# STAT 213: ANOVA Lab

Colin Reimer Dawson

Last Revised March 1, 2016

The goal of this lab is to carry out the CHOOSE-FIT-ASSESS-USE process for data with a single categorical explanatory variable and a quantitative response variable, through one-way ANOVA.

(Adapted from Exercise 5.28) Fenthion is a pesticide used against the olive fruit fly in olive groves. It is toxic to humans, so it is important that there be no residue left on the fruit or in olive oil that will be consumed. One theory was that, if there is residue of the pesticide left in the olive oil, it would dissipate over time. Chemists set out to test that theory by taking a random sample of small amounts of olive oil with fenthion residue and measuring the amuont of fenthion in the oil at 3 different times over the year — day 0, day 281, and day 365.

- (a) Two variables given in the dataset Olives (in the Stat2Data package) are fenthion and time. Which variable is the response variable and which is the explanatory variable? Explain.
- (b) Check the conditions necessary for conducting an ANOVA to analyze the amount of fenthion present in the samples. If the conditions are met, report the results of the analysis.
- (c) Transform the amount of fenthion using an exponential transformation (the exp() function in R). Check the conditions necessary for conducting an ANOVA to analyze the exponential of the amount of fenthion present in the samples. If the conditions are met, report the results of the analysis.

## **R** Markdown Workflow

- From the RStudio File menu, select New File > R Markdown.... Select From Template, and pick mosaic fancy. You will want to delete the example stuff in the middle, but leave in place the first two chunks and the last chunk, as well as the bulleted list with session info.
- 2. The first thing you should do after creating a new Markdown document is to save the file. Do File > Save As... and give your document a name.
- 3. Before editing anything, press the Knit HTML button in the document toolbar. If on the server you may get a message about a blocked pop-up. Allow it to show.
- 4. The first thing you should edit is the Title and Author fields at the top of the document. Give the document a title and put your name inside the quotes on the Author line. Re-Knit to see your change reflected in the document.
- 5. In general, every time you make a change, it is a good idea to Knit again. That way, if you get an error message, you know what caused it. As you get more experienced you may find yourself going longer between re-Knits, but at first you should do it very frequently.
- 6. You should **not** include any lines of code in the Markdown document that (1) access documentation for packages or functions (e.g., ?histogram) (2) otherwise cause external windows (other than plots) to pop up, or (3) cause new packages to be downloaded and installed.
- 7. You must include lines of code in the Markdown document that (1) cause packages or datasets to become *visible* (for example, library() and data() lines), (2) read datasets from files (e.g, read.file()), or (3) define variables that you use in later lines.
- 8. In general, include all and only code that is needed to produce the output you want, beginning from a fresh R session at a computer on which needed packages are already installed.

#### Essential R Commands for ANOVA (some require mosaic)

Case selection / defining new variables (e.g., ratio of two others)

```
filter(DATANAME, CONDITION) %>%
    some.function(..., data = .)
mutate(DATANAME, NewVariableName = OldVariable1 / OldVariable2) %>%
    some.function(..., data = .)
## To reorder the levels of the predictor (for plotting purposes)
## (use the actual text values in place of FIRST, SECOND, THIRD, etc.)
newData <-
    mutate(DATANAME,</pre>
```

```
NewX = factor(OldX, levels = c("FIRST", "SECOND", ...)))
```

## Plots

```
## Box and wisker plot of response by group
bwplot(Y ~ X, data = DATANAME)
## Dot plot of response by group (width and layout are optional)
dotPlot(~Y | X, data = DATANAME, width = WIDTH, layout = c(COLUMNS, ROWS))
## Strip plots of response by group
xyplot(Y ~ X, data = DATANAME, type = c("p", "a"))
## Confidence intervals of group means
plotmeans(Y ~ X, data = DATANAME) #requires library("gplots")
```

### Summary Statistics

mean(Y ~ X, data = DATANAME)
sd(Y ~ X, data = DATANAME)
favstats(Y ~ X, data = DATANAME)

#### Modeling and testing

```
my.model <- lm(Y ~ X, data = DATANAME)
plot(my.model, which = 1) ## fitted by residuals
plot(my.model, which = 2) ## quantile-quantile plot
anova(my.model) ## ANOVA table</pre>
```