Homework 6 (due May 30)

In your solution include your name and the homework number. Please staple your pages together. When solving the problems below, give detailed derivations in order to receive credit.

1. *(Sampling Distribution.)* Before the midterm, we learned that for i.i.d. observations $y_1, ..., y_n \sim N(0, \sigma^2)$, $\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sigma^2} \sim \chi^2_{n-1}$. Review the derivation of this result in your notes.

2. *(ANOVA.)* Today, we learned one-way ANOVA. In this setting, we have independent observations $y_{ij} \sim N(\mu + \alpha_i, \sigma^2)$ for $i = 1, ..., n$ and $j = 1, ..., m$. We also imposed the constraint $\sum_{i=1}^n \alpha_i = 0$ for the sake of identifiability.

   (a) Under the null hypothesis that $\alpha_1 = \alpha_2 = ... = \alpha_n$, together with the constraint $\sum_{i=1}^n \alpha_i = 0$, show $\alpha_i = 0$ for each $i = 1, ..., n$.

   (b) Show $\frac{\sum_{j=1}^m (y_{ij} - \bar{y}_{i\cdot})^2}{\sigma^2} \sim \chi^2_{m-1}$ under the null. You can use the sampling distribution theorem.

   (c) Show $\frac{\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y}_{i\cdot})^2}{\sigma^2} \sim \chi^2_{n(m-1)}$ under the null. This is the first conclusion in the theorem you learned today.

   (d) Define $x_i = \sqrt{m}\bar{y}_{i\cdot}$, and show $m \sum_{i=1}^n (\bar{y}_{i\cdot} - \bar{y})^2 = \sum_{i=1}^n (x_i - \bar{x})^2$.

   (e) What is the distribution of x_i under the null?

   (f) Show $\frac{m \sum_{i=1}^n (\bar{y}_{i\cdot} - \bar{y})^2}{\sigma^2} \sim \chi^2_{n-1}$ under the null using the sampling distribution theorem. This proves the second conclusion.

   (g) Calculate $\mathsf{Cov}(y_{ij} - \bar{y}_{i\cdot}, \bar{y}_{i\cdot} - \bar{y})$. Argue that $\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y}_{i\cdot})^2$ is independent of $m \sum_{i=1}^n (\bar{y}_{i\cdot} - \bar{y})^2$. This proves the third conclusion.

   (h) Prove the fourth conclusion.

   (i) Prove the fifth conclusion.