# Homework 5 (Written Section)

1. Suppose that for the model $y_i = \alpha + e_i$, the errors are independent with mean 0. Also suppose that measurements are taken using one device for the first $n_1$ measurements, and then a more precise instrument was used for the next $n_2$ measurements. Thus $\text{Var}(e_i) = \sigma^2$, $i = 1, \ldots, n_1$ and $\text{Var}(e_i) = \sigma^2/2$, $i = n_1 + 1, \ldots, n_1 + n_2$.

   (a) Ignore the fact that the errors have different variances, and derive the least squares estimator for $\hat{\alpha}$ using matrix notation and $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$.

   (b) Derive the weighted least squares estimator for $\alpha$.

   (c) Suppose that $n_1 = n_2$. Compute the expected values and variances of the two estimators above. Which is a better estimator and why?

2. Question 2, Chapter4

3. Question 3, Chapter 4

4. Return to Question 4 from the third homework, about coins being put on a scale. Now suppose that the variance in $Y$ is proportional to the number of coins put on the scale.

   (a) Design an appropriate matrix of weights $\mathbf{W}$.

   (b) Calculated the new least-squares estimates of the weights of the coins using weighted least squares.

   (c) Do these estimates make sense? Explain first why the estimates are unbiased, and second why the individual measurements are now more heavily weighted in the parameter estimates than they were before we used the weightings.

5. Check to see whether each of the following estimators are BLUE for their respective models. Explain why they are or aren't BLUE.

   (a) For the model $y_i = \alpha + e_i$, the errors are iid with mean zero. The estimator of $\alpha$ is $(max(y_i) + min(y_i))/2$.

   (b) The model: $y_{ij} = \alpha_i + e_{ij}$, $i = 1, 2, 3$, $j = 1, \ldots, 5$ is the ANOVA model with 3 groups, and 5 observations per group. The errors are iid with mean 0. The estimator of $\alpha_1$ is $3y_{11} - y_{12} - y_{13} - y_{14} - y_{15}$.

   (c) For the model $y_i = \beta_0 + \beta_1 x_i + e_i$, the errors are iid with mean 0. The four observed values of $x_i$ are $\mathbf{x} = [1\ 2\ 3\ 4]$. The estimator of $\beta_1$ is $\tilde{\beta}_1 = (y_4 + 2y_3 - 2y_2 - y_1)/5$. For this model, do the following:

      i. Is $\tilde{\beta}_1$ unbiased? Show why or why not.

      ii. What is the sampling variance of $\tilde{\beta}_1$?

      iii. Get the usual least squares estimator of $\beta$ and calculate its sampling variance.

iv. Compare the sampling variance of $\tilde{\beta}_1$ with the sampling variance of the usual least squares estimator of $\beta_1$.

6. A mom noticed that girls' shoes are narrower than boys' shoes and wondered whether their feet were also narrower. More specifically, she wondered whether 4th grade girls actually had narrower feet than boys, given the same length foot. She went to her daughter's classroom to measure the children's feet. The measurements taken for the children included foot width, foot length, and sex.

   (a) In order to answer the mom's question, is it necessary to have an interaction term? Why or why not?

   (b) Develop a linear model for this study, interpreting all parameters in the context of the problem.

   (c) Make up some example data, supposing that 5 children were measured, and write down a design matrix that matches your data.

   (d) Write down your hypotheses to be tested in terms of your model parameters. (You don't have any data to conduct the test; just write down the hypotheses.)