STA 215 R Sheet ggformula How-To

Note: In the How-To Sheet wherever you see something inside < > you want to replace that in the code. For example, if you see < EXPLANATORY > then you want to replace all of this, including the < >, with the name of the explanatory variable.

Note: R does not recognize “ “ marks when copied from Word. You must re-type them inside of R. For example, if you see “<CATEGORY>” you would replace <CATEGORY> with the name of the category, and you would retype the “ at the beginning and the “ at the end within R.

Utilities

Accessing RStudio Server

See the Handout: Accessing RStudio Server

Opening a Software Investigation Program File

See the Handout: Software Investigation Starter Programs

Working in R

See the Handout: Working in RMarkdown

Loading Packages

```
library(mosaic)
```

Note: The Software Investigation starter program will always include a code chunk that loads the mosaic package that includes the functions we need for STA215.

Reading in Data

```
<DATAFRAME> <-read.csv("<FILEPATH>/<DATASET>.csv", header=TRUE)
```

Note: The Software Investigation starter program will always include a code chunk that reads in the data file.

CH 2: Categorical Data

Frequency Table

```
tally(~ <VARIABLE>, data = <DATAFRAME>, format = "percent", margin = TRUE)
```

Note: If you use format = ”count” you will get counts instead of percents

Bar Graph using Percent

```
gf_percents(~ <VARIABLE>, data = <DATAFRAME> )
```

Bar Graph using Count

```
gf_bar(~ <VARIABLE>, data = <DATAFRAME> )
```

Two-Way Table

```
tally(~ <EXPLANATORY> + <RESPONSE>, data = <DATAFRAME>, margin=TRUE)
```

Clustered Bar Graph

```
gf_percents(~ <RESPONSE>, data = <DATAFRAME>, fill = ~<EXPLANATORY>, position = "dodge", denom = ~fill)
```

CH 3: One Quantitative

Basic Numerical Summaries

```
favstats(~ <VARIABLE>, data = <DATAFRAME> , na.rm=TRUE)
```

Percentile

```
quantile(~ <VARIABLE>, data = <DATAFRAME>, prob = c( <PERCENTILES> ), na.rm = TRUE)
```

Note: Replace percentiles with the values you want separated by commas. For ex., c(0.80,0.90,0.95)

Boxplot

```
gf_boxplot(~ <VARIABLE>, data = <DATAFRAME> )
```

Histogram

```
gf_histogram(~ <VARIABLE>, data = <DATAFRAME>, breaks = seq( <START, END, JUMP> ) )
```

Note: start, end, jump define the bars that make up the histogram. You can leave these out and R will make a default histogram.

Basic Numerical Summaries By-Group

```
favstats( <RESPONSE> ~ <EXPLANATORY>, data = <DATAFRAME> )
```

Boxplot By-Group

```
gf_boxplot( <RESPONSE> ~ <EXPLANATORY>, data = <DATAFRAME> )
```

Histogram By-Group

```
gf_histogram(~ <RESPONSE> | <EXPLANATORY>, data = <DATAFRAME>, breaks = seq( <START, END, JUMP> ) )
```

CH 5: Estimation

Confidence Interval on p

```
prop.test(~ <VARIABLE>, data = <DATAFRAME>, conf.level = <CONFIDENCE LEVEL>, success = “<CATEGORY>”)
```

Confidence Interval on µ

```
t_test(~ <VARIABLE>, data = <DATAFRAME>, conf.level = <CONFIDENCE LEVEL> )
```

STA 215 R Sheet ggformula How-To

CH 6: Two Quantitative

Scatterplot

```
gf_point( <RESPONSE> ~ <EXPLANATORY>, data =
<DATAFRAME> )
```

Linear Correlation

```
cor( <RESPONSE>  ~ <EXPLANATORY>, data = <DATAFRAME>,
use = "complete.obs")
```

Linear Regression

```
Model = lm( <RESPONSE> ~ <EXPLANATORY>, data =
<DATAFRAME> )
summary(Model)
```

Note: Instead of printing results of the lm( ) we have saved them to an R object named Model. The summary(Model) prints results.

Scatterplot By-Group

```
gf_point( <RESPONSE> ~ <EXPLANATORY> , data =
<DATAFRAME>, color = ~ <GROUPING VARIABLE> )
```

CH 7: Hypothesis Testing Introduction

χ^2-Test

```
chisq.test(tally( <RESPONSE> ~ <EXPLANATORY>, data =
<DATAFRAME> ))
```

Expected Counts

```
Output = chisq.test(tally( <EXPLANATORY>, ~<RESPONSE> ,
data = <DATAFRAME> ))
Output$expected
```

Note: Instead of printing results of the code we have saved them to an R object named Output. The Output$expected prints out the expected cell counts.

Confidence Interval for the Difference in Two Proportions

```
prop.test( <RESPONSE> ~ <EXPLANATORY> ,
data = <DATAFRAME>, conf.level = <CONFIDENCE LEVEL>,
success = "<CATEGORY>", correct = FALSE)
```

Note: For this code to work the explanatory variable must only have two categories.

CH 8: Hypothesis Testing Means

Create Paired Data Difference Variable

```
<DATAFRAME>  = mutate( <DATAFRAME>, Difference = <X1
VARIABLE> - <X2 VARIABLE> )
```

Paired T-Test and Confidence Interval

```
t_test(~ Difference, data = <DATAFRAME>, conf.level =
<CONFIDENCE LEVEL> )
```

Independent T-Test and CI

```
t_test( <RESPONSE> ~ <EXPLANATORY>, data =
<DATAFRAME>, conf.level = <CONFIDENCE LEVEL> )
```

ANOVA

```
ANOVA = lm( <RESPONSE> ~ <EXPLANATORY>, data =
<DATAFRAME> )
anova(ANOVA)
```

Note: Instead of printing results of the lm( ) we have saved the results to an R object named ANOVA. The anova(ANOVA) prints results.