STAT 8004 Midterm Exam I February 26, 2015

Name: ____________________________

1. Suppose that we have observable random variables y_1, y_2, y_3 and y_4 satisfying $E(y_1) = 2\beta_1 - \beta_2 + \beta_3 - \beta_4$, $E(y_2) = 2\beta_1 + \beta_3$, and $E(y_3) = \beta_2$, $E(y_4) = 2\beta_1 + \beta_2 + \beta_3$. Let $\mathbf{Y} = (y_1, y_2, y_3, y_4)^T$, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)^T$. Answer part (a)–(e) in this scenario.

   (a) Find $\mathbf{X}$ and $\boldsymbol{\varepsilon}$ such that a model for $\mathbf{Y}$ can be expressed in the form of $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

   (b) Is $\mathbf{X}$ in your model in (a) of full rank? Why or why not?

   (c) Clearly and precisely state minimal condition(s) under which your model in part (a) is a Gauss-Markov model.

(Question 1, continued...)

(d) Is β_4 estimable? If yes, find a linear unbiased estimator for β_4. If no, why?

(e) Let $\theta_1 = \beta_1$, $\theta_2 = 2\beta_2 - \beta_3$, $\theta_3 = \beta_2 + 2\beta_3$, $\theta_4 = \beta_4$, and $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)^T$. Find $\mathbf{Z}$ such that your model in (a) for $\mathbf{Y}$ can be written in the form of $\mathbf{Y} = \mathbf{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$.

2. Consider an experiment with two factors: A (with levels A_1 and A_2) and B (with levels B_1 and B_2). Let y_{ijk} be the outcome of the kth unit at the level of A_i factor and B_j factor $(i, j = 1, 2)$. Data are collected as in the following table:

Factor A	Factor B	Outcome
A_1	B_1	20
A_1	B_1	25
A_1	B_2	30
A_1	B_2	35
A_2	B_1	55
A_2	B_2	40
A_2	B_2	30

Consider a full rank Gauss-Markov model for the outcome data that takes the form

$$\mathbf{Y} = \begin{pmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{221} \\ y_{222} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

In this question, assume that $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Answer part (a)–(h) in this scenario.

(a) Is this data set a balanced one?

(b) Express the mean outcomes $\mu_{ij} = E(y_{ijk})$ $(i, j = 1, 2)$ corresponding to all possible combinations of the factors A and B as functions of $\beta_1, \beta_2, \beta_3, \beta_4$.

(c) Express the overall mean of the outcomes as a function of $\beta_1, \beta_2, \beta_3, \beta_4$.

(Question 2, continued ...)

Let $\mathbf{P}$ be the projection matrix associated with $\mathbf{X}$, and $\mathbf{I}$ be the identity matrix. Now you are given the following:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} = \begin{pmatrix} 36.25 \\ -8.75 \\ 2.5 \\ -7.5 \end{pmatrix}, (\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{32} \cdot \begin{pmatrix} 5 & -1 & 1 & -1 \\ -1 & 5 & -1 & 1 \\ 1 & -1 & 5 & -1 \\ -1 & 1 & -1 & 5 \end{pmatrix} \text{ and } \mathbf{Y}^T(\mathbf{I} - \mathbf{P})\mathbf{Y} = 75$$

where $\hat{\beta}_i$ $(i = 1, \ldots, 4)$ are the ordinary least squares estimates for the parameters. Please best simplifying your expression in answering the following parts. You **<u>don't</u>** have to calculate the exact numbers. Most importantly you need to clearly define any notation used in your expression and exactly specify which distribution and what quantile you are using.

(d) Find a 95% level confidence interval for $\mu_{21} - \mu_{11}$.

(e) Find the F statistic for $H_0 : \mu_{12} = \mu_{22} = 35$ vs H_a : not H_0, and give its degrees of freedom.

d.f.=____________________

(Question 2, continued ...)

(f) Suppose three new outcomes are observed at the condition with respectively level A_1 and B_2, find a 95% prediction interval for the average of the three new observations.

(g) Suppose that the fifth row (outcome=55) in the data table is now removed. Does it change the estimability of any of the parameters β_j ($j = 1, \dots, 4$) in the model? Why or why not?

(h) Does the change in the previous part (g) have any impact on the least square estimation of μ_{11} and μ_{12}? Explain your answer.

3. Consider the model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$ where ϵ_i independently follow standard normal distribution for $i = 1, \ldots, n$. Let x_0 be the value at which the function $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2$ is maximized (or minimized). Answer parts (a) and (b) in this scenario.

   (a) Find the maximum likelihood estimator for x_0.

   (b) Find a $(1 - \alpha)$ level confidence interval for x_0. You may assume n large here. If you solve this part without assuming n large, there will be 3 extra points.