Biost/Stat 533
Spring 2008

Homework Assignment #9
(due in class Wednesday, June 4, 2008)

Homework:

1. A manufacturer of locknuts found unwanted differences in the torque values of its product. (Torqe is the work (force $\times$ distance) required to tighten the nut.) The manufacturer did an experiment to study two factors that might affect torque values. The first factor was the type of manufacturing process. The second factor was the medium onto which the locknut would be threaded (a bolt or a mandrel). The following table gives the experimental data.

	Manufacturing Process		
	A	B	C
bolt	20,16,17,18,15, 16,19,14,15,24	26,40,28,38,38, 30,26,38,45,38	25,40,30,17,16, 45,49,33,30,20
mandrel	24,18,17,17,15, 23,14,18,12,11	32,22,30,35,32, 28,27,28,30,30	10,13,17,16,15, 14,11,14,15,16

(a) Give the ANOVA table for the two-way ANOVA model with interactions.

(b) Is there an interaction between these two factors for torque value?

(c) Regardless of your answer to (b), give the ANOVA table for the two-way ANOVA model without interactions.

(d) Descriptively, does manufacturing process or medium appear to be the more important factor? (*Hint:* consider the mean square.)

2. In class we did an example about orthogonal contrasts in one-way ANOVA. The same ideas can be used with continuous variables. Suppose we model $E[Y]$ as a function of x using data from a planned experiment where x takes on exactly 3 values. The data matrix is

x	y
40	25.66
50	29.15
60	35.73
40	28.00
50	35.09
60	39.56
40	20.65
50	29.79
60	35.66

Consider the following design matrix for a linear model with parameters β_0, L and Q. (The rows correspond to the same order of the data as in the data table above.)

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \end{bmatrix}$$

(a) Interpret the parameters β_0, L and Q.

(b) A common practice is to use the same design matrix but divide the second column by $\sqrt{2}$ and divide the third column by $\sqrt{6}$. What is the purpose of doing this?

(c) Find the least-squares estimates of β_0, L and Q using the re-scaled design matrix described in (b).

(d) Here is a different model: $E[y] = \beta_0 + \beta_1 x + \beta_2 x^2$. Consider the relationship between the two models. What is an advantage of using the first model compared to using $E[y] = \beta_0 + \beta_1 x + \beta_2 x^2$?

3.. Suppose an experiment is done to compare two varieties of tomatoes in plots that can hold two plants each. There are two such plots available for the study, so that the experimental design is

1	2
1	2

(plants will be randomized within a plot).

The statistical model is $y_{ij} = \mu + P_i + V_j + \epsilon_{ij}$ for $i = 1, 2$ plots and $j = 1, 2$ varieties. Assume the errors ϵ are uncorrelated and have variance σ^2.

For (a) and (c) use the data vector $\mathbf{Y}$ written in the order $\begin{matrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{matrix}$

(a) What is $\text{cov}(\mathbf{Y})$ when the "plot effects" P_i are considered fixed effects?

(b) What is the formula for the best linear unbiased estimator of $\hat{V}_1 - \hat{V}_2$ in terms of the Y_{ij} when the P_i are considered fixed effects?

(c) What is $\text{cov}(\mathbf{Y})$ when the "plot effects" P_i are considered random effects with variance τ^2?

(d) What is the formula for the best linear unbiased estimator of $\hat{V}_1 - \hat{V}_2$ in terms of the Y_{ij} when the P_i are considered random effects with variance τ^2? Treat σ^2 and τ^2 as known.

4. Suppose an experiment is done to compare three varieties of tomatoes in plots that can hold two plants each. There are three such plots available for the study, and the experimental design is

1	2
2	3
1	3

(plants will be randomized within a plot).

The statistical model is $y_{ij} = \mu + P_i + V_j + \epsilon_{ij}$ for $i = 1, 2, 3$ plots and $j = 1, 2, 3$ varieties. Assume the errors ϵ are uncorrelated and have variance σ^2.

For (a) and (c) use the data vector $\mathbf{Y}$ written in the order $\begin{matrix} y_{11} \\ y_{12} \\ y_{22} \\ y_{23} \\ y_{31} \\ y_{33} \end{matrix}$

(a) What is $\text{cov}(\mathbf{Y})$ when the "plot effects" P_i are considered fixed effects?

(b) What is the formula for the best linear unbiased estimator of $\hat{V}_1 - \hat{V}_2$ in terms of the Y_{ij} when the P_i are considered fixed effects?

(c) What is $\text{cov}(\mathbf{Y})$ when the "plot effects" P_i are considered random effects with variance τ^2?

(d) What is the formula for the best linear unbiased estimator of $\hat{V}_1 - \hat{V}_2$ in terms of the Y_{ij} when the P_i are considered random effects with variance τ^2? Treat σ^2 and τ^2 as known.

5. This is an exercise. The problem won't be graded, but if you hand in evidence of a "good faith" effort, it will be worth a few points of extra credit. A solution will be included in the homework key for you to check your work. The problem concerns the context in which regressors are considered random.

Consider the model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n;$$

where

$$\begin{pmatrix} x_i \\ \varepsilon_i \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_x \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix} \right).$$

We have

$$\begin{pmatrix} Y_i \\ x_i \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \sigma_{xY} \\ \sigma_{xY} & \sigma_x^2 \end{pmatrix} \right),$$

where $\sigma_y^2 = \beta_1^2 \sigma_x^2 + \sigma_\varepsilon^2$, $\mu_y = \beta_0 + \beta_1 \mu_x$, and $\sigma_{xY} = \beta_1 \sigma_x^2$.

(a) Derive $E[Y_i|x_i]$ and $\text{var}[Y_i|x_i]$.

Now suppose one does not observe $x_i, i = 1, 2, \ldots, n$; but observes $w_i = x_i + u_i$, where

$$\begin{pmatrix} x_i \\ \varepsilon_i \\ u_i \end{pmatrix} \sim N_3\left(\begin{pmatrix} \mu_x \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_\varepsilon^2 & 0 \\ 0 & 0 & \sigma_\varepsilon^2 \end{pmatrix} \right).$$

Assume that Y is conditionally independent of w: $E[Y_i|x_i, w_i] = E[Y_i|x_i]$. Suppose the true model is $E[Y_i \mid x_i] = \beta_0 + \beta_1 x_i$ but the observed data are $(Y_i, w_i), i = 1, 2, \ldots, n$.

(b) Relate $E[Y_i \mid w_i]$ to $E[x_i \mid w_i]$.

(c) What is the joint distribution of x_i and w_i and what is $E[x_i \mid w_i]$?

(d) Combine your answers to (b) and (c) to show that $E[Y_i|w_i] = \beta_0^* + \beta_1^* w_i$.

(e) What is the relationship between β_0^*, β_1^* and β_0, β_1?