# STAT120C Homework 3

**Assigned Thursday April 18, 2019**

**Due Thursday April 25, 2019 by 5pm in the Dropbox in DBH**

1. Posted on the course webpage in the *Code* folder is an R script called `OneWaySimulation.R`. Using the script, this problem will explore the Type I error rate of the one-way ANOVA F-test under various settings. In particular, we will consider what happens when some of the ANOVA assumptions are violated.

   (a) Download the script and load it in RStudio. Run through the script line by line, read the comments, and examine the code output.

   (b) For each of the following scenarios, conduct at least 5000 simulations to estimate the Type I error rate. Report the observed Type I error rate for each setting, and compare it to the intended level of the test. What do the simulations results suggest about the robustness of the F-test when the assumption of constant variance is violated?

   i.
   $$\begin{aligned} &Y_{ij} \sim \mathcal{N}(\mu_i, \sigma_i^2) \\ &J_1 = J_2 = J_3 = 30 \\ &\mu_1 = \mu_2 = \mu_3 = 0 \\ &\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1 \end{aligned}$$

   ii.
   $$\begin{aligned} &Y_{ij} \sim \mathcal{N}(\mu_i, \sigma_i^2) \\ &J_1 = J_2 = J_3 = 30 \\ &\mu_1 = \mu_2 = \mu_3 = 0 \\ &\sigma_1^2 = 0.5, \sigma_2^2 = 2, \sigma_3^2 = 4 \end{aligned}$$

   iii.
   $$\begin{aligned} &Y_{ij} \sim \mathcal{N}(\mu_i, \sigma_i^2) \\ &J_1 = 20, J_2 = 40, J_3 = 60 \\ &\mu_1 = \mu_2 = \mu_3 = 0 \\ &\sigma_1^2 = 1, \sigma_2^2 = 3, \sigma_3^2 = 6 \end{aligned}$$

2. (a) Complete the given ANOVA table.

| Source | SS | df | MS | F |
|---|---|---|---|---|
| Treatment | 4.66 | | | 1.674 |
| Error | | | 1.392 | |
| Total | 202.3 | | | |

   (b) What is the $P$-value from the ANOVA F-test for this table?

3. This problem will examine interaction plots for data generated from the two-way ANOVA model using the R script `InteractionPlots.R` posted on the course website.

   (a) Download the script and open it in RStudio. Run through the script, read the comments, and examine the output. The default settings will generate data for a two way ANOVA model with no interactions (i.e. $\delta = 0$ for all $i, j$). Run the script multiple times and observe the variability in the interaction plot. Describe or sketch what the true interaction plot looks like based on the population parameters.

   (b) Set $K = 1000$ and run the script and provide the interaction plot. How does the resulting plot compare to your description of the true interaction plot in (a)?

   (c) Experiment with different choices of $\delta$ to recreate the following interaction plot. You may want to set $K = 100$ or larger to reduce variability in your plots. Provide your resulting plot and the choice of $\delta$ that produced it.

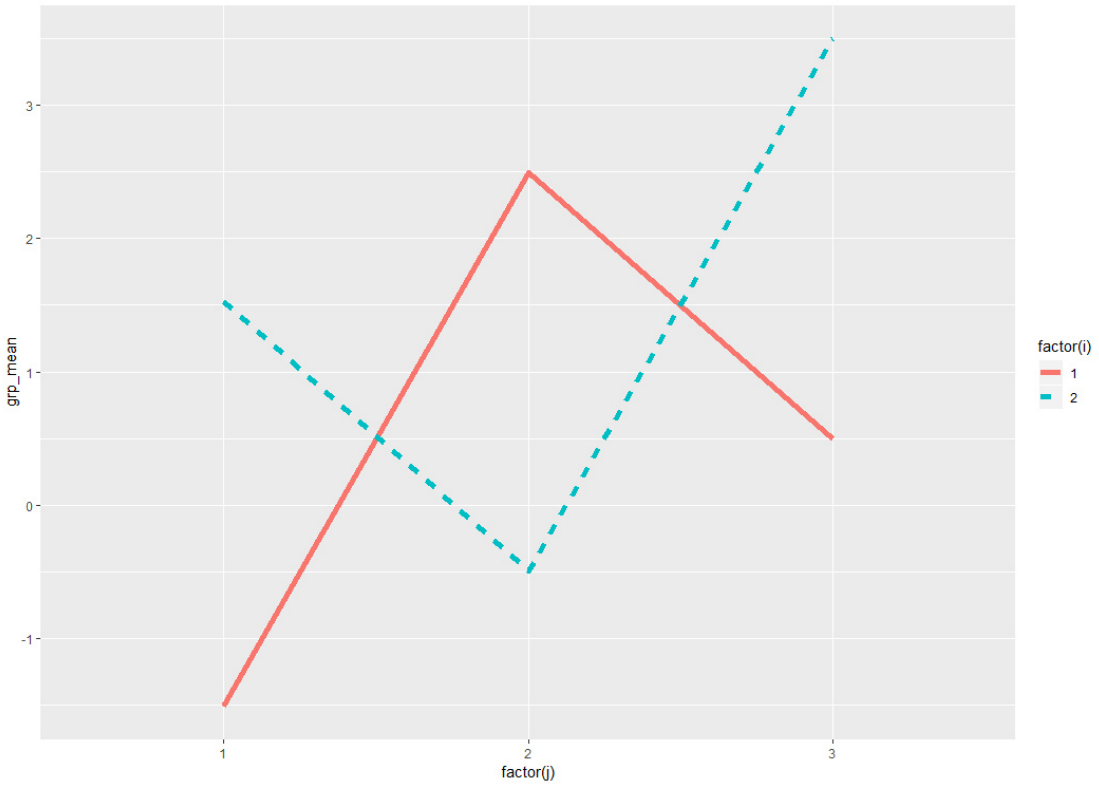


Figure 1: Figure for 3 (c).

4. This problem will analyze the iron retention data discussed in class.

   (a) Use the following R code to load the iron retention data into RStudio. Create boxplots of the mean retention for each combination of dose and form.

```
dat <-
read.csv(url("https://raw.githubusercontent.com/dspluta/STAT120C/master/Data/Chapter%2012/iron.csv"))
```

   (b) Conduct a one-way ANOVA to determine if mean iron retention is different across forms ($Fe2^+$ and $Fe3^+$). Provide the ANOVA table nicely formatted (not just code output), the $P$-value of the test, and precisely state the hypothesis and test conclusion in the context of the data and question of interest.

   (c) Conduct a one-way ANOVA to determine if mean retention is different across doses. Provide the ANOVA table, the $P$-value of the test, and precisely state the hypothesis and test conclusion in the context of the data and question of interest.

   (d) Conduct a two-way ANOVA to test for the presence of interaction effects. Provide the ANOVA table, $P$-value, and precisely state the hypothesis and conclusion of the test.

5. (a) Consider a hypothetical two-way layout with three factors (A, B, C) each at two levels (I, II). Construct a table of cell means for which there is no interaction.

   (b) (Rice 12.5.12) Again consider a hypothetical two-way layout with three factors (A, B, C) each at two levels (I, II). Is it possible for there to interactions but no main effects? Describe the difference between interaction and main effects in your explanation.