HW5: Chapter 3, Section 5.5, Section 6.3, Section 6.4

Your Name Here

The code below just loads some packages and makes it so that enough digits are printed that you won't get confused by rounding errors.

```
library(dplyr) # functions like summarize
library(ggplot2) # for making plots
library(mosaic) # convenient interface to t.test function
```

```
## Warning: package 'mosaic' was built under R version 3.5.2
```

```
## Warning: package 'ggformula' was built under R version 3.5.2
```

```
library(readr)
library(gmodels)

options("pillar.sigfig" = 10) # print 10 significant digits in summarize output
```

Problem 1: Bird Calcium

For many animals, the body's ability to use calcium depends on the level of certain sex-related hormones in the blood. The following data set looks at the relationship between hormone supplement (present or absent) and level of calcium in the blood. The subjects were 20 birds. Half the birds got a hormone supplement and the others served as controls. The response is the level of plasma calcium in mg/100 ml.

```
birds <- read.csv("http://www.evanlray.com/data/cobb_doe/bird_calcium_p160.csv") %>%
  transmute(
    hormone = ifelse(hormone == 1, "control", "supplement"),
    plasma = plasma
  )
```

(a) Check the conditions for conducting an analysis of these data with an ANOVA model. You should write an explicit sentence for each condition explaining why it is or isn't satisfied, with justification; if you need more information to make a determination, explain what else you would need to know. If necessary, find a transformation of the data so that the conditions are as well satisfied as possible.

(b) For the purpose of this problem, let's assume that the conditions you checked in part (b) were fairly well satisfied (perhaps after suitable transformation). Conduct a test to find out whether there were any differences in the mean level of plasma calcium for birds taking the hormones and the control group (perhaps after suitable transformation). Please define all parameters involved, state your hypotheses in terms of equations involving the parameters and written sentences explaining what the hypotheses mean in context, and interpret the p-value for your test in terms of strength of evidence against the null hypothesis of the test, stated in context.

(c) Find a 95% confidence interval describing the difference in the centers of the distributions of calcium concentrations between birds without the hormone supplement and birds with the hormone supplement. Interpret your confidence interval in context on the original (untransformed) scale of the data.

Problem 2: Pesticides in olive oil

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 amount of fenthion in the oil at 3 different times over the year: day 0 (the day the sample was taken), day 281, and day 365.

The following R code reads in the data:

```
olives <- read_csv("http://www.evanlray.com/data/stat2/Olives.csv") %>%
  mutate(
    Day = factor(paste0("Day", Day))
  )
```

```
## Parsed with column specification:
## cols(
##   SampleNumber = col_double(),
##   Group = col_double(),
##   Day = col_double(),
##   Fenthion = col_double(),
##   FenthionSulphoxide = col_double(),
##   FenthionSulphone = col_double(),
##   Time = col_double()
## )
```

(a) Two variables in the model are `Fenthion` and `Day`; we will analyze these variables in this problem. Of these variables, which is the explanatory variable and which is the response? Explain.

(b) Check the conditions for conducting an analysis of these data with an ANOVA model. You should write an explicit sentence for each condition explaining why it is or isn't satisfied, with justification; if you need more information to make a determination, explain what else you would need to know. If necessary, find a transformation of the data so that the conditions are as well satisfied as possible.

(c) For the purpose of this problem, let's assume that the conditions you checked in part (b) were fairly well satisfied (perhaps after suitable transformation). Conduct a test to find out whether there were any differences in the mean amount of fenthion at the three different times of year (if necessary, conduct a test about means on the transformed scale). Please define all parameters involved, state your hypotheses in terms of equations involving the parameters and written sentences explaining what the hypotheses mean in context, and interpret the p-value for your test in terms of strength of evidence against the null hypothesis of the test, stated in context.

(d) Find three confidence intervals with a familywise confidence level of 95%: one for the difference between the mean amount of fenthion present at day 0 and the mean amount present at day 281; a second for the difference between the mean amount of fenthion present at day 0 and the mean amount present at day 365; and a third for the difference between the mean amount of fenthion present at day 281 and the mean amount present at day 365. Find the confidence intervals using the Bonferroni adjustment for the familywise confidence level. Interpret your confidence intervals in context. For which pairs of days do the data provide

statistically significant evidence of a difference in means? All of your inferences can be on the transformed scale, if you selected a transformation in part (b).