Math 505 Lab 5

1. The file `GLScows.txt` contains (hypothetical) data for the milk production of 110 cows. The expected value of the milk production for each cow is α. The first 10 cows are from Farm U, and their production has variance σ^2, while the remaining 100 cows are from Farm V, where the variance is τ^2.

   (a) A naive way to estimate α is to find the sample mean of the 110 production values. Calculate this value.

   (b) Use generalized least squares to estimate α, σ, and τ. Use the theory from problem 1 on page 66 (the equation given in the back of the book for $\hat{\alpha}$ is more efficient than matrix calculation).

   (c) The values of the parameters used to generate the above data were $\alpha = 7000$, $\sigma = 10$, and $\tau = 1000$. With this in mind, compare your results from parts (a) and (b).

2. Consider the generalized linear model,

$$Y = X\beta + \epsilon,$$

   where $\beta \in \mathbb{R}^4$, $E(\epsilon \mid X) = 0$, and $\text{cov}(\epsilon \mid X) = G$, as in Example 2 on page 65. The first column of X is a column of ones. The first column of `GLS2.txt` contains values of Y_i, and the other three columns contain values of X_{ij}, $i = 1, \ldots, 100$, $j = 2, \ldots, 4$.

   (a) Estimate β using ordinary least squares.

   (b) Estimate β and K using generalized least squares. To create the matrix

$$G = \begin{pmatrix} K & & 0 \\ & \ddots & \\ 0 & & K \end{pmatrix},$$

   use the command `kronecker(I,K)`, where `I` is the 50×50 identity matrix.

   (c) The values of the parameters used to generate the above data were $\beta = (50, 12, -20, 2)'$ and

$$K = \begin{pmatrix} 1 & 5 \\ 5 & 100 \end{pmatrix}.$$

   With this in mind, compare your results from parts (a) and (b).

3. Consider the generalized linear model

$$Y_i = \beta_1 + \beta_2 X_i + \epsilon_i, \text{ for } i = 1, \ldots, 5,$$

   where $E(\epsilon \mid X) = 0$, and

$$\text{cov}(\epsilon \mid X) = \begin{pmatrix} \frac{\sigma^2}{w_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\sigma^2}{w_5} \end{pmatrix},$$

   for a vector of weights w. The columns of the file `GLS3.txt` are the vectors Y, X, and w.

(a) Estimate β using OLS.

(b) Estimate β and σ using GLS, as described in class or in problem 2 on page 67 (Notation: $\lambda = \sigma^2$ and $c_i = \frac{1}{w_i}$).

(c) Estimate β and σ using the command `lm(Y~X,weights=w)`. Your results should be the same as those from part (b).

(d) The parameters used to generate this data where $\beta = (100, 15)'$ and $\sigma = 50$. Which method provided better estimates, OLS or GLS?