BIOSTATISTICS 667 (Fall 2019)
Homework 5

1. This question relates to the issue of "adjusting for baseline". Suppose that, in a longitudinal study, "group" is coded as $x_i = 0$ or $x_i = 1$. The index i is for subjects. The response vector $(Y_{i1}, Y_{i,2})^\top$ is distributed as multivariate normal. In group 0 ($x_i = 0$) the mean vector is $(50, 40)^\top$ and the covariance matrix is

$$\begin{bmatrix} 20 & 10 \\ 10 & 30 \end{bmatrix}.$$

In group 1 ($x_i = 1$) the mean vector is $(50, 30)^\top$ and the covariance matrix is

$$\begin{bmatrix} 40 & 20 \\ 20 & 30 \end{bmatrix}.$$

Define $D_i = Y_{i2} - Y_{i1}$.

   (a) Does this seem to be a randomized study? Why?

   (b) Compute α_1 and α_2 in
$$\mathrm{E}[D_i; x_i] = \alpha_1 + \alpha_2 x_i.$$
   Also compute $\mathrm{var}(D_i; x_i)$ within each group.

   (c) Interpret α_2.

   (d) Obtain the equation for $\mathrm{E}[D_i | Y_{i1}; x_i]$ separately for each group.

   (e) Write the two equations as a single equation.

   (f) Write a single equation for $\mathrm{E}[Y_{i2} | Y_{i1}; x_i]$.

   (g) Is $\mathrm{var}(Y_{i2} | Y_{i1}; x_i) = \mathrm{var}(D_i | Y_{i1}; x_i)$?

   (h) Explain whether the "parallel lines" (no interaction) assumption holds or not.

   (i) If the model
$$\mathrm{E}[D_i | Y_{i1}; x_i] = \beta_1 + \beta_2 x_i + \beta_3 Y_{i1}$$
   is estimated using OLS, will the estimator of β_2 be an unbiased estimator of α_2?

   (j) Compute $\mathrm{var}(D_i | Y_{i1}; x_i)$ within each group.

   (k) If the two groups are of equal size, estimate the efficiency gain brought about by adjusting for the baseline.

2. Describe and explain what program "iml04.sas" does.

3. TBA