R CHEATSHEET • MTH107

Class R FAQ

by Derek H. Ogle, revised Oct-16

Data

Get Data

ENTER RAW DATA:

1. Enter data in Excel (variables in columns, individuals in rows, first row has variable names, no spaces or special characters).
2. Save as "Comma Separated Values (*.CSV)" file in your local directory/folder.

DATA PROVIDED BY PROFESSOR:

1. Goto Data Specific to MTH107 on Resources page.
2. Right-click on "data" link and save to your local directory/folder.

Load CSV

1. Start script and save it in the same folder that contains the CSV file.
2. Select Session, Set Working Directory, To Source File Location menus.
3. Copy resulting setwd() code to script.
4. Use read.csv() to load data into dfobj.

```
dfobj <- read.csv("filename.csv")
```

5. Observe structure of data.frame.

```
str(dfobj)
```

Filter Individuals

Individuals that meet a certain condition (or conditions) are filtered from the dfobj data.frame with filterD().

```
newdf <- filterD(dfobj,cond)
```

where cond may be as follows:

```
var == value          # equal to
var != value          # not equal to
var > value           # greater than
var >= value          # greater than or equal
var < value           # less than
var <=v alue          # less than or equal
var %in% c("val","val","val")   # in the list
cond | cond           # either condition met
cond, cond            # both conditions met
```

Individual in row rownum is selected with:

```
dfobj[rownum,]
```

Individual in row rownum is excluded with:

```
dfobj[-rownum,]
```

Exploratory Data Analysis

Univariate

QUANTITATIVE – Summary statistics (mean, median, SD, IQR, etc.) and a histogram for the qvar variable.

```
Summarize(~qvar,data=dfobj,digits=3)
hist(~qvar,data=dfobj,xlab="var label")
```

CATEGORICAL – Frequency and percentage tables and bar chart for the fvar variable.

```
freq1 <- xtabs(~fvar,data=dfobj)
percTable(freq1)
barplot(freq1,xlab="var label",
        ylab="Frequency")
```

QUANTITATIVE BY GROUP – Summary statistics and histograms for the qvar variable separated by groups in the fvar variable.

```
Summarize(qvar~fvar,data=dfobj,digits=3)
hist(qvar~fvar,data=dfobj,xlab="var label")
```

Bivariate

QUANTITATIVE – Correlation (r) and scatterplot for the qvarY and qvarX variables.

```
corr(~qvarY+qvarX,data=dfobj)
plot(qvarY~qvarX,data=dfobj,
     xyab="yvar label",xlab="xvar label")
```

CATEGORICAL – Frequency and percentage tables for the fvarRow and fvarCol variables.

```
freq2 <- xtabs(~fvarRow+fvarCol,
               data=dfobj)
percTable(freq2)              # total/table %
percTable(freq2,margin=1)     # row %
percTable(freq2,margin=2)     # column %
```

Models

Normal Distributions

```
distrib(val,mean=meanval,sd=sdval,
        lower.tail=FALSE,type="q")
```

where

- val is a value of the quantitative variable or area (i.e., percentage as a proportion).
- meanval is the population mean (μ)
- sdval is the standard deviation (σ) or error
- lower.tail=FALSE is included for "right-of" calculations
- type="q" is included for reverse calculations

For SE use (where nval=sample size):

```
sd=sdval/sqrt(nval)
```

Linear Regression

The best-fit line between the respvar response and expvar explanatory variables.

```
( bfl <- lm(respvar~expvar,data=dfobj) )
```

A visual of the best-fit line.

```
fitPlot(bfl,ylab="yvar label",xlab="xvar label")
```

The r^2 value.

```
rSquared(bfl)
```

Predict a value of respvar given the expval value of expvar.

```
predict(bfl,data.frame(expvar=expval)
```

Hypothesis Testing

Quantitative

ONE SAMPLE:

```
z.test(dfobj$qvar,mu=mu0,alt=HAtype,
       conf.level=confval,sd=sdval)
t.test(dfobj$qvar,mu=mu0,alt=HAtype,
       conf.level=confval)
```

- qvar is a quantitative variable in dfobj
- mu0 is the population mean in H_0
- HAtype is "two.sided", "less", or "greater" for not equals, less than, and greater than H_A
- confval is the confidence level (e.g., 0.95)
- sdval is the population stand deviation (σ)

TWO SAMPLE:

```
levenesTest(qvar~fvar,data=dfobj)
t.test(qvar~fvar,data=dfobj,alt=HAtype ,
       conf.level=confval,var.equal=TRUE)
```

- qvar is a quantitative variable in dfobj
- fvar is a factor (categorical) variable in dfobj
- var.equal=TRUE if the population variances are thought to be equal

Categorical

ONE SAMPLE:

Goodness-of-fit test for observed frequencies in freq1 and expected values (or proportions) in exp.p.

```
( gof <- chisq.test(freq1,p=exp.p,
        rescale.p=TRUE,correct=FALSE) )
```

Extract the expected values.

```
gof$expected
```

Extract the residuals.

```
gof$residuals
```

Follow-up confidence intervals.

```
gofCI(gof,digits=3)
```

TWO SAMPLE:

Chi-square from freq2 two-way observed frequency table.

```
( chi <- chisq.test(freq2,correct=FALSE) )
```

Extract the expected values and residuals as for one-sample situation (but using chi instead of gof).