MATH 422 – SPRING 2020 – HOMEWORK 6

1. This is exercise 15.1 slightly modified.

   Let $X_{ij} \sim \mathrm{N}(\mu + \alpha_i, \sigma^2)$ for $j = 1, 2, \ldots, n$ and $i = 1, 2, \ldots, k$. This means we have k independent samples, each of size n. The i^{th} sample comes from a normal population with mean $\mu + \alpha_i$ and variance σ^2. Let $\sum_{i=1}^{k} \alpha_i = 0$ so that the average of all the population means is μ.

   Define the sample means $\overline{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$ and the grand mean $\overline{X} = \frac{1}{kn}\sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}$.

   Show that:

$$\mathrm{E}\left[\frac{n}{k-1}\sum_{i=1}^{k}(\overline{X}_i - \overline{X})^2\right] = \sigma^2 + \frac{n}{k-1}\sum_{i=1}^{k}\alpha_i^2$$

   I provide some hints/steps:

   1. Factor out $\frac{n}{k-1}$ and set it aside.
   2. Foil $(\overline{X}_i - \overline{X})^2$.
   3. Since $X_i \sim N(\mu + \alpha_i, \sigma^2)$ for each $i = 1, 2, ..., k$ we have that $\overline{X}_i \sim N(\mu + \alpha_i, \frac{\sigma^2}{n})$
   4. Since we have k samples each of size n, we have that $\overline{X} \sim N(\mu, \frac{\sigma^2}{kn})$.
   5. Also notice that $\overline{X} = \frac{1}{k}\sum_{i=1}^{k}\overline{X}_i$. Thus you can simplify $\sum_{i=1}^{k}\overline{X}_i\overline{X}$.
   6. Distribute the summation $\sum_{i=1}^{k}$ across the foiled expression.
   7. The last two terms should be $-2\sum_{i=1}^{k}\overline{X}_i\overline{X} + \sum_{i=1}^{k}\overline{X}^2$. Show that this simplifies to $-k\overline{X}^2$.
   8. Recall that $\mathrm{E}(X) = \sigma^2 + \mu^2$.
   9. Now you calculate $\mathrm{E}(\overline{X}^2)$ and $\mathrm{E}(\overline{X}_i^2)$ in terms of their means and variances.
   10. Carefully apply the summation to the first term of the foiled expression and take its expected value.
   11. Now bring back in the $\frac{n}{k-1}$ factor and simplify.

2. Exercise 15.17. Perform the one-way ANOVA by your choice of method and interpret the result.

3. Exercise 15.19 - Just perform the one-way ANOVA and interpret the result.

   (a) Perform the one-way ANOVA by your choice of method and interpret the result.

   (b) Calculate SST, $SSTr$, and SSE. Using my R code provided may help. But you are free to calculate these any way you prefer. Check your answers against the built in R method `summary(aov())`.

4. (a) Load the state covid data into R using whatever method you are familiar with. Here are some options:

   - Open it in Excel and copy it to clipboard and use: `d = read.table("clipboard",header=T)`
   - Navigate to it from R's file window and right click it and select "Import Dataset".

     ```
     library(readxl)
     d <- read_excel("C:/<...insert-file-location...>/state_covid_data.xlsx")
     ```

   (b) Perform analysis of variance on the daily proportion of positive tests out of all tests for that day. We'll group them by state and also by data.

   1. Calculate the daily percentage of positive tests: `x=d$day_pos_tests/d$day_tot_tests`. We'll have a problem if any of the data has zero new tests on a particular day. We can fix this by setting the proportion as zero for those days using `x[is.nan(x)] <- 0`.
   2. Perform one-way ANOVA on this against State and Date.
   3. Interpret your results.