STAT 8260 Final Exam – Thursday, May 1, 2008
SHOW ALL WORK

Name:____

1. Consider a randomized complete block design in which two blocks are used with four treatments per block. The four treatments are combinations of two factors: factor A (with levels A1 and A2), and factor B (with levels B1 and B2). Let y_{ijk} be the response in the ith block, at the jth level of A combined with the kth level of B. The design and data can be represented schematically as follows:

   Block 1

	B1	B2
A1	y_{111}	y_{112}
A2	y_{121}	y_{122}

   Block 2

	B1	B2
A1	y_{211}	y_{212}
A2	y_{221}	y_{222}

   The classical model for such a design would be $y_{ijk} = \mu+\tau_i+\alpha_j+\beta_k+(\alpha\beta)_{jk}+e_{ijk}$. However, we will consider a model in which we assume that factors A and B do not interact. That is, consider the model

   $$y_{ijk} = \mu + \tau_i + \alpha_j + \beta_k + e_{ijk}, \quad i = 1, 2; j = 1, 2; k = 1, 2;$$

   with the usual assumption of independent errors where $\mathrm{E}(e_{ijk}) = 0$, $\mathrm{var}(e_{ijk}) = \sigma^2$ for all i, j, k. Here τ_i, α_j and β_k are effects for blocks, levels of factor A, and levels of factor B, respectively.

   a. **(7 pts)** Write down the model matrix $\mathbf{X}$ and parameter vector $\boldsymbol{\beta}$ for this situation. What is rank($\mathbf{X}$) here?

b. (**6 pts**) Write down an appropriate set of side conditions for reparameterizing this model as a full rank model.

c. (**6 pts**) Write down the reparameterized model matrix and parameter vector based on your side conditions from part (b).

d. **(9 pts)** Find the BLUE of $\alpha_1 - \alpha_2$, which is an estimable quantity in this model. Your estimator should be written as a simple, non-vector/matrix function of the y_{ijk}s.

e. **(9 pts)** Obtain a simple formula in terms of the y_{ijk}s for the Type I sum of squares for block effects in this model (you may assume that terms are entered into the model in the order in which they appear in the model equation above).

f. **(6 pts)** Consider now the classical model $y_{ijk} = \mu + \tau_i + \alpha_j + \beta_k + (\alpha\beta)_{jk} + e_{ijk}$ for this design. For this model, how would you represent the Type II sum of squares for block effects in terms of the $SS(\cdot|\cdot)$ notation?

2. Suppose we collect data on fifty 10 to 13-year-old children, half of whom are male, and the rest female. Let y_{ij} be the height of the jth child of the ith gender, and let x_{ij} be the corresponding age of that child (in months). Consider the following model for these data:

$$y_{ij} = \alpha_i + (\mu + \beta_i)x_{ij} + e_{ij} \quad i = 1, 2; j = 1, \ldots, 25, \qquad (*)$$

where $e_{11}, e_{12}, \ldots, e_{2,25}$ are independent, each with mean 0 and variance σ^2.

a. **(7 pts)** Write this model down in $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ form, specifying all matrices, vectors and assumptions that define the model for these data. (Use $\vdots$'s in your matrices and vectors to deal with the large number of rows resulting from the large sample size here.)

b. **(7 pts)** Is the hypothesis $H_0 : \alpha_1 - \alpha_2 = \mu = 0$ testable? Why or why not?

c. **(6 pts)** Based on model (*), suppose we want to test whether a simple linear regression of y on x holds for these data with common intercept and slope for the two genders. Express such a hypothesis as a testable general linear hypothesis on the parameters of model (*).

d. **(6 pts)** Show that α_1 is an estimable parameter in model (*). (It really is; I promise.)

3. **(12 pts)** Consider the analysis of variance model for the balanced one-way layout: $y_{ij} = \mu + \alpha_i + e_{ij}$, where $i = 1, \ldots, a$, $j = 1, \ldots, n$, and $\mathrm{E}(e_{ij}) = 0$, $\mathrm{var}(e_{ij}) = \sigma^2$ for all i, j. Show that $\sum_{i=1}^{a} c_i \alpha_i$ is estimable **if and only if** $\sum_{i=1}^{a} \alpha_i = 0$.

4. **(9 pts)** Consider the model $y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$, $i = 1, 2$, $j = 1, 2$, where the e_{ij} are independent with spherical variance-covariance structure. Suppose that I wish to use $\alpha_1 + \alpha_2 = \beta_1 - \beta_2 = 0$ as a set of side conditions for this model. Is this a valid and complete set of side conditions here? Justify your answer.

5. **(10 pts)** Consider the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ where $\mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$ and $\mathbf{X}$ is $n \times p$ with $\text{rank}(\mathbf{X}) = k < p$. Suppose that $\hat{\boldsymbol{\beta}}_1 = \mathbf{G}_1\mathbf{X}^T\mathbf{y}$ and $\hat{\boldsymbol{\beta}}_2 = \mathbf{G}_2\mathbf{X}^T\mathbf{y}$ are two different least squares estimators of $\boldsymbol{\beta}$ corresponding to $\mathbf{G}_1$ and $\mathbf{G}_2$, respectively, where $\mathbf{G}_1$ and $\mathbf{G}_2$ are different, **non-symmetric** generalized inverses of $\mathbf{X}^T\mathbf{X}$. Suppose we are interested in two distinct estimable functions of $\boldsymbol{\beta}$: $\boldsymbol{\lambda}_1^T\boldsymbol{\beta}$ and $\boldsymbol{\lambda}_2^T\boldsymbol{\beta}$ where $\boldsymbol{\lambda}_1^T\boldsymbol{\beta} \neq \boldsymbol{\lambda}_2^T\boldsymbol{\beta}$. Show that $\text{cov}(\boldsymbol{\lambda}_1^T\hat{\boldsymbol{\beta}}_1, \boldsymbol{\lambda}_2^T\hat{\boldsymbol{\beta}}_2) = \sigma^2\boldsymbol{\lambda}_1^T\mathbf{G}_2\boldsymbol{\lambda}_2$.