STAT 8260 — Theory of Linear Models
Homework 6 – Due Thursday, April 24

Homework Guidelines:

- Homework is due by 4:30pm on the due date specified above. You may turn it in at the beginning of class or place it in my mailbox in the Statistics Building. **No late homeworks will be accepted without permission granted prior to the due date.**

- Use only standard (8.5 × 11 inch) paper and use only one side of each sheet.

- Homework should show enough detail so that the reader can clearly understand the procedures of the solutions. This is **absolutely essential** for you to receive full credit for your answer since the answers to most of the problems in Rencher appear in the back of the book.

- Problems should appear in the order that they were assigned.

- Some problem numbers differ between the first and second editions of the book. When a problem number is given in parentheses, owners of the second edition should do the problem in parentheses, first edition user's should do the problem not in parentheses. E.g., if I assign problems 1,4, 7(8), 9(11), then 1st edition users should do problems 1,4,7,9, and 2nd edition users should do problems 1,4,8,11.

- You may use a statistical software package or (better) a matrix manipulation language like Matlab, R, S-PLUS or SAS PROC IML) to help you do some of the problems, but please include your code and output with your answers.

Assignment:

Read chapter 11 from our text (or chapter 12 if you have the second edition). In addition, do the following problems from that chapter: 5, 6, 19 (but in this problem use $\boldsymbol{\gamma} = \begin{pmatrix} \tau_2 - \tau_1 \\ \mu + \tau_2 \end{pmatrix}$ instead of the value of $\boldsymbol{\gamma}$ given in the statement of the problem), and 25. In addition, do the following six problems.

1. Prove the theorem on the bottom of p.172 of our class notes.

2. In the example on pp.179–182 of our notes,

   a. verify that $\mathbf{Z}\boldsymbol{\gamma} = \mathbf{X}\boldsymbol{\beta}$ and that $\mathbf{Z}\mathbf{u} = \mathbf{X}$; and

   b. show that the marginal contrast between the first level of B and the average of the second and third levels of B given by

$$\begin{aligned}\boldsymbol{\lambda}^T\boldsymbol{\beta} &= \{\beta_1 + \frac{1}{2}[(\alpha\beta)_{11} + (\alpha\beta)_{21}]\} - \frac{1}{2}\{\beta_2 + \frac{1}{2}[(\alpha\beta)_{12} + (\alpha\beta)_{22}]\} \\ &- \frac{1}{2}\{\beta_3 + \frac{1}{2}[(\alpha\beta)_{13} + (\alpha\beta)_{23}]\}\end{aligned}$$

can be written as $\mathbf{b}^T\boldsymbol{\gamma}$ for some $\mathbf{b}$.

3. Consider the classical linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad \mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n).$$

Let $\mathbf{G}$ be any nonsingular generalized inverse of $\mathbf{X}^T\mathbf{X}$. Show that

$$\mathbf{G}^{-1}\boldsymbol{\beta} - \mathbf{X}^T\mathbf{X}\boldsymbol{\beta}$$

is not estimable. (Note that a vector-valued function of $\boldsymbol{\beta}$ is estimable if and only if each of its elements are estimable functions of $\boldsymbol{\beta}$.)

4. An incomplete block design with a treatments and b blocks is similar to a randomized complete block design except that the block size k, is too small to allow the complete set of a treatments to be observed in each block. Instead, $k < a$ treatments are observed in each block. A balanced incomplete block design (BIBD) is an incomplete block design with the property that every possible pair of treatments occurs together (that is, together in the same block) the same number of times. A nice feature of a BIBD is that all pairwise differences among the treatment effects are estimated with equal precision.

Consider the following BIBD:

Table 1

Block					
1		2		3	
A	y_{11}	A	y_{12}	B	y_{23}
B	y_{21}	C	y_{32}	C	y_{33}

Here, y_{ij} denotes a response on the i^{th} treatment in the j^{th} block, and A, B, C denote the $a = 3$ treatments in the experiment. Note that the data are $\{y_{ij}\}$, $i = 1, 2, 3$, $j = 1, 2, 3$, but only $n = 6$ observations are present, so not all combinations of i and j are observed. Notice that the block size here is $k = 2$ which is too small to accommodate all $a = 3$ treatments. However, it is a *balanced* incomplete block design because all pairs $((A, B), (A, C), (B, C))$ occur the same number of times (once).

The usual model for this design is

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}, \quad \{e_{ij}\} \overset{iid}{\sim} N(0, \sigma^2) \qquad (*)$$

a. Write this model in matrix notation for the design presented above in Table 1.

b. Show that the three pairwise contrasts in the treatment effects, $\psi_1 = \alpha_1 - \alpha_2$, $\psi_2 = \alpha_1 - \alpha_3$, and $\psi_3 = \alpha_2 - \alpha_3$ are estimable.

c. Suppose we add the constraints $\mu = 0$ and $\beta_3 = 0$ to model (*). Is this a reparameterization, or a real restriction that changes the model? Justify your answer.

d. Show that the best linear unbiased estimators of ψ_1, ψ_2, and ψ_3 from part (b) all have equal variance. (*Hint:* Consider the constrained version of (*) with constraints as given in part (c).)

5. Consider the model

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, 2, \quad j = 1, 2,$$

where $e_{11}, \ldots, e_{22}$ are i.i.d. with mean 0, and constant variance σ^2. Let $\boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2)^T$ and let $\mathbf{X}$ be the model matrix here.

a. For $\boldsymbol{\lambda}^T \boldsymbol{\beta} = \mu + \alpha_1$, show that

$$\mathbf{r} = c \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 1/2 \\ 0 \end{pmatrix},$$

where c is an arbitrary constant, represents all solutions to $\mathbf{X}^T\mathbf{X}\mathbf{r} = \boldsymbol{\lambda}$.

b. Obtain the BLUE for $\mu + \alpha_1$ using $\mathbf{r}$ from part (a).

c. Prove that if $\boldsymbol{\lambda}_1^T\boldsymbol{\beta}$ and $\boldsymbol{\lambda}_2^T\boldsymbol{\beta}$ are both estimable, then so is $\boldsymbol{\lambda}_1^T\boldsymbol{\beta} + \boldsymbol{\lambda}_2^T\boldsymbol{\beta}$. Use this result and the result of part (b) to obtain the BLUE of $\alpha_1 - \alpha_2$.

d. In this model, is the hypothesis $H_0 : \mu + \alpha_1 = \mu + \alpha_2 = \alpha_1 - \alpha_2 = 0$ a testable hypothesis? Why, or why not?

6. Consider the following data from an unbalanced two-way layout:

High Protein			Low Protein		
Beef	Cereal	Pork	Beef	Cereal	Pork
73	98	94	90	107	49
102	74	79	76	95	82
	56		90		

Suppose we want to use the two factor model with interaction to analyze these data. This model is given by

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, \quad \begin{array}{l} i = 1, \ldots, a \\ j = 1, \ldots, b \\ k = 1, \ldots, n_{ij} \end{array} \qquad (**)$$

Here $a = 2$ is the number of levels of factor A, protein level; $b = 3$ is the number of levels of factor B, food type, and n_{ij} is the sample size in at the i^{th} level of A combined with the j^{th} level of B (e.g., $n_{11} = 2$, $n_{12} = 3$, etc.). y_{ijk} represents the response for the k^{th} rat receiving the i^{th} level of A and j^{th} level of B. α_i is an effect associated with the i^{th} level of A, β_j is an effect associated with the j^{th} level of B, and γ_{ij} is the effect of the i, j^{th} combination of A and B (γ_{ij} is an interaction effect). Note that model (**) is not a constrained model and it is overparameterized.

a. Write down model (**) in the matrix/vector form $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$. That is, specify $\mathbf{y}$, $\mathbf{X}$, $\boldsymbol{\beta}$ and $\mathbf{e}$.

b. Suppose we wanted to add side conditions to model (**) to make it a full rank model. How many side conditions are necessary? Identify a complete set of side conditions for this model and write them in the form $\mathbf{T}\boldsymbol{\beta} = \mathbf{0}$.

c. Substitute the constraints given by your set of side conditions into the model equation and simplify so that the constrained model can be written as $\mathbf{y} = \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}$ where $\tilde{\mathbf{X}}$ is full rank (I want you to write down the new model matrix $\tilde{\mathbf{X}}$ and parameter vector $\tilde{\boldsymbol{\beta}}$).

d. Find the least-squares estimator $\hat{\tilde{\boldsymbol{\beta}}}$ based on your reparameterized model from part (c). In addition, compute the constrained least-squares estimator $\hat{\boldsymbol{\beta}}$ using the formula $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X} + \mathbf{T}^T\mathbf{T})^{-1}\mathbf{X}^T\mathbf{y}$. Do your estimators $\hat{\tilde{\boldsymbol{\beta}}}$ and $\hat{\boldsymbol{\beta}}$ agree?

e. For the specific model matrix $\mathbf{X}$ and constraint matrix $\mathbf{A}$ of this problem, verify that $(\mathbf{X}^T\mathbf{X} + \mathbf{T}^T\mathbf{T})^{-1}$ is a generalized inverse of $\mathbf{X}^T\mathbf{X}$ by direct computation of the defining property of generalized inverses.