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(  fun  ction name)
```

Load a csv file

```
read.c sv( " sna ils.cs v", header = TRUE, row.names = NULL)
```

Install a library

```
insta  ll.p  a c k ag  es  ( "  library name")
```

Load an installed library

```
libra  ry  ( l  ibrary name)
```

Resize images in Jupyter and Google Collab

```
optio  ns  ( r  ep  r.p  l o t.width = x, repr.p  lo  t.h  eight = y)
```

Return the amount of values in x

```
lengt h(x)
```

Return the number of rows in a dataframe

```
nrow(df)
```

Return the absolute value(s) in x

```
abs(x)
```

By **Jeffrey M. Pisklak**
(non_human_entity)
cheatography.com/non-human-entity/

Not published yet.
Last updated 28th December, 2023.

Sponsored by **Readable.com**
Measure your website readability!
https://readable.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

```
quant  ile(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 **Readable.com**
Measure your website readability!
https://readable.com

Data Frames

Create a new data frame

```
Column_1 = c("A  ", "  B", " C")
Column_2 = c(21, 22, NA)
new_df = data.f  ra m e (C  olu  mn_1, Column_2)
```

Add a column

```
new_d  f$C  ol umn_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_d  f$C  ol u mn_2
```

or

```
new_df[ , 2]
```

Isolate column values that are not NAs

```
new_d  f$C  ol u m n_  2[!  is.n  a(  new  _d f $ Co  lum  n_2)]
```

Subset Function

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

```
subse  t(d  at a f rame, subset rule, select = ("co  lumns to keep"))
```

Example:

```
outliers = subset  (h e i gh  tData, Father < 60.1 | Father > 75.3, select = c("F  at h e r"))
```

Library Functions

library(car)

Levene's Test

```
levene Tes t(d ata _fr ame $Re sponse, data_f ram e$P red ictor, center = median)
```

Bootstrapping a Regression Model

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

Type III Sum of Squares ANOVA

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


Sponsored by **Readable.com**
Measure your website readability!
https://readable.com

Library Functions (cont)

library(effsize)

Cohen's d and Hedges g

```
cohen.d (y~x, data, hedges.co  rr e ction = FALSE)
```

library(plyr)

Aggregate data frames

```
new_df = ddply(  da t a frame, c("P red ictor1, Predic tor 2"), summarise,
   n = length  (S c o re  _Co  lumn),
   Means = mean(S cor e_C olumn) )
```

library(polycor)

Biserial Correlation

```
polyse rial(y, x)
```

library(pwr)

Sample Size for a Two-Sample T-test

```
pwr.t.t est(d, sig.level, power, type = c("t wo.s ample, " pai red "))
```

Sample Size for a One-Way ANOVA

```
pwr.an ova.te st(k, f, sig.level, power)
```

library(rcompanion)

Calculates lambda for Tukey's ladder of powers

```
trans  for  mT u k ey(x, plotit = FALSE, return  Lambda = 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

```
t1wayb t(R esponse ~ Predictor, data = data, tr = 0.2, nboot = 2000)
```

One-Way Robust Independent ANOVA with bootstrapping: Post Hocs

```
mcppb2 0(R esp onse~ Predictor, data = data, tr = 0.2, nboot = 2000)
```


Sponsored by **Readable.com**
Measure your website readability!
https://readable.com

Library Functions (cont)

Two-Way Robust Independent ANOVA: F-tests

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

Two-Way Robust Independent ANOVA: Post-Hocs

```
x = mcp2at m(R esponse ~ Predictor A+ Predictor B + Predictor A : Predictor B, data = depress, tr = 0.2)


 x$cont rasts
 x
```

Distribution Functions

Return the the corresponding *quantile* for a given probability

Normal Distribution

```
qnorm  (pr  ob a b ility, mean, sd)
```

T Distribution

```
qt(pr  oba  bi lity, df, lower.t ail)
```

F Distribution

```
qf(pro bab ility, df1, df2, lower.t ail)
```

Chi-Square Distribution

```
qchisq (pr oba bility, df, lower.t ail)
```

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

Normal Distribution

```
pnorm  (qu antile, mean, sd)
```

T Distribution

```
pt(qua ntile, df, lower.t ail)
```

F Distribution

```
pf(qua ntile, df1, df2, lower.t ail)
```

Chi-Square Distribution

```
pchisq (qu antile, df, lower.t ail)
```


Sponsored by **Readable.com**
Measure your website readability!
https://readable.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$P red icto r))`
Post Hoc Tests ("bonferroni", "holm", "BH")	`pairwi se.t.t est (da ta_ fra me$ 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 **Readable.com**
Measure your website readability!
https://readable.com

Common Statistical Tests and Calculations (cont)

Pearson's Chi-squared test (Two Variables)

```
chisq.t es t(table , correct = FALSE)
```

Fisher's Exact Test

```
fisher.te st( table)
```

Plotting: library(ggplot2)

Histogram

```
ggplo  t(d  at a F rame, aes(x = Dep_Var)) +
     geo m_ h i st  ogr  am ( c olour = "  bl a c k",
     fill = " whi te")
```

Density Plot

```
ggplo  t(d  at a F rame, aes(x = Dep_Var)) +
      geo m_ d e ns  ity  (c olour = "  bl a c k",fill = "  pi n k ", adjust = 1)
```

Boxplots

```
ggplo  t(d  at a F rame, aes(x = Indep_Var, y = Dep_Var)) +
      geo m_ b o xp  lot()
```

Barplot with errorbars

```
ggplo  t(p  lo t Data, aes(x = Indep_Var, y = Dep_Var, fill = Indep_  Var)) +
     geom_b  ar (stat = "  id e n ti  ty", colour = "  bl a c k") +
     geom_e  rr o r ba  r(a  es (ymin = bottom _value, ymax = top_va lue), width = .25)
```

Q-Q Plot For two independent samples
Remove `+ facet_  wrap()` for a single sample

```
ggplo  t(d  at a F rame, aes(sample = Dep_Var)) +
     stat_qq() +
     stat_q  q_ l ine() +
     facet_  wrap(~ Indep_  Var)
```

Line Plot of Means with Two Predictors

```
ggplot (pl otData, aes(x = Predic torA, y = Means, group = Predic torB, colour = Predic torB)) +
     geom_l ine (po sition = positi on_ dod ge( width = 0.4)) +
       geom_p oin t(p osition = positi on_ dod ge( width = 0.4))
```

Scatterplot with Regression Line

```
ggplot (da taf rame, aes(x = predictor, y = response)) +
     geom_p oint() +
     geom_a bli ne( int ercept = b0, slope = b1)
```

By **Jeffrey M. Pisklak**
(non_human_entity)
cheatography.com/non-human-entity/

Not published yet.
Last updated 28th December, 2023.

Sponsored by **Readable.com**
Measure your website readability!
https://readable.com