff)
t.test(Kinder$kinder, Kinder$nokinder, paired = TRUE)
# Or
t.test(diff)
rm(diff)
```

---

Laminect

*Median costs of laminectomies at hospitals across North Carolina in 1992*

Description
Data for Exercise 10.18

Usage
Laminect

Format
A data frame/tibble with 138 observations on two variables

- **area** a character vector indicating the area of the hospital with Rural, Regional, and Metropol
- **cost** a numeric vector indicating cost of a laminectomy

Source

Lead

References


Examples

```r
boxplot(cost ~ area, data = Laminect, col = topo.colors(3))
anova(lm(cost ~ area, data = Laminect))
```

---

**Description**

Data for Example 1.17

**Usage**

Lead

**Format**

A data frame/tibble with 66 observations on the two variables

- **group**: a character vector with values exposed and control
- **lead**: a numeric vector indicating the level of lead in children’s blood (in micrograms/dl)

**Source**


**References**


**Examples**

```r
boxplot(lead ~ group, data = Lead, col = topo.colors(2))
```
**Leader**  
*Leadership exam scores by age for employees on an industrial plant*

**Description**
Data for Exercise 7.31

**Usage**
Leader

**Format**
A data frame/tibble with 34 observations on two variables

- **age** a character vector indicating age with values under 35 and over 35
- **score** score on a leadership exam

**References**

**Examples**

```r
boxplot(score ~ age, data = Leader, col = c("gray", "green"))
t.test(score ~ age, data = Leader)
```

---

**Lethal**  
*Survival time of mice injected with an experimental lethal drug*

**Description**
Data for Example 6.12

**Usage**
Lethal

**Format**
A data frame/tibble with 30 observations on one variable

- **survival** a numeric vector indicating time survived after injection (in seconds)
References


Examples

```r
SIGN.test(Lethal$survival, md = 45, alternative = "less")
```

---

**Life**  
*Life expectancy of men and women in U.S.*

Description

Data for Exercise 1.31

Usage

```r
Life
```

Format

A data frame/tibble with eight observations on three variables

- `year`: a numeric vector indicating year
- `men`: life expectancy for men (in years)
- `women`: life expectancy for women (in years)

Source

National Center for Health Statistics.

References


Examples

```r
plot(men ~ year, type = "l", ylim = c(min(men, women), max(men, women)),
     col = "blue", main = "Life Expectancy vs Year", ylab = "Age",
     xlab = "Year", data = Life)
lines(women ~ year, col = "red", data = Life)
text(1955, 65, "Men", col = "blue")
text(1955, 70, "Women", col = "red")
```
Lifespan

*Life span of electronic components used in a spacecraft versus heat*

### Description

Data for Exercise 2.4, 2.37, and 2.49

### Usage

Lifespan

### Format

A data frame/tibble with six observations two variables

- **heat** temperature (in Celcius)
- **life** lifespan of component (in hours)

### References


### Examples

```r
plot(life ~ heat, data = Lifespan)
model <- lm(life ~ heat, data = Lifespan)
abline(model, col = "red")
resid(model)
sum((resid(model))^2)
anova(model)
rm(model)
```

---

Ligntmonth

*Relationship between damage reports and deaths caused by lightning*

### Description

Data for Exercise 2.6

### Usage

Ligntmonth
Lodge

**Format**

A data frame/tibble with 12 observations on four variables


- **deaths** number of deaths due to lightning strikes

- **injuries** number of injuries due to lightning strikes

- **damage** damage due to lightning strikes (in dollars)

**Source**


**References**


**Examples**

```r
plot(deaths ~ damage, data = Lightmonth)
model = lm(deaths ~ damage, data = Lightmonth)
abline(model, col = "red")
rm(model)
```

<table>
<thead>
<tr>
<th>Lodge</th>
<th>Measured traffic at three prospective locations for a motor lodge</th>
</tr>
</thead>
</table>

**Description**

Data for Exercise 10.33

**Usage**

Lodge

**Format**

A data frame/tibble with 45 observations on six variables

- **traffic** a numeric vector indicating the amount of vehicles that passed a site in 1 hour

- **site** a numeric vector with values 1, 2, and 3

- **ranks** ranks for variable traffic
References


Examples

```r
boxplot(traffic ~ site, data = Lodge, col = cm.colors(3))
anova(lm(traffic ~ factor(site), data = Lodge))
```

---

Longtail  
*Long-tailed distributions to illustrate Kruskal Wallis test*

Description

Data for Exercise 10.45

Usage

Longtail

Format

A data frame/tibble with 60 observations on three variables

- **score**: a numeric vector
- **group**: a numeric vector with values 1, 2, and 3
- **ranks**: ranks for variable **score**

References


Examples

```r
boxplot(score ~ group, data = Longtail, col = heat.colors(3))
kruskal.test(score ~ factor(group), data = Longtail)
anova(lm(score ~ factor(group), data = Longtail))
```
Lowabil

Reading skills of 24 matched low ability students

Description

Data for Example 7.18

Usage

Lowabil

Format

A data frame/tibble with 12 observations on three variables

- **pair** a numeric indicator of pair
- **experiment** score of the child with the experimental method
- **control** score of the child with the standard method

References


Examples

```r
diff = Lowabil$experiment - Lowabil$control
qqnorm(diff)
qqline(diff)
shapiro.test(diff)
t.test(Lowabil$experiment, Lowabil$control, paired = TRUE)
# OR
t.test(diff)
rm(diff)
```

Magnesiu

Magnesium concentration and distances between samples

Description

Data for Exercise 9.9

Usage

Magnesiu
Malpract

Format
A data frame/tibble with 20 observations on two variables

- **distance** distance between samples
- **magnesium** concentration of magnesium

Source

References

Examples

```r
plot(magnesium ~ distance, data = Magnesiu)
model = lm(magnesium ~ distance, data = Magnesiu)
abline(model, col = "red")
summary(model)
rm(model)
```

---

**Malpract**

*Amounts awarded in 17 malpractice cases*

Description
Data for Exercise 5.73

Usage
Malpract

Format
A data frame/tibble with 17 observations on one variable

- **award** malpractice reward (in $1000)

References
Examples

SIGN.test(Malpract$award, conf.level = 0.90)

Manager

Advertised salaries offered general managers of major corporations in 1995

Description

Data for Exercise 5.81

Usage

Manager

Format

A data frame/tibble with 26 observations on one variable

salary random sample of advertised annual salaries of top executives (in dollars)

References


Examples

stem(Manager$salary)
SIGN.test(Manager$salary)

Marked

Percent of marked cars in 65 police departments in Florida

Description

Data for Exercise 6.100

Usage

Marked
**Math**

**Format**

A data frame/tibble with 65 observations on one variable

- **percent** percentage of marked cars in 65 Florida police departments

**Source**


**References**


**Examples**

```r
EDA(Marked$percent)
SIGN.test(Marked$percent, md = 60, alternative = "greater")
t.test(Marked$percent, mu = 60, alternative = "greater")
```

---

**Math**  

<table>
<thead>
<tr>
<th>Math</th>
<th>Standardized math test scores for 30 students</th>
</tr>
</thead>
</table>

**Description**

Data for Exercise 1.69

**Usage**

Math

**Format**

A data frame/tibble with 30 observations on one variable

- **score** scores on a standardized test for 30 tenth graders

**References**

Examples

```r
stem(Math$score)
hist(Math$score, main = "Math Scores", xlab = "score", freq = FALSE)
lines(density(Math$score), col = "red")
CharlieZ <- (62 - mean(Math$score))/sd(Math$score)
CharlieZ
scale(Math$score)[which(Math$score == 62)]
```

Mathcomp

<table>
<thead>
<tr>
<th>Mathcomp</th>
<th>Standardized math competency for a group of entering freshmen at a small community college</th>
</tr>
</thead>
</table>

Description

Data for Exercise 5.26

Usage

Mathcomp

Format

A data frame/tibble with 31 observations one variable

`score` scores of 31 entering freshmen at a community college on a national standardized test

References


Examples

```r
stem(Mathcomp$score)
EDA(Mathcomp$score)
```
Description

Data for Exercise 9.24, Example 9.1, and Example 9.6

Usage

Mathpro

Format

A data frame/tibble with 51 observations on four variables

state a factor with levels Conn, D.C., Del, Ga, Hawaii, Ind, Maine, Mass, Md, N.C., N.H., N.J., N.Y., Ore, Pa, R.I., S.C., Va, and Vt

sat_math SAT math scores for high school seniors

profic math proficiency scores for eighth graders

group a numeric vector

Source

National Assessment of Educational Progress and The College Board.

References


Examples

model <- lm(sat_math ~ profic, data = Mathpro)
plot(sat_math ~ profic, data = Mathpro, ylab = "SAT", xlab = "proficiency")
abline(model, col = "red")
summary(model)
rm(model)
Maze

Error scores for four groups of experimental animals running a maze

Description

Data for Exercise 10.13

Usage

Maze

Format

A data frame/tibble with 32 observations on two variables

- **score**: error scores for animals running through a maze under different conditions
- **condition**: a factor with levels CondA, CondB, CondC, and CondD

References


Examples

```r
boxplot(score ~ condition, data = Maze, col = rainbow(4))
anova(lm(score ~ condition, data = Maze))
```

Median

Illustrates test of equality of medians with the Kruskal Wallis test

Description

Data for Exercise 10.52

Usage

Median

Format

A data frame/tibble with 45 observations on two variables

- **sample**: a vector with values Sample1, Sample 2, and Sample 3
- **value**: a numeric vector
**References**


**Examples**

```r
boxplot(value ~ sample, data = Mental, col = rainbow(3))
anova(lm(value ~ sample, data = Mental))
kruskal.test(value ~ factor(sample), data = Mental)
```

<table>
<thead>
<tr>
<th>Mental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median mental ages of 16 girls</td>
</tr>
</tbody>
</table>

**Description**

Data for Exercise 6.52

**Usage**

```r
Mental
```

**Format**

A data frame/tibble with 16 observations on one variable

- **age** mental age of 16 girls

**References**


**Examples**

```r
SIGN.test(Mental$age, md = 100)
```
**Mercury**

*Concentration of mercury in 25 lake trout*

**Description**

Data for Example 1.9

**Usage**

`Mercury`

**Format**

A data frame/tibble with 25 observations on one variable

- `mercury` a numeric vector measuring mercury (in parts per million)

**References**


**Examples**

```r
stem(Mercury$mercury)
```

**Metrent**

*Monthly rental costs in metro areas with 1 million or more persons*

**Description**

Data for Exercise 5.117

**Usage**

`Metrent`

**Format**

A data frame/tibble with 46 observations on one variable

- `rent` monthly rent in dollars
Source

References

Examples

```r
boxplot(Metrent$rent, col = "magenta")
t.test(Metrent$rent, conf.level = 0.99)$conf
```

<table>
<thead>
<tr>
<th>Miller</th>
<th>Miller personality test scores for a group of college students applying for graduate school</th>
</tr>
</thead>
</table>

Description
Data for Example 5.7

Usage
Miller

Format
A data frame/tibble with 25 observations on one variable

miller scores on the Miller Personality test

References

Examples

```r
stem(Miller$miller)
fivenum(Miller$miller)
boxplot(Miller$miller)
qqnorm(Miller$miller,col = "blue")
qqline(Miller$miller, col = "red")
```
**Miller1**

*Twenty scores on the Miller personality test*

**Description**
Data for Exercise 1.41

**Usage**
Miller1

**Format**
A data frame/tibble with 20 observations on one variable

miller scores on the Miller personality test

**References**

**Examples**

```r
stem(Miller1$miller)
stem(Miller1$miller, scale = 2)
```

---

**Moisture**

*Moisture content and depth of core sample for marine muds in eastern Louisiana*

**Description**
Data for Exercise 9.32

**Usage**
Moisture

**Format**
A data frame/tibble with 16 observations on four variables

depth a numeric vector
moisture g of water per 100 g of dried sediment
lnmoist a numeric vector
depthsq a numeric vector
Monoxide

Source


References


Examples

```r
plot(moisture ~ depth, data = Moisture)
model <- lm(moisture ~ depth, data = Moisture)
abline(model, col = "red")
plot(resid(model) ~ depth, data = Moisture)
rm(model)
```

---

| Monoxide | Carbon monoxide emitted by smoke stacks of a manufacturer and a competitor |

Description

Data for Exercise 7.45

Usage

`Monoxide`

Format

A data frame/tibble with ten observations on two variables

- **company**: a vector with values `manufacturer` and `competitor`
- **emission**: carbon monoxide emitted

References

Examples

```r
boxplot(emission ~ company, data = Monoxide, col = topo.colors(2))
t.test(emission ~ company, data = Monoxide)
wilcox.test(emission ~ company, data = Monoxide)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Monoxide, aes(x = company, y = emission)) +
  geom_boxplot() +
  theme_bw()
## End(Not run)
```

Movie

Moral attitude scale on 15 subjects before and after viewing a movie

Description

Data for Exercise 7.53

Usage

Movie

Format

A data frame/tibble with 12 observations on three variables

- **before**: moral aptitude before viewing the movie
- **after**: moral aptitude after viewing the movie
- **differ**: a numeric vector

References


Examples

```r
qqnorm(Movie$differ)
qqline(Movie$differ)
shapiro.test(Movie$differ)
t.test(Movie$after, Movie$before, paired = TRUE, conf.level = 0.99)
wilcox.test(Movie$after, Movie$before, paired = TRUE)
```
Description

Data for Exercise 7.59

Usage

Music

Format

A data frame/tibble with 12 observations on three variables

- **method1** a numeric vector measuring the improvement scores on a music recognition test
- **method2** a numeric vector measuring the improvement scores on a music recognition test
- **differ** method1 - method2

References


Examples

```r
qqnorm(Music$differ)
qqline(Music$differ)
shapiro.test(Music$differ)
t.test(Music$method1, Music$method2, paired = TRUE)
# Or
t.test(Music$differ)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Music, aes(x = differ)) +
geom_dotplot() +
theme_bw()
## End(Not run)
```
Name

Estimated value of a brand name product and the company’s revenue

Description

Data for Exercises 2.28, 9.19, and Example 2.8

Usage

Name

Format

A data frame/tibble with 42 observations on three variables

- **brand** a factor with levels Band-Aid, Barbie, Birds Eye, Budweiser, Camel, Campbell, Carlsberg, Coca-Cola, Colgate, Del Monte, Fisher-Price, Gordon’s, Green Giant, Guinness, Haagen-Dazs, Heineken, Heinz, Hennessy, Hermes, Hershey, Ivory, Jell-o, Johnnie Walker, Kellogg, Kleenex, Kraft, Louis Vuitton, Marlboro, Nescafe, Nestle, Nivea, Oil of Olay, Pampers, Pepsi-Cola, Planters, Quaker, Sara Lee, Schweppes, Smirnoff, Tampax, Winston, and Wrigley's

- **value** value in billions of dollars

- **revenue** revenue in billions of dollars

Source

Financial World.

References


Examples

```r
plot(value ~ revenue, data = Name)
model <- lm(value ~ revenue, data = Name)
abline(model, col = "red")
cor(Name$value, Name$revenue)
summary(model)
rm(model)
```
### Nascar

**Description**

Data for Exercise 10.53

**Usage**

Nascar

**Format**

A data frame/tibble with 36 observations on six variables

- **time**: duration of pit stop (in seconds)
- **team**: a numeric vector representing team 1, 2, or 3
- **ranks**: a numeric vector ranking each pit stop in order of speed

**References**


**Examples**

```r
boxplot(time ~ team, data = Nascar, col = rainbow(3))
model <- lm(time ~ factor(team), data = Nascar)
summary(model)
anova(model)
rm(model)
```

### Nervous

**Description**

Data for Example 10.3

**Usage**

Nervous
**Newsstand**

**Format**

A data frame/tibble with 25 observations on two variables

- **react** a numeric vector representing reaction time
- **drug** a numeric vector indicating each of the 4 drugs

**References**


**Examples**

```r
boxplot(react ~ drug, data = Nervous, col = rainbow(4))
model <- aov(react ~ factor(drug), data = Nervous)
summary(model)
TukeyHSD(model)
plot(TukeyHSD(model), las = 1)
```

---

**Newsstand**

**Daily profits for 20 newsstands**

**Description**

Data for Exercise 1.43

**Usage**

Newsstand

**Format**

A data frame/tibble with 20 observations on one variable

- **profit** profit of each newsstand (in dollars)

**References**


**Examples**

```r
stem(Newsstand$profit)
stem(Newsstand$profit, scale = 3)
```
### Nfldraf2

**Rating, time in 40-yard dash, and weight of top defensive linemen in the 1994 NFL draft**

**Description**

Data for Exercise 9.63

**Usage**

Nfldraf2

**Format**

A data frame/tibble with 47 observations on three variables

- **rating**: rating of each player on a scale out of 10
- **forty**: forty yard dash time (in seconds)
- **weight**: weight of each player (in pounds)

**References**


**Examples**

```r
plot(rating ~ forty, data = Nfldraf2)
summary(lm(rating ~ forty, data = Nfldraf2))
```

### Nfldraft

**Rating, time in 40-yard dash, and weight of top offensive linemen in the 1994 NFL draft**

**Description**

Data for Exercises 9.10 and 9.16

**Usage**

Nfldraft
Nicotine

Format
A data frame/tibble with 29 observations on three variables

rating rating of each player on a scale out of 10
forty forty yard dash time (in seconds)
weight weight of each player (in pounds)

Source
USA Today, April 20, 1994.

References

Examples
plot(rating ~ forty, data = Nfldraft)
cor(Nfldraft$rating, Nfldraft$forty)
summary(lm(rating ~ forty, data = Nfldraft))
Examples

model <- lm(sales ~ nicotine, data = Nicotine)
plot(sales ~ nicotine, data = Nicotine)
abline(model, col = "red")
summary(model)
predict(model, newdata = data.frame(nicotine = 1),
       interval = "confidence", level = 0.99)

normarea  Normal Area

Description

Function that computes and draws the area between two user specified values in a user specified normal distribution with a given mean and standard deviation

Usage

normarea(lower = -Inf, upper = Inf, m, sig)

Arguments

lower  the lower value
upper  the upper value
m      the mean for the population
sig    the standard deviation of the population

Author(s)

Alan T. Arnholt

Examples

normarea(70, 130, 100, 15)
# Finds and P(70 < X < 130) given X is N(100,15).
Description

Function to determine required sample size to be within a given margin of error.

Usage

\[ \text{nsize}(b, \text{sigma} = \text{NULL}, p = 0.5, \text{conf.level} = 0.95, \text{type} = \text{"mu"}) \]

Arguments

- `b`: the desired bound.
- `sigma`: population standard deviation. Not required if using type "pi".
- `p`: estimate for the population proportion of successes. Not required if using type "mu".
- `conf.level`: confidence level for the problem, restricted to lie between zero and one.
- `type`: character string, one of "mu" or "pi", or just the initial letter of each, indicating the appropriate parameter. Default value is "mu".

Details

Answer is based on a normal approximation when using type "pi".

Value

Returns required sample size.

Author(s)

Alan T. Arnholt

Examples

\[
\text{nsize}(b=0.03, p=708/1200, \text{conf.level}=0.90, \text{type}=\text{"pi"})
\]

# Returns the required sample size (n) to estimate the population proportion of successes with a 0.9 confidence interval so that the margin of error is no more than 0.03 when the estimate of the population proportion of successes is 708/1200. This is problem 5.38 on page 257 of Kitchen's BSDA.

\[
\text{nsize}(b=0.15, \text{sigma}=0.31, \text{conf.level}=0.90, \text{type}=\text{"mu"})
\]

# Returns the required sample size (n) to estimate the population mean with a 0.9 confidence interval so that the margin of error is no more than 0.15. This is Example 5.17 on page 261 of Kitchen's BSDA.
Normality Tester

Description
Q-Q plots of randomly generated normal data of the same size as the tested data are generated and plotted on the perimeter of the graph while a Q-Q plot of the actual data is depicted in the center of the graph.

Usage
ntester(actual.data)

Arguments

actual.data a numeric vector. Missing and infinite values are allowed, but are ignored in the calculation. The length of actual.data must be less than 5000 after dropping nonfinite values.

Details
Q-Q plots of randomly generated normal data of the same size as the tested data are generated and plotted on the perimeter of the graph sheet while a Q-Q plot of the actual data is depicted in the center of the graph. The p-values are calculated form the Shapiro-Wilk W-statistic. Function will only work on numeric vectors containing less than or equal to 5000 observations.

Author(s)

Alan T. Arnholt

References

Examples

ntester(rexp(50,1))
# Q-Q plot of random exponential data in center plot
# surrounded by 8 Q-Q plots of randomly generated
# standard normal data of size 50.
**Orange**

*Price of oranges versus size of the harvest*

---

**Description**

Data for Exercise 9.61

**Usage**

Orange

**Format**

A data frame/tibble with six observations on two variables

- **harvest**: harvest in millions of boxes
- **price**: average price charged by California growers for a 75-pound box of navel oranges

**References**


**Examples**

```r
plot(price ~ harvest, data = Orange)
model <- lm(price ~ harvest, data = Orange)
abline(model, col = "red")
summary(model)
rm(model)
```

---

**Orioles**

*Salaries of members of the Baltimore Orioles baseball team*

---

**Description**

Data for Example 1.3

**Usage**

Orioles
Oxytocin

Format

A data frame/tibble with 27 observations on three variables

**first name** a factor with levels Albert, Arthur, B.J., Brady, Cal, Charles, dl-Delino, dl-Scott, Doug, Harold, Heathcliff, Jeff, Jesse, Juan, Lenny, Mike, Rich, Ricky, Scott, Sidney, Will, and Willis

**last name** a factor with levels Amaral, Anderson, Baines, Belle, Bones, Bordick, Clark, Conine, Deshields, Erickson, Fetters, Garcia, Guzman, Johns, Johnson, Kamieniecki, Mussina, Orosco, Otanez, Ponson, Reboulet, Rhodes, Ripken Jr., Slocumb, Surhoff, Timlin, and Webster

**1999salary** a numeric vector containing each player’s salary (in dollars)

References


Examples

```r
stripchart(Orioles$`1999salary`, method = "stack", pch = 19)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Orioles, aes(x = `1999salary`)) +
  geom_dotplot(dotsize = 0.5) +
  labs(x = "1999 Salary") +
  theme_bw()
## End(Not run)
```

---

<table>
<thead>
<tr>
<th>Oxytocin</th>
<th>Arterial blood pressure of 11 subjects before and after receiving oxytocin</th>
</tr>
</thead>
</table>

Description

Data for Exercise 7.86

Usage

Oxytocin

Format

A data frame/tibble with 11 observations on three variables

**subject** a numeric vector indicating each subject

**before** mean arterial blood pressure of subject before receiving oxytocin

**after** mean arterial blood pressure of subject after receiving oxytocin
References


Examples

```r
diff = Oxytocin$after - Oxytocin$before
qqnorm(diff)
qqline(diff)
shapiro.test(diff)
t.test(Oxytocin$after, Oxytocin$before, paired = TRUE)
rm(diff)
```

---

Parented

*Education backgrounds of parents of entering freshmen at a state university*

Description

Data for Exercise 1.32

Usage

`Parented`

Format

A data frame/tibble with 200 observations on two variables

- **education**: a factor with levels `4yr college degree`, `Doctoral degree`, `Grad degree`, `H.S grad or less`, `Some college`, and `Some grad school`
- **parent**: a factor with levels `mother` and `father`

References


Examples

```r
T1 <- xtabs(~education + parent, data = Parented)
T1
barplot(t(T1), beside = TRUE, legend = TRUE, col = c("blue", "red"))
rm(T1)
```

---

## Not run:

```r
library(ggplot2)
```
ggplot2::ggplot(data = Parented, aes(x = education, fill = parent)) +
  geom_bar(position = "dodge") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 85, vjust = 0.5)) +
  scale_fill_manual(values = c("pink", "blue")) +
  labs(x = ", y = ")

## End(Not run)

<table>
<thead>
<tr>
<th>Patrol</th>
<th>Years of experience and number of tickets given by patrolpersons in New York City</th>
</tr>
</thead>
</table>

**Description**

Data for Example 9.3

**Usage**

Patrol

**Format**

A data frame/tibble with ten observations on three variables

- **tickets** number of tickets written per week
- **years** patrolperson’s experience (in years)
- **log_tickets** natural log of tickets

**References**


**Examples**

```r
model <- lm(tickets ~ years, data = Patrol)
summary(model)
confint(model, level = 0.98)
```
Pearson

Karl Pearson's data on heights of brothers and sisters

Description
Data for Exercise 2.20

Usage
Pearson

Format
A data frame/tibble with 11 observations on three variables

family number indicating family of brother and sister pair
brother height of brother (in inches)
sister height of sister (in inches)

Source

References

Examples
plot(brother ~ sister, data = Pearson, col = "lightblue")
cor(Pearson$brother, Pearson$sister)

Phone

Length of long-distance phone calls for a small business firm

Description
Data for Exercise 6.95

Usage
Phone
Poison

Format

A data frame/tibble with 20 observations on one variable

time duration of long distance phone call (in minutes)

References


Examples

```r
time <- c(120, 135, 150, 165, 180, 200, 210, 220, 230, 240)
qqnorm(time)
qqline(time)
shapiro.test(time)
SIGN.test(time, md = 5, alternative = "greater")
```

---

Poison

Number of poisonings reported to 16 poison control centers

Description

Data for Exercise 1.113

Usage

Poison

Format

A data frame/tibble with 226,361 observations on one variable

type a factor with levels Alcohol, Cleaning agent, Cosmetics, Drugs, Insecticides, and Plants

Source

Centers for Disease Control, Atlanta, Georgia.

References

Examples

```r
T1 <- xtabs(~type, data = Poison)
T1
par(mar = c(5.1 + 2, 4.1, 4.1, 2.1))
barplot(sort(T1, decreasing = TRUE), las = 2, col = rainbow(6))
par(mar = c(5.1, 4.1, 4.1, 2.1))
rm(T1)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Poison, aes(x = type, fill = type)) +
  geom_bar() +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 85, vjust = 0.5)) +
  guides(fill = FALSE)
## End(Not run)
```

---

**Politic**

*Political party and gender in a voting district*

### Description

Data for Example 8.3

### Usage

```r
Politic
```

### Format

A data frame/tibble with 250 observations on two variables

- **party** a factor with levels republican, democrat, and other
- **gender** a factor with levels female and male

### References


### Examples

```r
T1 <- xtabs(~party + gender, data = Politic)
T1
chisq.test(T1)
rm(T1)
```
Pollutio

Air pollution index for 15 randomly selected days for a major western city

Description
Data for Exercise 5.59

Usage
Pollutio

Format
A data frame/tibble with 15 observations on one variable

inde air pollution index

References

Examples

stem(Pollutio$inde)
t.test(Pollutio$inde, conf.level = 0.98)$conf

Porosity

Porosity measurements on 20 samples of Tensleep Sandstone, Pennsylvanian from Bighorn Basin in Wyoming

Description
Data for Exercise 5.86

Usage
Porosity

Format
A data frame/tibble with 20 observations on one variable

porosity porosity measurement (percent)
Poverty

Source

References

Examples

```r
glimpse(Porosity)
pov <- Poverty[, c(3,4)]
```

```
source <- c('Atlanta', 'Buffalo', 'Cincinnati', 'Cleveland', 'Dayton, O', 'Detroit', 'Flint, Mich', 'Fresno, C.Gary', 'Ind', 'Hartford, C', 'Laredo, Macon, Ga', 'Miami, Milwaukee', 'New Orleans, Newark, NJ', 'Rochester, NY', 'Shreveport, St. Louis, and Waco, Tx')
poverty <- as.character(pov$poverty)
crime <- as.character(pov$crime)
```

Description
Data for Exercise 9.11 and 9.17

Usage
Poverty

Format
A data frame/tibble with 20 observations on four variables

- **city**: a factor with levels Atlanta, Buffalo, Cincinnati, Cleveland, Dayton, O, Detroit, Flint, Mich, Fresno, C.Gary, Ind, Hartford, C, Laredo, Macon, Ga, Miami, Milwaukee, New Orleans, Newark, NJ, Rochester, NY, Shreveport, St. Louis, and Waco, Tx
- **poverty**: percent of children living in poverty
- **crime**: crime rate (per 1000 people)
- **population**: population of city

Source
Children’s Defense Fund and the Bureau of Justice Statistics.

References
Examples

```r
plot(poverty ~ crime, data = Poverty)
model <- lm(poverty ~ crime, data = Poverty)
abline(model, col = "red")
summary(model)
rm(model)
```

<table>
<thead>
<tr>
<th>Precinct</th>
<th>Robbery rates versus percent low income in eight precincts</th>
</tr>
</thead>
</table>

Description

Data for Exercise 2.2 and 2.38

Usage

`Precinct`

Format

A data frame/tibble with eight observations on two variables

- **rate**: robbery rate (per 1000 people)
- **income**: percent with low income

References


Examples

```r
plot(rate ~ income, data = Precinct)
model <- lm(rate ~ income, data = Precinct)
abline(model, col = "red")
rm(model)
```
Prejudic

Racial prejudice measured on a sample of 25 high school students

Description

Data for Exercise 5.10 and 5.22

Usage

Prejudic

Format

A data frame with 25 observations on one variable

prejud racial prejudice score

References


Examples

stem(Prejudic$prejud)
EDA(Prejudic$prejud)

Presiden

Ages at inauguration and death of U.S. presidents

Description

Data for Exercise 1.126

Usage

Presiden
Press

Degree of confidence in the press versus education level for 20 randomly selected persons

Description
Data for Exercise 9.55

Usage
Press

Format
A data frame/tibble with 20 observations on two variables

education_yrs years of education
confidence degree of confidence in the press (the higher the score, the more confidence)

References

Examples

```r
pie(xtabs(~birth_state, data = Presiden))
stem(Presiden$inaugural_age)
stem(Presiden$death_age)
par(mar = c(5.1, 4.1 + 3, 4.1, 2.1))
stripchart(x=list(Presiden$inaugural_age, Presiden$death_age),
    method = "stack", col = c("green","brown"), pch = 19, las = 1)
par(mar = c(5.1, 4.1, 4.1, 2.1))
```
References


Examples

```r
plot(confidence ~ education_yrs, data = Press)
model <- lm(confidence ~ education_yrs, data = Press)
abline(model, col = "purple")
summary(model)
rm(model)
```

---

Prognost

*Kloper's prognostic rating scale for subjects receiving behavior modification therapy*

Description

Data for Exercise 6.61

Usage

`Prognost`

Format

A data frame/tibble with 15 observations on one variable

- **kprs_score** Kloper’s Prognostic Rating Scale score

Source


References


Examples

```r
EDA(Prognost$kprs_score)
t.test(Prognost$kprs_score, mu = 9)
```
Program

Effects of four different methods of programmed learning for statistics students

Description

Data for Exercise 10.17

Usage

Program

Format

A data frame/tibble with 44 observations on two variables

method  a character variable with values method1, method2, method3, and method4
score   standardized test score

References


Examples

boxplot(score ~ method, col = c("red", "blue", "green", "yellow"), data = Program)
anova(lm(score ~ method, data = Program))
TukeyHSD(aov(score ~ method, data = Program))
par(mar = c(5.1, 4.1 + 4, 4.1, 2.1))
plot(TukeyHSD(aov(score ~ method, data = Program)), las = 1)
par(mar = c(5.1, 4.1, 4.1, 2.1))

Psat

PSAT scores versus SAT scores

Description

Data for Exercise 2.50

Usage

Psat
Psych

Format

A data frame/tibble with seven observations on the two variables

psat  PSAT score
sat   SAT score

References


Examples

model <- lm(sat ~ psat, data = Psat)
par(mfrow = c(1, 2))
plot(Psats$psat, resid(model))
plot(model, which = 1)
rm(model)
par(mfrow = c(1, 1))

---

Psych

Correct responses for 24 students in a psychology experiment

Description

Data for Exercise 1.42

Usage

Psych

Format

A data frame/tibble with 23 observations on one variable

score  number of correct responses in a psychology experiment

References


Examples

stem(Psych$score)
EDA(Psych$score)
Puerto

Weekly incomes of a random sample of 50 Puerto Rican families in Miami

Description
Data for Exercise 5.22 and 5.65

Usage
Puerto

Format
A data frame/tibble with 50 observations on one variable

income  weekly family income (in dollars)

References

Examples

stem(Puerto$income)
boxplot(Puerto$income, col = "purple")
t.test(Puerto$income,conf.level = .90)$conf

Quail

Plasma LDL levels in two groups of quail

Description
Data for Exercise 1.53, 1.77, 1.88, 5.66, and 7.50

Usage
Quail

Format
A data frame/tibble with 40 observations on two variables

group  a character variable with values placebo and treatment
level  low-density lipoprotein (LDL) cholestrol level
Quality

Source


References


Examples

```r
boxplot(level ~ group, data = Quail, horizontal = TRUE, xlab = "LDL Level",
col = c("yellow", "lightblue"))
```

---

**Quality**

*Quality control test scores on two manufacturing processes*

**Description**

Data for Exercise 7.81

**Usage**

```r
Quality
```

**Format**

A data frame/tibble with 15 observations on two variables

- **process** a character variable with values *Process1* and *Process2*
- **score** results of a quality control test

**References**


**Examples**

```r
boxplot(score ~ process, data = Quality, col = "lightgreen")
t.test(score ~ process, data = Quality)
```
Rainks

Rainfall in an area of west central Kansas and four surrounding counties

Description

Data for Exercise 9.8

Usage

Rainks

Format

A data frame/tibble with 35 observations on five variables

rain rainfall (in inches)
x1 rainfall (in inches)
x2 rainfall (in inches)
x3 rainfall (in inches)
x4 rainfall (in inches)

Source


References


Examples

```r
cor(Rainks)
model <- lm(rain ~ x2, data = Rainks)
summary(model)
```
Randd  
Research and development expenditures and sales of a large company

**Description**

Data for Exercise 9.36 and Example 9.8

**Usage**

Randd

**Format**

A data frame/tibble with 12 observations on two variables

- **rd** research and development expenditures (in million dollars)
- **sales** sales (in million dollars)

**References**


**Examples**

```r
plot(sales ~ rd, data = Randd)
model <- lm(sales ~ rd, data = Randd)
abline(model, col = "purple")
summary(model)
plot(model, which = 1)
rm(model)
```

Rat  
Survival times of 20 rats exposed to high levels of radiation

**Description**

Data for Exercise 1.52, 1.76, 5.62, and 6.44

**Usage**

Rat
**Ratings**

**Format**

A data frame/tibble with 20 observations on one variable

**survival_time** survival time in weeks for rats exposed to a high level of radiation

**Source**


**References**


**Examples**

```r
hist(Rat$survival_time)
norm(Rat$survival_time)
sline(Rat$survival_time)
summary(Rat$survival_time)
t.test(Rat$survival_time)
t.test(Rat$survival_time, mu = 100, alternative = "greater")
```

---

**Ratings**

*Grade point averages versus teacher’s ratings*

**Description**

Data for Example 2.6

**Usage**

Ratings

**Format**

A data frame/tibble with 250 observations on two variables

**rating** character variable with students’ ratings of instructor (A-F)

**gpa** students’ grade point average

**References**

**Examples**

```r
boxplot(gpa ~ rating, data = Ratings, xlab = "Student rating of instructor", ylab = "Student GPA")
```

```r
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Ratings, aes(x = rating, y = gpa, fill = rating)) +
  geom_boxplot() +
  theme_bw() +
  theme(legend.position = "none") +
  labs(x = "Student rating of instructor", y = "Student GPA")

## End(Not run)
```

---

**Reaction**

*Threshold reaction time for persons subjected to emotional stress*

---

**Description**

Data for Example 6.11

**Usage**

Reaction

**Format**

A data frame/tibble with 12 observations on one variable

- **time** threshold reaction time (in seconds) for persons subjected to emotional stress

**References**


**Examples**

```r
stem(Reaction$time)
SIGN.test(Reaction$time, md = 15, alternative = "less")
```
Reading

Standardized reading scores for 30 fifth graders

Description

Data for Exercise 1.72 and 2.10

Usage

Reading

Format

A data frame/tibble with 30 observations on four variables

- **score**  standardized reading test score
- **sorted**  sorted values of score
- **trimmed**  trimmed values of sorted
- **winsoriz**  winsorized values of score

References


Examples

```r
hist(Reading$score, main = "Exercise 1.72",
     col = "lightgreen", xlab = "Standardized reading score")
summary(Reading$score)
sd(Reading$score)
```

Readiq

Reading scores versus IQ scores

Description

Data for Exercises 2.10 and 2.53

Usage

Readiq
Referend

Format

A data frame/tibble with 14 observations on two variables

reading  reading achievement score
iq  IQ score

References


Examples

```r
plot(reading ~ iq, data = Readiq)
model <- lm(reading ~ iq, data = Readiq)
abline(model, col = "purple")
predict(model, newdata = data.frame(iq = c(100, 120)))
residuals(model)[c(6, 7)]
rm(model)
```

Referend  Opinion on referendum by view on freedom of the press

Description

Data for Exercise 8.20

Usage

Referend

Format

A data frame with 237 observations on two variables

choice  a factor with levels A, B, and C
response  a factor with levels for, against, and undecided

References

Examples

```r
t1 <- xtabs(~choice + response, data = Referend)
t1
chisq.test(t1)
chisq.test(t1)$expected
```

---

<table>
<thead>
<tr>
<th>Region</th>
<th>Pollution index taken in three regions of the country</th>
</tr>
</thead>
</table>

Description

Data for Exercise 10.26

Usage

Region

Format

A data frame/tibble with 48 observations on three variables

- **pollution** pollution index
- **region** region of a county (west, central, and east)
- **ranks** ranked values of pollution

References


Examples

```r
boxplot(pollution ~ region, data = Region, col = "gray")
anova(lm(pollution ~ region, data = Region))
```
Register

Maintenance cost versus age of cash registers in a department store

Description
Data for Exercise 2.3, 2.39, and 2.54

Usage
Register

Format
A data frame/tibble with nine observations on two variables

- **age** age of cash register (in years)
- **cost** maintenance cost of cash register (in dollars)

References

Examples

```r
plot(cost ~ age, data = Register)
model <- lm(cost ~ age, data = Register)
abline(model, col = "red")
predict(model, newdata = data.frame(age = c(5, 10)))
plot(model, which = 1)
rm(model)
```

Rehab

Rehabilitative potential of 20 prison inmates as judged by two psychiatrists

Description
Data for Exercise 7.61

Usage
Rehab
Remedial

Format

A data frame/tibble with 20 observations on four variables

- **inmate**: inmate identification number
- **psych1**: rating from first psychiatrist on the inmates rehabilitative potential
- **psych2**: rating from second psychiatrist on the inmates rehabilitative potential
- **differ**: psych1 - psych2

References


Examples

```r
boxplot(Rehab$differ)
qqnorm(Rehab$differ)
qqline(Rehab$differ)
t.test(Rehab$differ)
# Or
t.test(Rehab$psych1, Rehab$psych2, paired = TRUE)
```

---

Remedial

*Math placement test score for 35 freshmen females and 42 freshmen males*

Description

Data for Exercise 7.43

Usage

Remedial

Format

A data frame/tibble with 84 observations on two variables

- **gender**: a character variable with values female and male
- **score**: math placement score

References

Examples

```r
boxplot(score ~ gender, data = Remedial, 
        col = c("purple", "blue"))
t.test(score ~ gender, data = Remedial, conf.level = 0.98)
t.test(score ~ gender, data = Remedial, conf.level = 0.98)$conf
wilcox.test(score ~ gender, data = Remedial, 
             conf.int = TRUE, conf.level = 0.98)
```

---

### Rentals

**Weekly rentals for 45 apartments**

---

### Description

Data for Exercise 1.122

### Usage

Rentals

### Format

A data frame/tibble with 45 observations on one variable

- **rent**: weekly apartment rental price (in dollars)

### References


### Examples

```r
stem(Rentals$rent)
sum(Rentals$rent < mean(Rentals$rent) - 3*sd(Rentals$rent) | 
    Rentals$rent > mean(Rentals$rent) + 3*sd(Rentals$rent))
```
**Repair**

*Recorded times for repairing 22 automobiles involved in wrecks*

**Description**

Data for Exercise 5.77

**Usage**

Repair

**Format**

A data frame/tibble with 22 observations on one variable

- **time**  time to repair a wrecked in car (in hours)

**References**


**Examples**

```r
stem(Repair$time)
SIGN.test(Repair$time, conf.level = 0.98)
```

---

**Retail**

*Length of employment versus gross sales for 10 employees of a large retail store*

**Description**

Data for Exercise 9.59

**Usage**

Retail

**Format**

A data frame/tibble with 10 observations on two variables

- **months**  length of employment (in months)
- **sales**  employee gross sales (in dollars)
References


Examples

```r
plot(sales ~ months, data = Retail)
model <- lm(sales ~ months, data = Retail)
abline(model, col = "blue")
summary(model)
```

Oceanography data obtained at site 1 by scientist aboard the ship Ron Brown

Description

Data for Exercise 2.9

Usage

```r
Ronbrown1
```

Format

A data frame/tibble with 75 observations on two variables

- **depth**: ocean depth (in meters)
- **temperature**: ocean temperature (in Celsius)

References


Examples

```r
plot(temperature ~ depth, data = Ronbrown1, ylab = "Temperature")
```
### Ronbrown2

Oceanography data obtained at site 2 by scientist aboard the ship Ron Brown

**Description**

Data for Exercise 2.56 and Example 2.4

**Usage**

Ronbrown2

**Format**

A data frame/tibble with 150 observations on three variables

- **depth** ocean depth (in meters)
- **temperature** ocean temperature (in Celcius)
- **salinity** ocean salinity level

**References**


**Examples**

```r
plot(salinity ~ depth, data = Ronbrown2)
model <- lm(salinity ~ depth, data = Ronbrown2)
summary(model)
plot(model, which = 1)
rm(model)
```

---

### Rural

Social adjustment scores for a rural group and a city group of children

**Description**

Data for Example 7.16

**Usage**

Rural
Salary

Starting salaries for 25 new PhD psychologist

Description
Data for Exercise 3.66

Usage
Salary

Format
A data frame/tibble with 25 observations on one variable

salary starting salary for Ph.D. psychologists (in dollars)

References
Salinity

Examples

\[
\text{qqnorm(Salinity$salary, pch = 19, col = "purple")}
\]
\[
\text{qqline(Salinity$salary, col = "blue")}
\]

Salinity

Surface-water salinity measurements from Whitewater Bay, Florida

Description

Data for Exercise 5.27 and 5.64

Usage

Salinity

Format

A data frame/tibble with 48 observations on one variable

\textbf{salinity}  surface-water salinity value

Source


References


Examples

\[
\text{stem(Salinity$salinity)}
\]
\[
\text{qqnorm(Salinity$salinity, pch = 19, col = "purple")}
\]
\[
\text{qqline(Salinity$salinity, col = "blue")}
\]
\[
\text{t.test(Salinity$salinity, conf.level = 0.99)}
\]
\[
\text{t.test(Salinity$salinity, conf.level = 0.99)$conf}
\]
SAT scores, percent taking exam and state funding per student by state for 1994, 1995 and 1999

Description
Data for Statistical Insight Chapter 9

Usage
Sat

Format
A data frame/tibble with 102 observations on seven variables

- **state**: U.S. state
- **verbal**: verbal SAT score
- **math**: math SAT score
- **total**: combined verbal and math SAT score
- **percent**: percent of high school seniors taking the SAT
- **expend**: state expenditure per student (in dollars)
- **year**: year

Source

References

Examples

```r
Sat94 <- Sat[Sat$year == 1994, ]
Sat94
Sat99 <- subset(Sat, year == 1999)
Sat99
stem(Sat99$total)
plot(total ~ percent, data = Sat99)
model <- lm(total ~ percent, data = Sat99)
abline(model, col = "blue")
summary(model)
rm(model)
```
Scales

Description

Data for Exercise 10.34 and 10.49

Usage

Saving

Format

A data frame/tibble with 65 observations on two variables

par problem-asset-ratio for Savings & Loans that were listed as being financially troubled in 1992

state U.S. state

References


Examples

```r
boxplot(par ~ state, data = Saving, col = "red")
boxplot(par ~ state, data = Saving, log = "y", col = "red")
model <- aov(par ~ state, data = Saving)
summary(model)
plot(TukeyHSD(model))
kruskal.test(par ~ factor(state), data = Saving)
```

Scales

Description

Data for Exercise 1.89

Usage

Scales

Description

Readings obtained from a 100 pound weight placed on four brands of bathroom scales
Schizop2

**Format**

A data frame/tibble with 20 observations on two variables

- **brand**: variable indicating brand of bathroom scale (A, B, C, or D)
- **reading**: recorded value (in pounds) of a 100 pound weight

**References**


**Examples**

```r
boxplot(reading ~ brand, data = Scales, col = rainbow(4),
ylab = "Weight (lbs)"

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Scales, aes(x = brand, y = reading, fill = brand)) +
  geom_boxplot() +
  labs(y = "weight (lbs)") +
  theme_bw() +
  theme(legend.position = "none")

## End(Not run)
```

---

**Schizop2**

Exam scores for 17 patients to assess the learning ability of schizophrenics after taking a specified dose of a tranquilizer

**Description**

Data for Exercise 6.99

**Usage**

Schizop2

**Format**

A data frame/tibble with 17 observations on one variable

- **score**: schizophrenics score on a second standardized exam

**References**

Examples

```r
hist(Schizop2$score, xlab = "score on standardized test after a tranquilizer", 
main = "Exercise 6.99", breaks = 10, col = "orange")
EDA(Schizop2$score)
SIGN.test(Schizop2$score, md = 22, alternative = "greater")
```

---

**Schizoph**

*Standardized exam scores for 13 patients to investigate the learning ability of schizophrenics after a specified dose of a tranquilizer*

---

**Description**

Data for Example 6.10

**Usage**

`Schizoph`

**Format**

A data frame/tibble with 13 observations on one variable

- **score**  schizophrenia score on a standardized exam one hour after receiving a specified dose of a tranquilizer.

**References**


**Examples**

```r
hist(Schizoph$score, xlab = "score on standardized test", 
main = "Example 6.10", breaks = 10, col = "orange")
EDA(Schizoph$score)
t.test(Schizoph$score, mu = 20)
```
Description

Data for Exercise 8.24

Usage

Seatbelt

Format

A data frame/tibble with 86,759 observations on two variables

seatbelt a factor with levels No and Yes

injuries a factor with levels None, Minimal, Minor, or Major indicating the extent of the drivers injuries

Source


References


Examples

```r
T1 <- xtabs(~seatbelt + injuries, data = Seatbelt)
T1
chisq.test(T1)
rm(T1)
```
Selfdefe

Self-confidence scores for 9 women before and after instructions on self-defense

Description
Data for Example 7.19

Usage
Selfdefe

Format
A data frame/tibble with nine observations on three variables
- woman: number identifying the woman
- before: before the course self-confidence score
- after: after the course self-confidence score

References

Examples

```r
Selfdefe$differ <- Selfdefe$after - Selfdefe$before
Selfdefe

t.test(Selfdefe$differ, alternative = "greater")
t.test(Selfdefe$after, Selfdefe$before,
       paired = TRUE, alternative = "greater")
```

Senior

Reaction times of 30 senior citizens applying for drivers license renewals

Description
Data for Exercise 1.83 and 3.67

Usage
Senior
Format

A data frame/tibble with 31 observations on one variable

reaction reaction time for senior citizens applying for a driver’s license renewal

References


Examples

```r
stem(Senior$reaction)
fivenum(Senior$reaction)
boxplot(Senior$reaction, main = "Problem 1.83, part d",
       horizontal = TRUE, col = "purple")
```

Description

Data for Exercise 1.123

Usage

`Sentence`

Format

A data frame/tibble with 41 observations on one variable

months sentence length (in months) for prisoners convicted of homicide

Source


References

Examples

```r
stem(Sentence$months)  
ll <- mean(Sentence$months) - 2*sd(Sentence$months)  
ul <- mean(Sentence$months) + 2*sd(Sentence$months)  
limits <- c(ll, ul)  
limits  
rm(ll, ul, limits)
```

---

**Shkdrug**

*Effects of a drug and electroshock therapy on the ability to solve simple tasks*

Description

Data for Exercises 10.11 and 10.12

Usage

`Shkdrug`

Format

A data frame/tibble with 64 observations on two variables

- **treatment** type of treatment Drug/NoS, Drug/Shk, NoDg/NoS, or NoDrug/S
- **response** number of tasks completed in a 10-minute period

References


Examples

```r
boxplot(response ~ treatment, data = Shkdrug, col = "gray")  
model <- lm(response ~ treatment, data = Shkdrug)  
anova(model)  
rm(model)
```
**Shock**

**Effect of experimental shock on time to complete difficult task**

**Description**

Data for Exercise 10.50

**Usage**

Shock

**Format**

A data frame/tibble with 27 observations on two variables

- **group** grouping variable with values of Group1 (no shock), Group2 (medium shock), and Group3 (severe shock)

- **attempts** number of attempts to complete a task

**References**


**Examples**

```r
boxplot(attempts ~ group, data = Shock, col = "violet")
model <- lm(attempts ~ group, data = Shock)
anova(model)
rm(model)
```

---

**Shoplift**

**Sales receipts versus shoplifting losses for a department store**

**Description**

Data for Exercise 9.58

**Usage**

Shoplift
Format

A data frame/tibble with eight observations on two variables

- **sales** sales (in 1000 dollars)
- **loss** loss (in 100 dollars)

References


Examples

```r
plot(loss ~ sales, data = Shoplift)
model <- lm(loss ~ sales, data = Shoplift)
summary(model)
rm(model)
```

Short

*James Short’s measurements of the parallax of the sun*

Description

Data for Exercise 6.65

Usage

Short

Format

A data frame/tibble with 158 observations on two variables

- **sample** sample number
- **parallax** parallax measurements (seconds of a degree)

References

### Examples

```
hist(Short$parallax, main = "Problem 6.65",
     xlab = "", col = "orange")
SIGN.test(Short$parallax, md = 8.798)
t.test(Short$parallax, mu = 8.798)
```

<table>
<thead>
<tr>
<th>Shuttle</th>
<th>Number of people riding shuttle versus number of automobiles in the downtown area</th>
</tr>
</thead>
</table>

### Description

Data for Exercise 9.20

### Usage

Shuttle

### Format

A data frame/tibble with 15 observations on two variables

- **users**  number of shuttle riders
- **autos**  number of automobiles in the downtown area

### References


### Examples

```
plot(autos ~ users, data = Shuttle)
model <- lm(autos ~ users, data = Shuttle)
summary(model)
rm(model)
```
Description

This function will test a hypothesis based on the sign test and reports linearly interpolated confidence intervals for one sample problems.

Usage

SIGN.test(
  x, 
  y = NULL, 
  md = 0,
  alternative = "two.sided",
  conf.level = 0.95,
  ...
)

Arguments

x numeric vector; NAs and Infs are allowed but will be removed.
y optional numeric vector; NAs and Infs are allowed but will be removed.
md a single number representing the value of the population median specified by the null hypothesis
alternative is a character string, one of "greater", "less", or "two.sided", or the initial letter of each, indicating the specification of the alternative hypothesis. For one-sample tests, alternative refers to the true median of the parent population in relation to the hypothesized value of the median.
conf.level confidence level for the returned confidence interval, restricted to lie between zero and one
... further arguments to be passed to or from methods

Details

Computes a “Dependent-samples Sign-Test” if both x and y are provided. If only x is provided, computes the “Sign-Test”.

Value

A list of class htest_S, containing the following components:

statistic the S-statistic (the number of positive differences between the data and the hypothesized median), with names attribute “S”.
p.value the p-value for the test
**SIGN.test**

conf.int is a confidence interval (vector of length 2) for the true median based on linear interpolation. The confidence level is recorded in the attribute conf.level. When the alternative is not "two.sided", the confidence interval will be half-infinite, to reflect the interpretation of a confidence interval as the set of all values \( k \) for which one would not reject the null hypothesis that the true mean or difference in means is \( k \). Here infinity will be represented by Inf.

**estimate** is a vector of length 1, giving the sample median; this estimates the corresponding population parameter. Component estimate has a names attribute describing its elements.

null.value is the value of the median specified by the null hypothesis. This equals the input argument \( md \). Component null.value has a names attribute describing its elements.

alternative records the value of the input argument alternative: "greater", "less", or "two.sided"

**data.name** a character string (vector of length 1) containing the actual name of the input vector \( x \)

**Confidence.Intervals** a 3 by 3 matrix containing the lower achieved confidence interval, the interpolated confidence interval, and the upper achieved confidence interval.

**Null Hypothesis**

For the one-sample sign-test, the null hypothesis is that the median of the population from which \( x \) is drawn is \( md \). For the two-sample dependent case, the null hypothesis is that the median for the differences of the populations from which \( x \) and \( y \) are drawn is \( md \). The alternative hypothesis indicates the direction of divergence of the population median for \( x \) from \( md \) (i.e., "greater", "less", "two.sided").

**Note**

The reported confidence interval is based on linear interpolation. The lower and upper confidence levels are exact.

**Author(s)**

Alan T. Arnholt

**References**


See Also

z.test, zsum.test, tsum.test

Examples

```r
x <- c(7.8, 6.6, 6.5, 7.4, 7.3, 7., 6.4, 7.1, 6.7, 7.6, 6.8)
SIGN.test(x, md = 6.5)
# Computes two-sided sign-test for the null hypothesis
# that the population median for 'x' is 6.5. The alternative
# hypothesis is that the median is not 6.5. An interpolated 95%
# confidence interval for the population median will be computed.

             14.4, 15.8, 11.3, 15.0)
SIGN.test(reaction, md = 15, alternative = "less")
# Data from Example 6.11 page 330 of Kitchens BSDA.
# Computes one-sided sign-test for the null hypothesis
# that the population median is 15. The alternative
# hypothesis is that the median is less than 15.
# An interpolated upper 95% upper bound for the population
# median will be computed.
```

Simpson

Grade point averages of men and women participating in various
sports—an illustration of Simpson’s paradox

Description

Data for Example 1.18

Usage

Simpson

Format

A data frame/tibble with 100 observations on three variables

- **gpa** grade point average
- **sport** sport played (basketball, soccer, or track)
- **gender** athlete sex (male, female)

References

### Examples

```r
boxplot(gpa ~ gender, data = Simpson, col = "violet")
boxplot(gpa ~ sport, data = Simpson, col = "lightgreen")
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Simpson, aes(x = gender, y = gpa, fill = gender)) +
  geom_boxplot() +
  facet_grid(.~sport) +
  theme_bw()
## End(Not run)
```

### Description

Maximum number of situps by participants in an exercise class

### Usage

```r
Situp
```

### Format

A data frame/tibble with 20 observations on one variable

- **number**: maximum number of situps completed in an exercise class after 1 month in the program

### References


### Examples

```r
stem(Situp$number)
hist(Situp$number, breaks = seq(0, 70, 10), right = FALSE)
hist(Situp$number, breaks = seq(0, 70, 10), right = FALSE, freq = FALSE, col = "pink", main = "Problem 1.47",
     xlab = "Maximum number of situps")
lines(density(Situp$number), col = "red")
```
**Description**

Data for Exercise 7.65

**Usage**

Skewed

**Format**

A data frame/tibble with 21 observations on two variables

- **C1** values from a sample of size 16 from a particular population
- **C2** values from a sample of size 14 from a particular population

**References**


**Examples**

```r
boxplot(Skewed$C1, Skewed$C2, col = c("pink", "lightblue"))
wilcox.test(Skewed$C1, Skewed$C2)
```

---

**Description**

Data for Exercise 5.20

**Usage**

Skin

**Survival times of closely and poorly matched skin grafts on burn patients**
Format

A data frame/tibble with 11 observations on four variables

- **patient**: patient identification number
- **close**: graft survival time in days for a closely matched skin graft on the same burn patient
- **poor**: graft survival time in days for a poorly matched skin graft on the same burn patient
- **differ**: difference between close and poor (in days)

Source


References


Examples

```r
stem(Skin$differ)
boxplot(Skin$differ, col = "pink")
summary(Skin$differ)
```

---

**Slc**

*Sodium-lithium countertransport activity on 190 individuals from six large English kindred*

Description

Data for Exercise 5.116

Usage

Slc

Format

A data frame/tibble with 190 observations on one variable

- **slc**: Red blood cell sodium-lithium countertransport

Source

References


Examples

```r
EDA(Slc$s1c)
hist(Slc$s1c, freq = FALSE, xlab = "sodium lithium countertransport", main = "", col = "lightblue")
lines(density(Slc$s1c), col = "purple")
```

| Smokyph | Water pH levels of 75 water samples taken in the Great Smoky Mountains |

Description

Data for Exercises 6.40, 6.59, 7.10, and 7.35

Usage

Smokyph

Format

A data frame/tibble with 75 observations on three variables

- **waterph** water sample pH level
- **code** character variable with values low (elevation below 0.6 miles), and high (elevation above 0.6 miles)
- **elev** elevation in miles

Source


References

**Snore**

### Examples

```r
summary(Smokyph$waterph)
tapply(Smokyph$waterph, Smokyph$code, mean)
stripchart(waterph ~ code, data = Smokyph, method = "stack",
          pch = 19, col = c("red", "blue"))
t.test(Smokyph$waterph, mu = 7)
SIGN.test(Smokyph$waterph, md = 7)
t.test(waterph ~ code, data = Smokyph, alternative = "less")
t.test(waterph ~ code, data = Smokyph, conf.level = 0.90)
```

```r
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Smokyph, aes(x = waterph, fill = code)) +
  geom_dotplot() +
  facet_grid(code ~ .) +
  guides(fill = FALSE)

## End(Not run)
```

### Snore

**Snoring versus heart disease**

<table>
<thead>
<tr>
<th>Snore</th>
<th>Snoring versus heart disease</th>
</tr>
</thead>
</table>

### Description

Data for Exercise 8.21

### Usage

```r
Snore
```

### Format

A data frame/tibble with 2,484 observations on two variables

- **snore** factor with levels nonsnorer, occasional snorer, nearly every night, and snores every night
- **heartdisease** factor indicating whether the individual has heart disease (no or yes)

### Source


### References

Examples

```
T1 <- xtabs(~ heartdisease + snore, data = Snore)
T1
chisq.test(T1)
rm(T1)
```

**Description**

Data for Exercise 7.87

**Usage**

`Snow`

**Format**

A data frame/tibble with 34 observations on two variables

- `concent` concentration of microparticles from melted snow (in parts per billion)
- `site` location of snow sample (Antarctica or Greenland)

**Source**


**References**


**Examples**

```
boxplot(concent ~ site, data = Snow, col = c("lightblue", "lightgreen"))
```
**Soccer**

*Weights of 25 soccer players*

**Description**

Data for Exercise 1.46

**Usage**

Soccer

**Format**

A data frame/tibble with 25 observations on one variable

- **weight**: soccer players weight (in pounds)

**References**


**Examples**

```r
stem(Soccer$weight, scale = 2)
hist(Soccer$weight, breaks = seq(110, 210, 10), col = "orange",
    main = "Problem 1.46 \n Weights of Soccer Players",
    xlab = "weight (lbs)", right = FALSE)
```

---

**Social**

*Median income level for 25 social workers from North Carolina*

**Description**

Data for Exercise 6.63

**Usage**

Social

**Format**

A data frame/tibble with 25 observations on one variable

- **income**: annual income (in dollars) of North Carolina social workers with less than five years experience.
References


Examples

SIGN.test(Social$income, md = 27500, alternative = "less")

| Sophomor | Grade point averages, SAT scores and final grade in college algebra for 20 sophomores |

Description

Data for Exercise 2.42

Usage

Sophomor

Format

A data frame/tibble with 20 observations on four variables

- student  identification number
- gpa  grade point average
- sat  SAT math score
- exam  final exam grade in college algebra

References


Examples

cor(Sophomor)
plot(exam ~ gpa, data = Sophomor)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Sophomor, aes(x = gpa, y = exam)) + geom_point()
ggplot2::ggplot(data = Sophomor, aes(x = sat, y = exam)) + geom_point()

## End(Not run)
South

Murder rates for 30 cities in the South

Description
Data for Exercise 1.84

Usage
South

Format
A data frame/tibble with 31 observations on one variable
rate murder rate per 100,000 people

References

Examples
boxplot(South$rate, col = "gray", ylab = "Murder rate per 100,000 people")

Speed
Speed reading scores before and after a course on speed reading

Description
Data for Exercise 7.58

Usage
Speed

Format
A data frame/tibble with 15 observations on four variables
before reading comprehension score before taking a speed-reading course
after reading comprehension score after taking a speed-reading course
differ after - before (comprehension reading scores)
signranks signed ranked differences
References


Examples

```r
t.test(Speed$differ, alternative = "greater")
t.test(Speed$signranks, alternative = "greater")
wilcox.test(Speed$after, Speed$before, paired = TRUE, alternative = "greater")
```

---

### Spellers

**Standardized spelling test scores for two fourth grade classes**

**Description**

Data for Exercise 7.82

**Usage**

Spellers

**Format**

A data frame/tibble with ten observations on two variables

- **teacher** character variable with values Fourth and Colleague
- **score** score on a standardized spelling test

**References**


**Examples**

```r
boxplot(score ~ teacher, data = Spellers, col = "pink")
t.test(score ~ teacher, data = Spellers)
```
Spelling

Spelling scores for 9 eighth graders before and after a 2-week course of instruction

Description
Data for Exercise 7.56

Usage
Spelling

Format
A data frame/tibble with nine observations on three variables
- **before**: spelling score before a 2-week course of instruction
- **after**: spelling score after a 2-week course of instruction
- **differ**: after - before (spelling score)

References

Examples
```
qqnorm(Spelling$differ)
qqline(Spelling$differ)
shapiro.test(Spelling$differ)
t.test(Spelling$before, Spelling$after, paired = TRUE)
t.test(Spelling$differ)
```

Sports
Favorite sport by gender

Description
Data for Exercise 8.32

Usage
Sports
Spouse

Format
A data frame/tibble with 200 observations on two variables

- gender: a factor with levels male and female
- sport: a factor with levels football, basketball, baseball, and tennis

References

Examples

```r
T1 <- xtabs(~gender + sport, data = Sports)
T1
chiq.test(T1)
rm(T1)
```

---

Spouse

Convictions in spouse murder cases by gender

Description
Data for Exercise 8.33

Usage
Spouse

Format
A data frame/tibble with 540 observations on two variables

- result: a factor with levels not prosecuted, pleaded guilty, convicted, and acquitted
- spouse: a factor with levels husband and wife

Source
Bureau of Justice Statistics (September 1995), *Spouse Murder Defendants in Large Urban Counties*, Executive Summary, NCJ-156831.

References
Examples

```r
T1 <- xtabs(~result + spouse, data = Spouse)
T1
chisq.test(T1)
rm(T1)
```

Description

Computes all possible samples from a given population using simple random sampling.

Usage

```r
SRS(POPvalues, n)
```

Arguments

- `POPvalues`: vector containing the population values.
- `n`: the sample size.

Value

Returns a matrix containing the possible simple random samples of size `n` taken from a population `POPvalues`.

Author(s)

Alan T. Arnholt

See Also

Combinations

Examples

```r
SRS(c(5,8,3),2)
```

# The rows in the matrix list the values for the 3 possible
# simple random samples of size 2 from the population of 5,8, and 3.
Stable  

Times of a 2-year old stallion on a one mile run

Description
Data for Exercise 6.93

Usage
Stable

Format
A data frame/tibble with nine observations on one variable

time  time (in seconds) for horse to run 1 mile

References

Examples
SIGN.test(Stable$time, md = 98.5, alternative = "greater")

Stamp  

Thicknesses of 1872 Hidalgo stamps issued in Mexico

Description
Data for Statistical Insight Chapter 1 and Exercise 5.110

Usage
Stamp

Format
A data frame/tibble with 485 observations on one variable

thickness  stamp thickness (in mm)
Source

References

Examples
```
hist(Stamp$thickness, freq = FALSE, col = "lightblue",
     main = "", xlab = "stamp thickness (mm)"
lines(density(Stamp$thickness), col = "blue")
t.test(Stamp$thickness, conf.level = 0.99)
```

Statclas

<table>
<thead>
<tr>
<th>Statclas</th>
<th>Grades for two introductory statistics classes</th>
</tr>
</thead>
</table>

Description
Data for Exercise 7.30

Usage
Statclas

Format
A data frame/tibble with 72 observations on two variables
```
class  class meeting time (9am or 2pm)
score  grade for an introductory statistics class
```

References

Examples
```
str(Statclas)
boxplot(score ~ class, data = Statclas, col = "red")
t.test(score ~ class, data = Statclas)
```
### Statelaw

*Operating expenditures per resident for each of the state law enforcement agencies*

<table>
<thead>
<tr>
<th>Statelaw</th>
<th>Description</th>
<th>Usage</th>
<th>Format</th>
<th>Source</th>
<th>References</th>
<th>Examples</th>
</tr>
</thead>
</table>

### Statisti

*Test scores for two beginning statistics classes*

<table>
<thead>
<tr>
<th>Statisti</th>
<th>Description</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data for Exercises 1.70 and 1.87</td>
<td>Statisti</td>
</tr>
</tbody>
</table>
Step

**Format**

A data frame/tibble with 62 observations on two variables

- **class**: character variable with values Class1 and Class2
- **score**: test score for an introductory statistics test

**References**


**Examples**

```r
boxplot(score ~ class, data = Statisti, col = "violet")
tapply(Statisti$score, Statisti$class, summary, na.rm = TRUE)
```

```r
## Not run:
library(dplyr)
dplyr::group_by(Statisti, class) %>%
summarize(Mean = mean(score, na.rm = TRUE),
          Median = median(score, na.rm = TRUE),
          SD = sd(score, na.rm = TRUE),
          RS = IQR(score, na.rm = TRUE))
## End(Not run)
```

---

**Step**

*STEP science test scores for a class of ability-grouped students*

**Description**

Data for Exercise 6.79

**Usage**

`Step`

**Format**

A data frame/tibble with 12 observations on one variable

- **score**: State test of educational progress (STEP) science test score

**References**

Examples

```r
edascore

t.test(dascore, mu = 80, alternative = "less")
wilcox.test(dascore, mu = 80, alternative = "less")
```

---

| **Stress** | Short-term memory test scores on 12 subjects before and after a stressful situation |

---

Description

Data for Example 7.20

Usage

Stress

Format

A data frame/tibble with 12 observations on two variables

- **prestress** short term memory score before being exposed to a stressful situation
- **poststress** short term memory score after being exposed to a stressful situation

References


Examples

```r
diff <- Stresselect$prestress - Stresselect$poststress
qqnorm(diff)
qqline(diff)
t.test(diff)
t.test(Stresselect$prestress, Stresselect$poststress, paired = TRUE)
## Not run:
wilcox.test(Stresselect$prestress, Stresselect$poststress, paired = TRUE)
## End(Not run)
```
Study

Number of hours studied per week by a sample of 50 freshmen

Description

Data for Exercise 5.25

Usage

Study

Format

A data frame/tibble with 50 observations on one variable

hours number of hours a week freshmen reported studying for their courses

References


Examples

```
stem(Study$hours)
hist(Study$hours, col = "violet")
summary(Study$hours)
```
### Source


### References


### Examples

```r
model <- lm(actual ~ reported, data = Subway)
summary(model)
plot(actual ~ reported, data = Subway)
abline(model, col = "red")
rm(model)
```

---

### Subway

*Time it takes a subway to travel from the airport to downtown*

<table>
<thead>
<tr>
<th>Subway</th>
<th>Time it takes a subway to travel from the airport to downtown</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Description

Data for Exercise 5.19

### Usage

Subway

### Format

A data frame/tibble with 30 observations on one variable

- **time**: time (in minutes) it takes a subway to travel from the airport to downtown

### References


### Examples

```r
hist(Subway$time, main = "Exercise 5.19",
     xlab = "Time (in minutes)", col = "purple")
summary(Subway$time)
```
Description

Data for Example 1.7

Usage

Sunspot

Format

A data frame/tibble with 301 observations on two variables

year  year
sunspots  average number of sunspots for the year

References


Examples

plot(sunspots ~ year, data = Sunspot, type = "l")
## Not run:
library(ggplot2)
lattice::xplot(sunspots ~ year, data = Sunspot,
              main = "Yearly sunspots", type = "l")
lattice::xplot(sunspots ~ year, data = Sunspot, type = "l",
              main = "Yearly sunspots", aspect = "xy")
ggplot2::ggplot(data = Sunspot, aes(x = year, y = sunspots)) +
                geom_line() +
                theme_bw()
## End(Not run)
Superbowl

*Margin of victory in Superbowls I to XXXV*

**Description**

Data for Exercise 1.54

**Usage**

Superbowl

**Format**

A data frame/tibble with 35 observations on five variables

- **winning_team** name of Superbowl winning team
- **winner_score** winning score for the Superbowl
- **losing_team** name of Superbowl losing team
- **loser_score** score of losing team numeric vector
- **victory_margin** winner_score - loser_score

**References**


**Examples**

```
stem(Superbowl$victory_margin)
```

Supercar

*Top speeds attained by five makes of supercars*

**Description**

Data for Statistical Insight Chapter 10

**Usage**

Supercar
**Format**

A data frame/tibble with 30 observations on two variables

- **speed**: top speed (in miles per hour) of car without redlining
- **car**: name of sports car

**Source**

*Car and Drvier* (July 1995).

**References**


**Examples**

```r
boxplot(speed ~ car, data = Supercar, col = rainbow(6), ylab = "Speed (mph)")
summary(aov(speed ~ car, data = Supercar))
anova(lm(speed ~ car, data = Supercar))
```

---

**Tablrock**

Ozone concentrations at Mt. Mitchell, North Carolina

**Description**

Data for Exercise 5.63

**Usage**

Tablrock

**Format**

A data frame/tibble with 719 observations on the following 17 variables.

- **day**: date
- **hour**: time of day
- **ozone**: ozone concentration
- **tmp**: temperature (in Celcius)
- **vdc**: a numeric vector
- **wd**: a numeric vector
- **ws**: a numeric vector
amb  a numeric vector

dew  a numeric vector

so2  a numeric vector

no   a numeric vector

no2  a numeric vector

nox  a numeric vector

co   a numeric vector

co2  a numeric vector

gas  a numeric vector

air  a numeric vector

References


Examples

```r
summary(Tablrock$ozone)
boxplot(Tablrock$ozone)
qqnorm(Tablrock$ozone)
qqline(Tablrock$ozone)
par(mar = c(5.1 - 1, 4.1 + 2, 4.1 - 2, 2.1))
boxplot(ozone ~ day, data = Tablrock,
        horizontal = TRUE, las = 1, cex.axis = 0.7)
par(mar = c(5.1, 4.1, 4.1, 2.1))
## Not run:
library(ggplot2)
  ggplot2::ggplot(data = Tablrock, aes(sample = ozone)) +
  geom_qq() +
  theme_bw()
  ggplot2::ggplot(data = Tablrock, aes(x = as.factor(day), y = ozone)) +
  geom_boxplot(fill = "pink") +
  coord_flip() +
  labs(x = "") +
  theme_bw()
## End(Not run)
```
Teacher

**Average teacher’s salaries across the states in the 70s 80s and 90s**

**Description**

Data for Exercise 5.114

**Usage**

Teacher

**Format**

A data frame/tibble with 51 observations on three variables

- **state** U.S. state
- **year** academic year
- **salary** average salary (in dollars)

**Source**

National Education Association.

**References**


**Examples**

```r
par(mfrow = c(3, 1))
hist(Teacher$salary[Teacher$year == "1973-74"],
    main = "Teacher salary 1973-74", xlab = "salary",
    xlim = range(Teacher$salary, na.rm = TRUE))
hist(Teacher$salary[Teacher$year == "1983-84"],
    main = "Teacher salary 1983-84", xlab = "salary",
    xlim = range(Teacher$salary, na.rm = TRUE))
hist(Teacher$salary[Teacher$year == "1993-94"],
    main = "Teacher salary 1993-94", xlab = "salary",
    xlim = range(Teacher$salary, na.rm = TRUE))
par(mfrow = c(1, 1))
```

Not run:

```r
library(ggplot2)
ggplot(data = Teacher, aes(x = salary)) +
  geom_histogram(fill = "purple", color = "black") +
  facet_grid(year ~ .) +
  theme_bw()
```
Tenness

Tennessee self concept scores for 20 gifted high school students

Description

Data for Exercise 6.56

Usage

Tenness

Format

A data frame/tibble with 20 observations on one variable

score  Tennessee Self-Concept Scale score

References


Examples

```r
hist(Tenness$score, freq= FALSE, main = "", col = "green",
     xlab = "Tennessee Self-Concept Scale score")
lines(density(Tenness$score))

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Tenness, aes(x = score, y = ..density..)) +
   geom_histogram(binwidth = 2, fill = "purple", color = "black") +
   geom_density(color = "red", fill = "pink", alpha = 0.3) +
   theme_bw()

## End(Not run)
```
Tensile

*Tensile strength of plastic bags from two production runs*

**Description**

Data for Example 7.11

**Usage**

Tensile

**Format**

A data frame/tibble with 72 observations on two variables

- **tensile**: plastic bag tensile strength (pounds per square inch)
- **run**: factor with run number (1 or 2)

**References**


**Examples**

```r
boxplot(tensile ~ run, data = Tensile, col = c("purple", "cyan"))
t.test(tensile ~ run, data = Tensile)
```

---

Test1

*Grades on the first test in a statistics class*

**Description**

Data for Exercise 5.80

**Usage**

Test1

**Format**

A data frame/tibble with 25 observations on one variable

- **score**: score on first statistics exam
Thermal

References


Examples

```r
stem(Test1$score)
boxplot(Test1$score, col = "purple")
```

---

### Thermal

*Heat loss of thermal pane windows versus outside temperature*

Description

Data for Example 9.5

Usage

Thermal

Format

A data frame/tibble with 12 observations on the two variables

- `temp` temperature (degrees Celcius)
- `loss` heat loss (BTUs)

References


Examples

```r
model <- lm(loss ~ temp, data = Thermal)
summary(model)
plot(loss ~ temp, data = Thermal)
abline(model, col = "red")
rm(model)
```
**Tiaa**

**Description**

Data for your enjoyment

**Usage**

Tiaa

**Format**

A data frame/tibble with 365 observations on four variables

- **crefstk**: closing price (in dollars)
- **crefgwt**: closing price (in dollars)
- **tiaa**: closing price (in dollars)
- **date**: day of the year

**References**


**Examples**

```r
data(Tiaa)
```

---

**Ticket**

*Time to complete an airline ticket reservation*

**Description**

Data for Exercise 5.18

**Usage**

Ticket

**Format**

A data frame/tibble with 20 observations on one variable

- **time**: time (in seconds) to check out a reservation
Toaster

References


Examples

EDA(Ticket$time)

---

**Toaster**  
*Consumer Reports (Oct 94) rating of toaster ovens versus the cost*

Description

Data for Exercise 9.36

Usage

Toaster

Format

A data frame/tibble with 17 observations on three variables

- **toaster** name of toaster  
- **score** Consumer Reports score  
- **cost** price of toaster (in dollars)

Source

*Consumer Reports* (October 1994).

References


Examples

```r
plot(cost ~ score, data = Toaster)
model <- lm(cost ~ score, data = Toaster)
summary(model)
names(summary(model))
summary(model)$r.squared
plot(model, which = 1)
```
## Tonsils

Size of tonsils collected from 1,398 children

### Description

Data for Exercise 2.78

### Usage

Tonsils

### Format

A data frame/tibble with 1,398 observations on two variables

- **size** a factor with levels Normal, Large, and Very Large
- **status** a factor with levels Carrier and Non-carrier

### References


### Examples

```r
T1 <- xtabs(~size + status, data = Tonsils)
T1
prop.table(T1, 1)
prop.table(T1, 1)[2, 1]
barplot(t(T1), legend = TRUE, beside = TRUE, col = c("red", "green"))
## Not run:
library(dplyr)
library(ggplot2)
NDF <- dplyr::count(Tonsils, size, status)
ggplot2::ggplot(data = NDF, aes(x = size, y = n, fill = status)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_fill_manual(values = c("red", "green")) +
  theme_bw()
## End(Not run)
```
The number of torts, average number of months to process a tort, and county population from the court files of the nation’s largest counties

Description

Data for Exercise 5.13

Usage

Tort

Format

A data frame/tibble with 45 observations on five variables

- `county` U.S. county
- `months` average number of months to process a tort
- `population` population of the county
- `torts` number of torts
- `rate` rate per 10,000 residents

Source


References


Examples

EDA(Tort$months)
Toxic

Hazardous waste sites near minority communities

Description

Data for Exercises 1.55, 5.08, 5.109, 8.58, and 10.35

Usage

Toxic

Format

A data frame/tibble with 51 observations on five variables

state  U.S. state
region  U.S. region
sites  number of commercial hazardous waste sites
minority  percent of minorities living in communities with commercial hazardous waste sites
percent  a numeric vector

References


Examples

hist(Toxic$sites, col = "red")
hist(Toxic$minority, col = "blue")
qqnorm(Toxic$minority)
qqline(Toxic$minority)
boxplot(sites ~ region, data = Toxic, col = "lightgreen")
tapply(Toxic$sites, Toxic$region, median)
kruskal.test(sites ~ factor(region), data = Toxic)
**Track**

*National Olympic records for women in several races*

**Description**

Data for Exercises 2.97, 5.115, and 9.62

**Usage**

Track

**Format**

A data frame with 55 observations on eight variables

- **country**: athlete’s country
- **100m**: time in seconds for 100 m
- **200m**: time in seconds for 200 m
- **400m**: time in seconds for 400 m
- **800m**: time in minutes for 800 m
- **1500m**: time in minutes for 1500 m
- **3000m**: time in minutes for 3000 m
- **marathon**: time in minutes for marathon

**Source**


**References**


**Examples**

```r
plot("200m ~ 100m", data = Track)
plot("400m ~ 100m", data = Track)
plot("400m ~ 200m", data = Track)
cor(Track[, 2:8])
```
Description

Data for Exercise 1.36

Usage

Track15

Format

A data frame/tibble with 26 observations on two variables

- **year**: Olympic year
- **time**: Olympic winning time (in seconds) for the 1500-meter run

Source


References


Examples

```r
plot(time~ year, data = Track15, type = "b", pch = 19,
     ylab = "1500m time in seconds", col = "green")
```

---

Description

Data for Exercise 10.44

Usage

Treatments

Illustrates analysis of variance for three treatment groups

Description

Data for Exercise 10.44

Usage

Treatments
### Format

A data frame/tibble with 24 observations on two variables

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>score</strong></td>
<td>score from an experiment</td>
</tr>
<tr>
<td><strong>group</strong></td>
<td>factor with levels 1, 2, and 3</td>
</tr>
</tbody>
</table>

### References


### Examples

```r
boxplot(score ~ group, data = Treatments, col = "violet")
summary(aov(score ~ group, data = Treatments))
summary(lm(score ~ group, data = Treatments))
anova(lm(score ~ group, data = Treatments))
```

---

### Description

Data for Exercise 1.50

### Usage

Trees

### Format

A data frame/tibble with 20 observations on one variable

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>number</strong></td>
<td>number of trees in a grid</td>
</tr>
</tbody>
</table>

### References


### Examples

```r
stem(Trees$number)
hist(Trees$number, main = "Exercise 1.50", xlab = "number", col = "brown")
```
Trucks

Miles per gallon for standard 4-wheel drive trucks manufactured by Chevrolet, Dodge and Ford

Description

Data for Example 10.2

Usage

Trucks

Format

A data frame/tibble with 15 observations on two variables

mpg  miles per gallon
truck  a factor with levels chevy, dodge, and ford

References


Examples

boxplot(mpg ~ truck, data = Trucks, horizontal = TRUE, las = 1)
summary(aov(mpg ~ truck, data = Trucks))

---

tsum.test  Summarized t-test

Description

Performs a one-sample, two-sample, or a Welch modified two-sample t-test based on user supplied summary information. Output is identical to that produced with t.test.
Usage

tsum.test(
  mean.x,
  s.x = NULL,
  n.x = NULL,
  mean.y = NULL,
  s.y = NULL,
  n.y = NULL,
  alternative = "two.sided",
  mu = 0,
  var.equal = FALSE,
  conf.level = 0.95
)

Arguments

mean.x is a single number representing the sample mean of x
s.x is a single number representing the sample standard deviation for x
n.x is a single number representing the sample size for x
mean.y is a single number representing the sample mean of y
s.y is a single number representing the sample standard deviation for y
n.y is a single number representing the sample size for y
alternative is a character string, one of "greater", "less" or "two.sided", or just the initial letter of each, indicating the specification of the alternative hypothesis. For one-sample tests, alternative refers to the true mean of the parent population in relation to the hypothesized value mu. For the standard two-sample tests, alternative refers to the difference between the true population mean for x and that for y, in relation to mu. For the one-sample and paired t-tests, alternative refers to the true mean of the parent population in relation to the hypothesized value mu. For the standard and Welch modified two-sample t-tests, alternative refers to the difference between the true population mean for x and that for y, in relation to mu. For the one-sample t-tests, alternative refers to the true mean of the parent population in relation to the hypothesized value mu. For the standard and Welch modified two-sample t-tests, alternative refers to the difference between the true population mean for x and that for y, in relation to mu.
mu is a single number representing the value of the mean or difference in means specified by the null hypothesis.
var.equal is a logical flag: if TRUE, the variances of the parent populations of x and y are assumed equal. Argument var.equal should be supplied only for the two-sample tests.
conf.level is the confidence level for the returned confidence interval; it must lie between zero and one.
Details

If y is NULL, a one-sample t-test is carried out with x. If y is not NULL, either a standard or Welch modified two-sample t-test is performed, depending on whether var.equal is TRUE or FALSE.

Value

A list of class htest, containing the following components:

- **statistic**: the t-statistic, with names attribute “t”
- **parameters**: is the degrees of freedom of the t-distribution associated with statistic. Component parameters has names attribute “df”.
- **p.value**: the p-value for the test.
- **conf.int**: is a confidence interval (vector of length 2) for the true mean or difference in means. The confidence level is recorded in the attribute conf.level. When alternative is not "two.sided", the confidence interval will be half-infinite, to reflect the interpretation of a confidence interval as the set of all values \( k \) for which one would not reject the null hypothesis that the true mean or difference in means is \( k \). Here infinity will be represented by Inf.
- **estimate**: vector of length 1 or 2, giving the sample mean(s) or mean of differences; these estimate the corresponding population parameters. Component estimate has a names attribute describing its elements.
- **null.value**: the value of the mean or difference in means specified by the null hypothesis. This equals the input argument mu. Component null.value has a names attribute describing its elements.
- **alternative**: records the value of the input argument alternative: "greater", "less" or "two.sided".
- **data.name**: a character string (vector of length 1) containing the names x and y for the two summarized samples.

Null Hypothesis

For the one-sample t-test, the null hypothesis is that the mean of the population from which x is drawn is mu. For the standard and Welch modified two-sample t-tests, the null hypothesis is that the population mean for x less that for y is mu.

The alternative hypothesis in each case indicates the direction of divergence of the population mean for x (or difference of means for x and y) from mu (i.e., "greater", "less", or "two.sided").

Author(s)

Alan T. Arnholt

References


See Also

`z.test`, `zsum.test`

Examples

```r
tsum.test(mean.x=5.6, s.x=2.1, n.x=16, mu=4.9, alternative="greater")
# Problem 6.31 on page 324 of BSDA states: The chamber of commerce
# of a particular city claims that the mean carbon dioxide
# level of air polution is no greater than 4.9 ppm. A random
# sample of 16 readings resulted in a sample mean of 5.6 ppm,
# and s=2.1 ppm. One-sided one-sample t-test. The null
# hypothesis is that the population mean for 'x' is 4.9.
# The alternative hypothesis states that it is greater than 4.9.

x <- rnorm(12)
tsum.test(mean(x), sd(x), n.x=12)
# Two-sided one-sample t-test. The null hypothesis is that
# the population mean for 'x' is zero. The alternative
# hypothesis states that it is either greater or less
# than zero. A confidence interval for the population mean
# will be computed. Note: above returns same answer as:
# t.test(x)

x <- c(7.8, 6.6, 6.5, 7.4, 7.3, 7.0, 6.4, 7.1, 6.7, 7.6, 6.8)
y <- c(4.5, 5.4, 6.1, 6.1, 5.4, 5.0, 4.1, 5.5)
tsum.test(mean(x), s.x=sd(x), n.x=11, mean(y), s.y=sd(y), n.y=8, mu=2)
# Two-sided standard two-sample t-test. The null hypothesis
# is that the population mean for 'x' less that for 'y' is 2.
# The alternative hypothesis is that this difference is not 2.
# A confidence interval for the true difference will be computed.
# Note: above returns same answer as:
# t.test(x, y)

tsum.test(mean(x), s.x=sd(x), n.x=11, mean(y), s.y=sd(y), n.y=8, conf.level=0.90)
# Two-sided standard two-sample t-test. The null hypothesis
# is that the population mean for 'x' less that for 'y' is zero.
# The alternative hypothesis is that this difference is not
# zero. A 90% confidence interval for the true difference will
# be computed. Note: above returns same answer as:
# t.test(x, y, conf.level=0.90)
```
Tv

Percent of students that watch more than 6 hours of TV per day versus national math test scores

Description
Data for Examples 2.1 and 2.7

Usage
Tv

Format
A data frame/tibble with 53 observations on three variables

- **state**: U.S. state
- **percent**: percent of students who watch more than six hours of TV a day
- **test**: state average on national math test

Source
Educational Testing Services.

References

Examples

```r
plot(test ~ percent, data = Tv, col = "blue")
cor(Tv$test, Tv$percent)
```

Twin

*Intelligence test scores for identical twins in which one twin is given a drug*

Description
Data for Exercise 7.54

Usage
Twin
Format

A data frame/tibble with nine observations on three variables

- `twinA` score on intelligence test without drug
- `twinB` score on intelligence test after taking drug
- `differ` twinA - twinB

References


Examples

```r
qqnorm(Twin$differ)
qqline(Twin$differ)
shapiro.test(Twin$differ)
t.test(Twin$twinA, Twin$twinB, paired = TRUE)
```

---

Undergrad

Data set describing a sample of undergraduate students

Description

Data for Exercise 1.15

Usage

Undergrad

Format

A data frame/tibble with 100 observations on six variables

- `gender` character variable with values Female and Male
- `major` college major
- `class` college year group classification
- `gpa` grade point average
- `sat` Scholastic Assessment Test score
- `drops` number of courses dropped

References

Examples

```r
stripchart(gpa ~ class, data = Undergrad, method = "stack",
col = c("blue", "red", "green", "lightblue"),
pch = 19, main = "GPA versus Class")
stripchart(gpa ~ gender, data = Undergrad, method = "stack",
col = c("red", "blue"), pch = 19,
main = "GPA versus Gender")
stripchart(sat ~ drops, data = Undergrad, method = "stack",
col = c("blue", "red", "green", "lightblue"),
pch = 19, main = "SAT versus Drops")
stripchart(drops ~ gender, data = Undergrad, method = "stack",
col = c("red", "blue"), pch = 19, main = "Drops versus Gender")
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Undergrad, aes(x = sat, y = drops, fill = factor(drops))) +
  facet_grid(drops ~ .) +
  geom_dotplot() +
  guides(fill = FALSE)
## End(Not run)
```

---

**Vacation**

*Number of days of paid holidays and vacation leave for sample of 35 textile workers*

---

**Description**

Data for Exercise 6.46 and 6.98

**Usage**

`Vacation`

**Format**

A data frame/tibble with 35 observations on one variable

- **number**: number of days of paid holidays and vacation leave taken

**References**

Examples

```r
boxplot(Vacation$number, col = "violet")
hist(Vacation$number, main = "Exercise 6.46", col = "blue",
     xlab = "number of days of paid holidays and vacation leave taken")
t.test(Vacation$number, mu = 24)
```

---

**Vaccine**

Reported serious reactions due to vaccines in 11 southern states

---

**Description**

Data for Exercise 1.111

**Usage**

Vaccine

**Format**

A data frame/tibble with 11 observations on two variables

- **state**: U.S. state
- **number**: number of reported serious reactions per million doses of a vaccine

**Source**

Center for Disease Control, Atlanta, Georgia.

**References**


**Examples**

```r
stem(Vaccine$number, scale = 2)
fn <- fivenum(Vaccine$number)
fn
iqr <- IQR(Vaccine$number)
iqr
```
Vehicle 261

Vehicle Fatality ratings for foreign and domestic vehicles

Description
Data for Exercise 8.34

Usage
Vehicle

Format
A data frame/tibble with 151 observations on two variables

make a factor with levels domestic and foreign
rating a factor with levels Much better than average, Above average, Average, Below average, and Much worse than average

Source
Insurance Institute for Highway Safety and the Highway Loss Data Institute, 1995.

References

Examples

T1 <- xtabs(~make + rating, data = Vehicle)
T1
chisq.test(T1)

Verbal  Verbal test scores and number of library books checked out for 15 eighth graders

Description
Data for Exercise 9.30

Usage
Verbal
Format

A data frame/tibble with 15 observations on two variables

number number of library books checked out
verbal verbal test score

References


Examples

```r
plot(verbal ~ number, data = Verbal)
abline(lm(verbal ~ number, data = Verbal), col = "red")
summary(lm(verbal ~ number, data = Verbal))
```

<table>
<thead>
<tr>
<th>Victoria</th>
<th>Number of sunspots versus mean annual level of Lake Victoria Nyanza from 1902 to 1921</th>
</tr>
</thead>
</table>

Description

Data for Exercise 2.98

Usage

Victoria

Format

A data frame/tibble with 20 observations on three variables

year year
level mean annual level of Lake Victoria Nyanza
sunspot number of sunspots

Source


References

Examples

plot(level ~ sunspot, data = Victoria)
model <- lm(level ~ sunspot, data = Victoria)
summary(model)
rm(model)

Vviscosit

Viscosity measurements of a substance on two different days

Description

Data for Exercise 7.44

Usage

Viscosit

Format

A data frame/tibble with 11 observations on two variables

first viscosity measurement for a certain substance on day one
second viscosity measurement for a certain substance on day two

References


Examples

boxplot(Viscosit$first, Viscosit$second, col = “blue”)  
t.test(Viscosit$first, Viscosit$second, var.equal = TRUE)
Visual

*Visual acuity of a group of subjects tested under a specified dose of a drug*

**Description**

Data for Exercise 5.6

**Usage**

*Visual*

**Format**

A data frame/tibble with 18 observations on one variable

visual visual acuity measurement

**References**


**Examples**

```r
stem(Visual$visual)
boxplot(Visual$visual, col = "purple")
```

Vocab

*Reading scores before and after vocabulary training for 14 employees who did not complete high school*

**Description**

Data for Exercise 7.80

**Usage**

*Vocab*

**Format**

A data frame/tibble with 14 observations on two variables

first reading test score before formal vocabulary training

second reading test score after formal vocabulary training
Wastewat

References


Examples

t.test(Vocab$first, Vocab$second, paired = TRUE)

| Wastewat | Volume of injected waste water from Rocky Mountain Arsenal and number of earthquakes near Denver |

Description

Data for Exercise 9.18

Usage

Wastewat

Format

A data frame/tibble with 44 observations on two variables

gallons injected water (in million gallons)

number number of earthquakes detected in Denver

Source


References


Examples

plot(number ~ gallons, data = Wastewat)
model <- lm(number ~ gallons, data = Wastewat)
summary(model)
anova(model)
plot(model, which = 2)
Weather94  

Weather casualties in 1994

Description

Data for Exercise 1.30

Usage

Weather94

Format

A data frame/tibble with 388 observations on one variable

type factor with levels Extreme Temp, Flash Flood, Fog, High Wind, Hurricane, Lighting, Other, River Flood, Thunderstorm, Tornado, and Winter Weather

References


Examples

```
T1 <- xtabs(~type, data = Weather94)
T1
par(mar = c(5.1 + 2, 4.1 - 1, 4.1 - 2, 2.1))
barplot(sort(T1, decreasing = TRUE), las = 2, col = rainbow(11))
par(mar = c(5.1, 4.1, 4.1, 2.1))
## Not run:
library(ggplot2)
T2 <- as.data.frame(T1)
T2
ggplot2::ggplot(data = T2, aes(x = reorder(type, Freq), y = Freq)) +
  geom_bar(stat = "identity", fill = "purple") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 55, vjust = 0.5)) +
  labs(x = "", y = "count")
## End(Not run)
```
Wheat

Price of a bushel of wheat versus the national weekly earnings of production workers

Description

Data for Exercise 2.11

Usage

Wheat

Format

A data frame/tibble with 19 observations on three variables

- **year**: year
- **earnings**: national weekly earnings (in dollars) for production workers
- **price**: price for a bushel of wheat (in dollars)

Source


References


Examples

```r
par(mfrow = c(1, 2))
plot(earnings ~ year, data = Wheat)
plot(price ~ year, data = Wheat)
par(mfrow = c(1, 1))
```
Windmill

*Direct current produced by different wind velocities*

---

**Description**

Data for Exercise 9.34

**Usage**

Windmill

**Format**

A data frame/tibble with 25 observations on two variables

- **velocity** wind velocity (miles per hour)
- **output** power generated (DC volts)

**Source**


**References**


**Examples**

```r
summary(lm(output ~ velocity, data = Windmill))
anova(lm(output ~ velocity, data = Windmill))
```

---

Window

*Wind leakage for storm windows exposed to a 50 mph wind*

---

**Description**

Data for Exercise 6.54

**Usage**

Window
Wins

Format

A data frame/tibble with nine observations on two variables

**window** window number

**leakage** percent leakage from a 50 mph wind

References


Examples

```
SIGN.test(Window$leakage, md = 0.125, alternative = "greater")
```

<table>
<thead>
<tr>
<th>Wins</th>
<th>Baseball team wins versus seven independent variables for National league teams in 1990</th>
</tr>
</thead>
</table>

Description

Data for Exercise 9.23

Usage

`Wins`

Format

A data frame with 12 observations on nine variables

**team** name of team

**wins** number of wins

**batavg** batting average

**rbi** runs batted in

**stole** bases stole

**strkout** number of strikeouts

**caught** number of times caught stealing

**errors** number of errors

**era** earned run average
References


Examples

```r
plot(wins ~ era, data = Wins)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Wins, aes(x = era, y = wins)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  theme_bw()
## End(Not run)
```

---

### Wool

*Strength tests of two types of wool fabric*

---

**Description**

Data for Exercise 7.42

**Usage**

Wool

**Format**

A data frame/tibble with 20 observations on two variables

- **type**: type of wool (Type I, Type 2)
- **strength**: strength of wool

**References**


**Examples**

```r
boxplot(strength ~ type, data = Wool, col = c("blue", "purple"))
t.test(strength ~ type, data = Wool, var.equal = TRUE)
```
Yearsunspot

Monthly sunspot activity from 1974 to 2000

Description

Data for Exercise 2.7

Usage

Yearsunspot

Format

A data frame/tibble with 252 observations on two variables

- **number**: average number of sunspots
- **year**: date

Source

NASA/Marshall Space Flight Center, Huntsville, AL 35812.

References


Examples

```r
plot(number ~ year, data = Yearsunspot)
```

z.test

Z-test

Description

This function is based on the standard normal distribution and creates confidence intervals and tests hypotheses for both one and two sample problems.
z.test

Usage

z.test(
  x, y = NULL,
  alternative = "two.sided",
  mu = 0,
  sigma.x = NULL,
  sigma.y = NULL,
  conf.level = 0.95
)

Arguments

x numeric vector; NAs and Infs are allowed but will be removed.
y numeric vector; NAs and Infs are allowed but will be removed.
alternative character string, one of "greater", "less" or "two.sided", or the initial letter of each, indicating the specification of the alternative hypothesis. For one-sample tests, alternative refers to the true mean of the parent population in relation to the hypothesized value mu. For the standard two-sample tests, alternative refers to the difference between the true population mean for x and that for y, in relation to mu.
mu a single number representing the value of the mean or difference in means specified by the null hypothesis
sigma.x a single number representing the population standard deviation for x
sigma.y a single number representing the population standard deviation for y
conf.level confidence level for the returned confidence interval, restricted to lie between zero and one

Details

If y is NULL, a one-sample z-test is carried out with x. If y is not NULL, a standard two-sample z-test is performed.

Value

A list of class htest, containing the following components:

statistic the z-statistic, with names attribute "z"
p.value the p-value for the test
conf.int is a confidence interval (vector of length 2) for the true mean or difference in means. The confidence level is recorded in the attribute conf.level. When alternative is not "two.sided", the confidence interval will be half-infinite, to reflect the interpretation of a confidence interval as the set of all values k for which one would not reject the null hypothesis that the true mean or difference in means is k. Here infinity will be represented by Inf.
estimate vector of length 1 or 2, giving the sample mean(s) or mean of differences; these
estimate the corresponding population parameters. Component estimate has a
names attribute describing its elements.
null.value is the value of the mean or difference in means specified by the null hypothe-
sis. This equals the input argument mu. Component null.value has a names
attribute describing its elements.
alternative records the value of the input argument alternative: "greater", "less" or
"two.sided".
data.name a character string (vector of length 1) containing the actual names of the input
vectors x and y

Null Hypothesis

For the one-sample z-test, the null hypothesis is that the mean of the population from which x is
drawn is mu. For the standard two-sample z-tests, the null hypothesis is that the population mean
for x less that for y is mu.

The alternative hypothesis in each case indicates the direction of divergence of the population mean
for x (or difference of means for x and y) from mu (i.e., "greater", "less", "two.sided").

Author(s)

Alan T. Arnholt

References

Canada: Macmillan.
University Press.

See Also

zsum.test, tsum.test

Examples

x <- rnorm(12)
z.test(x, sigma.x=1)
  # Two-sided one-sample z-test where the assumed value for
  # sigma.x is one. The null hypothesis is that the population
  # mean for 'x' is zero. The alternative hypothesis states
  # that it is either greater or less than zero. A confidence
  # interval for the population mean will be computed.
x <- c(7.8, 6.6, 6.5, 7.4, 7.3, 7., 6.4, 7.1, 6.7, 7.6, 6.8)
y <- c(4.5, 5.4, 6.1, 6.1, 5.4, 5., 4.1, 5.5)
z.test(x, sigma.x=0.5, y, sigma.y=0.5, mu=2)
  # Two-sided standard two-sample z-test where both sigma.x and sigma.y are both assumed to equal 0.5.
  # The null hypothesis is that the population mean for 'x' less that for 'y' is 2.
  # The alternative hypothesis is that this difference is not 2.
  # A confidence interval for the true difference will be computed.

z.test(x, sigma.x=0.5, y, sigma.y=0.5, conf.level=0.90)
  # Two-sided standard two-sample z-test where both sigma.x and sigma.y are both assumed to equal 0.5.
  # The null hypothesis is that the population mean for 'x' less that for 'y' is zero.
  # The alternative hypothesis is that this difference is not zero. A 90% confidence interval for the true difference will be computed.

rm(x, y)

---

zsum.test

### Summarized z-test

#### Description

This function is based on the standard normal distribution and creates confidence intervals and tests hypotheses for both one and two sample problems based on summarized information the user passes to the function. Output is identical to that produced with `z.test`.

#### Usage

```r
zsum.test(
  mean.x,
  sigma.x = NULL,
  n.x = NULL,
  mean.y = NULL,
  sigma.y = NULL,
  n.y = NULL,
  alternative = "two.sided",
  mu = 0,
  conf.level = 0.95
)
```

#### Arguments

- **mean.x**: a single number representing the sample mean of x
- **sigma.x**: a single number representing the population standard deviation for x
- **n.x**: a single number representing the sample size for x
- **mean.y**: a single number representing the sample mean of y
sigma.y  a single number representing the population standard deviation for y
n.y     a single number representing the sample size for y
alternative is a character string, one of "greater", "less" or "two.sided", or the initial letter of each, indicating the specification of the alternative hypothesis. For one-sample tests, alternative refers to the true mean of the parent population in relation to the hypothesized value mu. For the standard two-sample tests, alternative refers to the difference between the true population mean for x and that for y, in relation to mu.
mu       a single number representing the value of the mean or difference in means specified by the null hypothesis
conf.level confidence level for the returned confidence interval, restricted to lie between zero and one

Details
If y is NULL, a one-sample z-test is carried out with x. If y is not NULL, a standard two-sample z-test is performed.

Value
A list of class htest, containing the following components:
statistic  the z-statistic, with names attribute z.
p.value    the p-value for the test
conf.int   is a confidence interval (vector of length 2) for the true mean or difference in means. The confidence level is recorded in the attribute conf.level. When alternative is not "two.sided", the confidence interval will be half-infinite, to reflect the interpretation of a confidence interval as the set of all values k for which one would not reject the null hypothesis that the true mean or difference in means is k. Here, infinity will be represented by Inf.
estimate  vector of length 1 or 2, giving the sample mean(s) or mean of differences; these estimate the corresponding population parameters. Component estimate has a names attribute describing its elements.
null.value the value of the mean or difference in means specified by the null hypothesis. This equals the input argument mu. Component null.value has a names attribute describing its elements.
alternative records the value of the input argument alternative: "greater", "less" or "two.sided".
data.name  a character string (vector of length 1) containing the names x and y for the two summarized samples

Null Hypothesis
For the one-sample z-test, the null hypothesis is that the mean of the population from which x is drawn is mu. For the standard two-sample z-tests, the null hypothesis is that the population mean for x less that for y is mu.

The alternative hypothesis in each case indicates the direction of divergence of the population mean for x (or difference of means of x and y) from mu (i.e., "greater", "less", "two.sided").
Author(s)

Alan T. Arnholt

References


See Also

*z.test*, *tsum.test*

Examples

```r
zsum.test(mean.x=56/30,sigma.x=2, n.x=30, alternative="greater", mu=1.8)
# Example 9.7 part a. from PASWR.
x <- rnorm(12)
zsum.test(mean(x),sigma.x=1,n.x=12)
  # Two-sided one-sample z-test where the assumed value for 
  # sigma.x is one. The null hypothesis is that the population 
  # mean for 'x' is zero. The alternative hypothesis states 
  # that it is either greater or less than zero. A confidence 
  # interval for the population mean will be computed. 
  # Note: returns same answer as: 
z.test(x,sigma.x=1)

x <- c(7.8, 6.6, 6.5, 7.4, 7.3, 7.0, 6.4, 7.1, 6.7, 7.6, 6.8)
y <- c(4.5, 5.4, 6.1, 6.1, 5.4, 5.0, 4.1, 5.5)
zsum.test(mean(x), sigma.x=0.5, n.x=11 ,mean(y), sigma.y=0.5, n.y=8, mu=2)
  # Two-sided standard two-sample z-test where both sigma.x 
  # and sigma.y are both assumed to equal 0.5. The null hypothesis 
  # is that the population mean for 'x' less that for 'y' is 2. 
  # The alternative hypothesis is that this difference is not 2. 
  # A confidence interval for the true difference will be computed. 
  # Note: returns same answer as: 
z.test(x, sigma.x=0.5, y, sigma.y=0.5)

zsum.test(mean(x), sigma.x=0.5, n.x=11, mean(y), sigma.y=0.5, n.y=8, conf.level=0.90)
  # Two-sided standard two-sample z-test where both sigma.x and 
  # sigma.y are both assumed to equal 0.5. The null hypothesis 
  # is that the population mean for 'x' less that for 'y' is zero. 
  # The alternative hypothesis is that this difference is not 
  # zero. A 90% confidence interval for the true difference will 
```

# be computed. Note: returns same answer as:
> z.test(x, sigma.x=0.5, y, sigma.y=0.5, conf.level=0.90)
> rm(x, y)
Index

* datasets
  Abbey, 8
  Abc, 9
  Abilene, 10
  Ability, 11
  Abortion, 12
  Absent, 13
  Achieve, 14
  Adsales, 14
  Aggress, 15
  Aid, 16
  Aids, 17
  Airdisasters, 18
  Airline, 19
  Alcohol, 20
  Allergy, 20
  Anesthet, 21
  Anxiety, 22
  Apolipop, 22
  Append, 23
  Appendec, 24
  Aptitude, 25
  Archaeo, 25
  Arthritis, 26
  Artifici, 27
  Aspin, 28
  Asthamati, 28
  Attorney, 29
  Autogear, 30
  Backtoback, 31
  Bbsalaries, 31
  Bigten, 32
  Biology, 33
  Birth, 34
  Blackedu, 35
  Blood, 35
  Board, 36
  Bones, 37
  Books, 38
  Bookstor, 38
  Brain, 39
  Bumpers, 40
  Bus, 41
  Bypass, 41
  Cabinets, 42
  Cancer, 43
  Carbon, 44
  Cat, 45
  Censored, 45
  Challeng, 46
  Chemist, 47
  Chesapea, 48
  Chevy, 49
  Chicken, 50
  Chipavg, 50
  Chips, 51
  Cigar, 52
  Cigarett, 53
  Citrus, 55
  Clean, 56
  Coaxial, 57
  Coffee, 58
  Coins, 58
  Commute, 60
  Concept, 61
  Concrete, 61
  Corn, 62
  Correlat, 63
  Counsel, 64
  Cpi, 64
  Crime, 65
  Darwin, 66
  Dealers, 67
  Defectiv, 67
  Degree, 68
  Depend, 70
  Detroit, 70

278
| Develop, 71       | Greenriv, 112     |
| Devmath, 72      | Grnriv2, 113     |
| Dice, 72         | Groupabc, 114    |
| Diesel, 73       | Groups, 114      |
| Diplomat, 74     | Gym, 115         |
| Disposal, 75     | Habits, 116      |
| Dogs, 76         | Haptoglo, 117    |
| Domestic, 77     | Hardware, 117    |
| Dopamine, 78     | Hardwood, 118    |
| Dowjones, 79     | Heat, 119        |
| Drink, 80        | Heating, 120     |
| Drug, 80         | Hodgkin, 121     |
| Dyslexia, 81     | Homes, 122       |
| Earthqk, 82      | Homework, 123    |
| Educat, 84       | Honda, 124       |
| Eggs, 85         | Hostile, 124     |
| Elderly, 85      | Housing, 125     |
| Energy, 86       | Hurrican, 126    |
| Engineer, 87     | Iceberg, 127     |
| Entrance, 88     | Income, 128      |
| Epaminicompact, 89 | Independent, 129 |
| Epatwoseater, 90 | Indian, 130      |
| Executiv, 91     | Indiapol, 131    |
| Exercise, 91     | Indy500, 131     |
| Fabric, 92       | Inflatio, 132    |
| Faithful, 93     | Inletoil, 133    |
| Family, 94       | Inmate, 134      |
| Ferrarol, 95     | Inspect, 135     |
| Ferraroo, 95     | Insulate, 136    |
| Fertility, 96    | Iqgpa, 137       |
| Firstchi, 97     | Irises, 137      |
| Fish, 98         | Jdpower, 138     |
| Fitness, 99      | Jobsat, 139      |
| Florida2000, 100 | Kidsmoke, 140    |
| Fluid, 101       | Kilowatt, 141    |
| Food, 102        | Kinder, 141      |
| Framingh, 102    | Laminect, 142    |
| Freshman, 103    | Lead, 143        |
| Funeral, 104     | Leader, 144      |
| Galaxie, 105     | Lethal, 144      |
| Gallup, 105      | Life, 145        |
| Gasoline, 106    | Lifespan, 146    |
| German, 107      | Lightmonth, 146  |
| Golf, 108        | Lodge, 147       |
| Governor, 109    | Longtail, 148    |
| Gpa, 110         | Lowabil, 149     |
| Grades, 111      | Magnesiu, 149    |
| Graduate, 112    | Malpract, 150    |
Manager, 151
Marked, 151
Math, 152
Mathcomp, 153
Mathpro, 154
Maze, 155
Median, 155
Mental, 156
Mercury, 157
Metrent, 157
Miller, 158
Miller1, 159
Moisture, 159
Monoxide, 160
Movie, 161
Music, 162
Name, 163
Nascar, 164
Nervous, 164
Newsstand, 165
Nfldraft, 166
Nfldraft2, 166
Nicotine, 167
Orange, 171
Orioles, 171
Oxytocin, 172
Parented, 173
Patrol, 174
Pearson, 175
Phone, 175
Poison, 176
Politic, 177
Pollutio, 178
Porosity, 178
Poverty, 179
Precinct, 180
Prejudic, 181
Presiden, 181
Press, 182
Prognost, 183
Program, 184
Psat, 184
Psych, 185
Puerto, 186
Quail, 186
Quality, 187
Rainks, 188
Randd, 189
Rat, 189
Ratings, 190
Reaction, 191
Reading, 192
Readiq, 192
Referend, 193
Region, 194
Register, 195
Rehab, 195
Remedial, 196
Rentals, 197
Repair, 198
Retail, 198
Ronbrown1, 199
Ronbrown2, 200
Rural, 200
Salary, 201
Salinity, 202
Sat, 203
Saving, 204
Scales, 204
Schizop, 205
Schizop2, 206
Seatbelt, 207
Selfdefe, 208
Senior, 208
Sentence, 209
Shkdrug, 210
Shock, 211
Shoplift, 211
Short, 212
Shuttle, 213
Simpson, 216
Situp, 217
Skewed, 218
Skin, 218
Slc, 219
Smokyp, 220
Snore, 221
Snow, 222
Soccer, 223
Social, 223
Sophomor, 224
South, 225
Speed, 225
Spellers, 226
Spelling, 227
Sports, 227
Spouse, 228
Stable, 230
Stamp, 230
Statclas, 231
Statelaw, 232
Statisti, 232
Step, 233
Stress, 234
Study, 235
Submarin, 235
Subway, 236
Sunspot, 237
Superbowl, 238
Supercar, 238
Tablrock, 239
Teacher, 241
Tenness, 242
Tensile, 243
Test1, 243
Thermal, 244
Tiaa, 245
Ticket, 245
Toaster, 246
Tonsils, 247
Tort, 248
Toxic, 249
Track, 250
Track15, 251
Treatments, 251
Trees, 252
Trucks, 253
Tv, 257
Twin, 257
Undergrad, 258
Vacation, 259
Vaccine, 260
Vehicle, 261
Verbal, 261
Victoria, 262
Viscosit, 263
Visual, 264
Vocab, 264
Wastewat, 265
Weather94, 266
Wheat, 267
Windmill, 268
Window, 268
Wins, 269

Wool, 270
Yearsunspot, 271

* distribution
  CIsim, 54
  Combinations, 59
  normarea, 168
  ntester, 170
  SRS, 229

* htest
  tsum.test, 253
  z.test, 271
  zsum.test, 274

* univar
  EDA, 83
  nsizer, 169

Abbey, 8
Abc, 9
Abilene, 10
Ability, 11
Abortion, 12
Absent, 13
Achieve, 14
Adsales, 14
Aggress, 15
Aid, 16
Aids, 17
Airdisasters, 18
Airline, 19
Alcohol, 20
Allergy, 20
Anesthet, 21
Anxiety, 22
Apolipop, 22
Append, 23
Appendec, 24
Aptitude, 25
Archaeo, 25
Arthriti, 26
Artifici, 27
Asprin, 28
Asthmati, 28
Attorney, 29
Autogear, 30

Backtoback, 31
Bbsalaries, 31
Bigten, 32
Biology, 33
INDEX

Birth, 34
Blackedu, 35
Blood, 35
Board, 36
Bones, 37
Books, 38
Bookstor, 38
Brain, 39
Bumpers, 40
Bus, 41
Bypass, 41
Cabinets, 42
Cancer, 43
Carbon, 44
Cat, 45
Censored, 45
Challeng, 46
Chemist, 47
Chesapea, 48
Chevy, 49
Chicken, 50
Chipavg, 50
Chips, 51
Cigar, 52
Cigarett, 53
CIsim, 54
Citrus, 55
Clean, 56
Coaxial, 57
Coffee, 58
Coins, 58
Combinations, 59, 229
Commute, 60
Concept, 61
Concrete, 61
Corn, 62
Correlat, 63
Counsel, 64
Cpi, 64
Crime, 65
Darwin, 66
Dealers, 67
Defectiv, 67
Degree, 68
Delay, 69
Depend, 70
Detroit, 70
Develop, 71
Devmath, 72
Dice, 72
Diesel, 73
Diplomat, 74
Disposal, 75
Dogs, 76
Domestic, 77
Dopamine, 78
Dowjones, 79
Drink, 80
Drug, 80
Dyslexia, 81
Earthqk, 82
EDA, 83
Educat, 84
Eggs, 85
Elderly, 85
Energy, 86
Engineer, 87
Entrance, 88
Epaminicompact, 89
Epatwoseater, 90
Executiv, 91
Exercise, 91
Fabric, 92
Faithful, 93
Family, 94
Ferraro1, 95
Ferraror2, 95
Fertility, 96
Firstchi, 97
Fish, 98
Fitness, 99
Florida2000, 100
Fluid, 101
Food, 102
Framingham, 102
Freshman, 103
Funeral, 104
Galaxie, 105
Gallup, 105
Gasoline, 106
German, 107
Golf, 108
Governor, 109
<table>
<thead>
<tr>
<th>Precinct, 180</th>
<th>Simpson, 216</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prejudic, 181</td>
<td>Situp, 217</td>
</tr>
<tr>
<td>President, 181</td>
<td>Skewed, 218</td>
</tr>
<tr>
<td>Press, 182</td>
<td>Skin, 218</td>
</tr>
<tr>
<td>Prognost, 183</td>
<td>SIC, 219</td>
</tr>
<tr>
<td>Program, 184</td>
<td>Smokephy, 220</td>
</tr>
<tr>
<td>Psat, 184</td>
<td>Snore, 221</td>
</tr>
<tr>
<td>Psych, 185</td>
<td>Snow, 222</td>
</tr>
<tr>
<td>Puerto, 186</td>
<td>Soccer, 223</td>
</tr>
<tr>
<td>Quail, 186</td>
<td>Social, 223</td>
</tr>
<tr>
<td>Quality, 187</td>
<td>Sophomore, 224</td>
</tr>
<tr>
<td>Rainks, 188</td>
<td>South, 225</td>
</tr>
<tr>
<td>Randd, 189</td>
<td>Speed, 225</td>
</tr>
<tr>
<td>Rat, 189</td>
<td>Spellers, 226</td>
</tr>
<tr>
<td>Ratings, 190</td>
<td>Spelling, 227</td>
</tr>
<tr>
<td>Reaction, 191</td>
<td>Sports, 227</td>
</tr>
<tr>
<td>Reading, 192</td>
<td>Spouse, 228</td>
</tr>
<tr>
<td>Readiq, 192</td>
<td>SRS, 59, 229</td>
</tr>
<tr>
<td>Referend, 193</td>
<td>Stable, 230</td>
</tr>
<tr>
<td>Region, 194</td>
<td>Stamp, 230</td>
</tr>
<tr>
<td>Register, 195</td>
<td>Statclass, 231</td>
</tr>
<tr>
<td>Rehab, 195</td>
<td>Statelaw, 232</td>
</tr>
<tr>
<td>Remedial, 196</td>
<td>Statisti, 232</td>
</tr>
<tr>
<td>Rentals, 197</td>
<td>Step, 233</td>
</tr>
<tr>
<td>Repair, 198</td>
<td>Stress, 234</td>
</tr>
<tr>
<td>Retail, 198</td>
<td>Study, 235</td>
</tr>
<tr>
<td>Ronbrown1, 199</td>
<td>Submarin, 235</td>
</tr>
<tr>
<td>Ronbrown2, 200</td>
<td>Subway, 236</td>
</tr>
<tr>
<td>Rural, 200</td>
<td>Sunspot, 237</td>
</tr>
<tr>
<td>Salary, 201</td>
<td>Superbowl, 238</td>
</tr>
<tr>
<td>Salinity, 202</td>
<td>Supercar, 238</td>
</tr>
<tr>
<td>Sat, 203</td>
<td>Tablrock, 239</td>
</tr>
<tr>
<td>Saving, 204</td>
<td>Teacher, 241</td>
</tr>
<tr>
<td>Scales, 204</td>
<td>Tenness, 242</td>
</tr>
<tr>
<td>Schizop2, 205</td>
<td>Tensile, 243</td>
</tr>
<tr>
<td>Schizoph, 206</td>
<td>Test1, 243</td>
</tr>
<tr>
<td>Seatbelt, 207</td>
<td>Thermal, 244</td>
</tr>
<tr>
<td>Selfdefe, 208</td>
<td>TIAA, 245</td>
</tr>
<tr>
<td>Senior, 208</td>
<td>Ticket, 245</td>
</tr>
<tr>
<td>Sentence, 209</td>
<td>Toaster, 246</td>
</tr>
<tr>
<td>Shkdrug, 210</td>
<td>Tonsils, 247</td>
</tr>
<tr>
<td>Shock, 211</td>
<td>Tort, 248</td>
</tr>
<tr>
<td>Shoplift, 211</td>
<td>Toxic, 249</td>
</tr>
<tr>
<td>Short, 212</td>
<td>Track, 250</td>
</tr>
<tr>
<td>Shuttle, 213</td>
<td>Track15, 251</td>
</tr>
<tr>
<td>SIGN.test, 214</td>
<td>Treatments, 251</td>
</tr>
<tr>
<td></td>
<td>Trees, 252</td>
</tr>
<tr>
<td></td>
<td>Trucks, 253</td>
</tr>
</tbody>
</table>
INDEX

tsum.test, 216, 253, 273, 276
Tv, 257
Twin, 257

Undergrad, 258

Vacation, 259
Vaccine, 260
Vehicle, 261
Verbal, 261
Victoria, 262
Viscosit, 263
Visual, 264
Vocab, 264

Wastewat, 265
Weather94, 266
Wheat, 267
Windmill, 268
Window, 268
Wins, 269
Wool, 270

Yearsunspot, 271

z.test, 216, 256, 271, 276
zsum.test, 216, 256, 273, 274