Cheatography

R Cheat Sheet V2: Electric Boogaloo Cheat Sheet

by Jeffrey M. Pisklak (non_human_entity) via cheatography.com/127433/cs/25553/

Operators

=	Assigns a value to an object
<-	
x > y	x greater than y
x < y	x is less than y
x >= y	x greater than or equal to y
x <= y	x is less than or equal to y
!= x	not equal to x
!x	not x
x | y	x OR y
x & y	x AND y

Basic R Functions

Access a function's help file

```
help(function name)
```

Load a csv file

```
read.csv("snails.csv", header = TRUE, row.names = NULL)
```

Install a library

```
install.packages("library name")
```

Load an installed library

```
library(library name)
```

Resize images in Jupyter and Google Collab

```
options(repr.plot.width = x, repr.plot.height = y)
```

Return the amount of values in x

```
length(x)
```

Return the number of rows in a dataframe

```
nrow(df)
```

Return the absolute value(s) in x

```
abs(x)
```

Sponsored by **CrosswordCheats.com**
Learn to solve cryptic crosswords!
http://crosswordcheats.com

Basic R Functions (cont)

Return the sum of all the values in x

```
sum(x)
```

Return the square-root of the value(s) in x

```
sqrt(x)
```

Return the mean of the values in x with optional arguments for trimming and removing NAs

```
mean(x, tr = 0, na.rm = FALSE)
```

Return the median of the values in x with optional arguments removing NAs

```
median(x, na.rm = FALSE)
```

Return the sample standard deviation of values in x with optional argument for removing NAs

```
sd(x, na.rm = FALSE)
```

Return the sample variance of values in x with optional argument for removing NAs

```
var(x, na.rm = FALSE)
```

Return the quartiles for x with optional argument for removing NAs

```
quantile(x, na.rm = FALSE)
```

Sort the values of x into ascending order

```
sort(x)
```

Compute the median absolute deviation of x with optional argument to remove NAs

```
mad(x, na.rm = FALSE)
```

Find NA values in x (returns TRUE/FALSE)

```
is.na(x)
```

Paste things together into a single string

```
paste(x, y, z, sep = " ")
```

Create a table of counts

Examples:

```
table(x)
table(x, y)
```

Sponsored by **CrosswordCheats.com**
Learn to solve cryptic crosswords!
http://crosswordcheats.com

Data Frames

Create a new data frame

```
Column_1 = c("A", "B", "C")
Column_2 = c(21, 22, NA)
new_df = data.frame(Column_1, Column_2)
```

Add a column

```
new_df$Column_3 = c(51, 52, 53)
```

Select a specific value (e.g., 52 = row 2, column 3)

```
new_df[2, 3]
```

Select a series of values (e.g., all of row 2)

```
new_df[2, c(1,2,3)]
```

or

```
new_df[2, ]
```

Select an entire column (e.g., column 2)

```
new_df$Column_2
```

or

```
new_df[ , 2]
```

Isolate column values that are not NAs

```
new_df$Column_2[!is.na(new_df$Column_2)]
```

Subset Function

Used to select specific observations from a dataframe according to a rule you specify.

```
subset(dataframe, subset rule, select = ("columns to keep"))
```

Example:

```
outliers = subset(heightData, Father < 60.1 | Father > 75.3, select = c("Father"))
```

Library Functions

library(car)

Levene's Test

```
leveneTest(data_frame$Response, data_frame$Predictor, center = median)
```

Bootstrapping a Regression Model

```
x = Boot(model, R = 2000)
 hist(x)
 confint(x)
 summary(x)
```

Type III Sum of Squares ANOVA

```
Anova(model, type = "III")
```

Sponsored by **CrosswordCheats.com**
Learn to solve cryptic crosswords!
http://crosswordcheats.com

Library Functions (cont)

library(effsize)

Cohen's d and Hedges g

```
cohen.d (y~x, data, hedges.correction = FALSE)
```

library(plyr)

Aggregate data frames

```
new_df = ddply( data frame, c("Predictor1, Predictor2"), summarise,
   n = length (Score_Column),
   Means = mean(Score_Column) )
```

library(polycor)

Biserial Correlation

```
polyserial(y, x)
```

library(pwr)

Sample Size for a Two-Sample T-test

```
pwr.t.test(d, sig.level, power, type = c("two.sample, " paired "))
```

Sample Size for a One-Way ANOVA

```
pwr.anova.test(k, f, sig.level, power)
```

library(rcompanion)

Calculates lambda for Tukey's ladder of powers

```
transformTukey(x, plotit = FALSE, returnLambda = TRUE)
```

library(WRS2)

Winsorized variance of x

```
winvar(x, tr = .2)
```

Yuen's two sample t-test for trimmed independent means

```
yuen(y ~ x, tr = .2)
```

One-Way Robust Independent ANOVA with bootstrapping: F-tests

```
t1waybt(Response ~ Predictor, data = data, tr = 0.2, nboot = 2000)
```

One-Way Robust Independent ANOVA with bootstrapping: Post Hocs

```
mcppb20(Response~ Predictor, data = data, tr = 0.2, nboot = 2000)
```

Sponsored by **CrosswordCheats.com**
Learn to solve cryptic crosswords!
http://crosswordcheats.com

Library Functions (cont)

Two-Way Robust Independent ANOVA: F-tests

```
t2way(Response ~ Predictor A+ Predictor B + Predictor A : Predictor B, data = depress, tr = 0.2)
```

Two-Way Robust Independent ANOVA: Post-Hocs

```
x = mcp2atm(Response ~ Predictor A+ Predictor B + Predictor A : Predictor B, data = depress, tr = 0.2)

 x$contrasts
 x
```

Distribution Functions

Return the the corresponding *quantile* for a given probability

Normal Distribution

```
qnorm (probability, mean, sd)
```

T Distribution

```
qt(probability, df, lower.tail)
```

F Distribution

```
qf(probability, df1, df2, lower.tail)
```

Chi-Square Distribution

```
qchisq (probability, df, lower.tail)
```

Return the the corresponding *probability* for a given quantile.

Normal Distribution

```
pnorm (quantile, mean, sd)
```

T Distribution

```
pt(quantile, df, lower.tail)
```

F Distribution

```
pf(quantile, df1, df2, lower.tail)
```

Chi-Square Distribution

```
pchisq (quantile, df, lower.tail)
```

Sponsored by **CrosswordCheats.com**
Learn to solve cryptic crosswords!
http://crosswordcheats.com

Regression and ANOVA Functions

Factoring a Predictor	`data_f ram e$P red ictor = factor (da ta_ fra me$ Pre dictor)`
Viewing levels of a factor	`levels (da ta_ fra me$ Pre dictor)`
Linear Model	`model = lm(Res ponse ~ Predictor1 + Predic tor2, data = data)`
Summary output of a linear model	`summar y(m odel)`
Linear Model Confidence Intervals	`confin t(m odel)`
F-test Model Comparisons	`anova( model1, model2, model3, etc...)`
Anova main effects	`summar y(a ov( model))`
Dummy Coding with 1s and 0s	`ifelse (da ta_ fra me$ Pre dictor == " X", 1, 0)`

Contrasts	`cont1 = c(1, 1, -2)`
	`cont2 = c(1, -1, 0)`
	`contra sts (da ta_ fra me$ Pre dictor) = cbind( cont1, cont2)`

Polynomial Contrasts	`contra sts (da ta_ fra me$ Pre dictor) = contr.p ol y(l eve ls( dat a_f ram e$P red ictor))`
Post Hoc Tests ("bonferroni", "holm", "BH")	`pairwi se.t.t est (da ta_ fra me$ Res ponse, data_f ram e$P red ictor, p.adju st.m ethod = c("h olm "))`
Tukey HSD	`TukeyH SD( aov (mo del), " Pre dic tor ")`

Note: the `lm()` function stores many useful things as attributes:

```
model$ res iduals
model$ coe ffi cients
```

Common Statistical Tests and Calculations

T-test

```
t.tes  t(y~x, altern  ative = c("t wo.s id ed"), mu = 0, var.equal = FALSE, conf.level = 0.95)
```

Correlation

```
cor(x, y)
```

Goodness-Of Fit (One Variable)

```
chisq.t est(x = observed, p = expected probab ili ties)
```

Sponsored by **CrosswordCheats.com**
Learn to solve cryptic crosswords!
http://crosswordcheats.com

Common Statistical Tests and Calculations (cont)

Pearson's Chi-squared test (Two Variables)

```
chisq.test(table , correct = FALSE)
```

Fisher's Exact Test

```
fisher.test( table)
```

Plotting: library(ggplot2)

Histogram

```
ggplot(dataFrame, aes(x = Dep_Var)) +
     geom_histogram(colour = "black",
     fill = "white")
```

Density Plot

```
ggplot(dataFrame, aes(x = Dep_Var)) +
      geom_density(colour = "black",fill = "pink", adjust = 1)
```

Boxplots

```
ggplot(dataFrame, aes(x = Indep_Var, y = Dep_Var)) +
      geom_boxplot()
```

Barplot with errorbars

```
ggplot(plotData, aes(x = Indep_Var, y = Dep_Var, fill = Indep_Var)) +
     geom_bar(stat = "identity", colour = "black") +
     geom_errorbar(aes(ymin = bottom_value, ymax = top_value), width = .25)
```

Q-Q Plot For two independent samples

Remove `+ facet_wrap()` for a single sample

```
ggplot(dataFrame, aes(sample = Dep_Var)) +
     stat_qq() +
     stat_qq_line() +
     facet_wrap(~ Indep_Var)
```

Line Plot of Means with Two Predictors

```
ggplot(plotData, aes(x = PredictorA, y = Means, group = PredictorB, colour = PredictorB)) +
     geom_line(position = position_dodge( width = 0.4)) +
       geom_point(position = position_dodge( width = 0.4))
```

Scatterplot with Regression Line

```
ggplot(dataframe, aes(x = predictor, y = response)) +
     geom_point() +
     geom_abline( intercept = b0, slope = b1)
```

Sponsored by **CrosswordCheats.com**
Learn to solve cryptic crosswords!
http://crosswordcheats.com