Preliminary Examination: 
LINEAR MODELS

Answer all questions and show all work. 
Q1 is 35 points; Q2 is 30 points, and Q3 is 35 points.

1. Assume each \( Y_i \) \((i = 1, \ldots, n)\) can be modeled by the following linear regression model:

\[
Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \epsilon_i,
\]

where \(\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \sim N(0, \Sigma) \) with \( \Sigma = \sigma^2 V \), \(\sigma^2 > 0\), and

\[
V = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} = (1 - \rho)I + \rho J;
\]

here, \(I\) is an \(n \times n\) identity matrix; \(J\) is an \(n \times n\) matrix whose elements are all 1s.

The ‘centered form’ of the model can be written as

\[
Y_i = \beta_0 + \beta_1 (X_{i1} - \bar{X}_1) + \cdots + \beta_p (X_{ip} - \bar{X}_p) + \epsilon_i,
\]

where \(\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ij}; j = 1, \ldots, p.\)

Define \(\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'; \beta_1 = (\beta_1, \ldots, \beta_n)'; \mathbf{j}\) is an \(n\)-dimensional vector of 1s;

\(\alpha = \beta_0 + \beta_1 \bar{X}_1 + \cdots + \beta_p \bar{X}_p; \mathbf{X}_c = (I - \frac{1}{n} J) \bar{X}\) with

\[
\mathbf{X} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix}.
\]

a. Show that the following is equivalent to the ‘centered’ form of the model:

\[
\mathbf{Y} = (\mathbf{j}, \mathbf{X}_c) \begin{pmatrix} \alpha \\ \beta_1 \end{pmatrix} + \mathbf{\epsilon}.
\]
b. Let \( X = (j, X_c) \) and \( \beta = \left( \frac{\alpha}{\beta_1} \right) \). Derive the generalized least squares (GLS) estimator for \( \beta \) in terms of \( X, Y, \) and \( V \).

c. Show that 
\[
X'V^{-1}X = \begin{pmatrix}
bn & 0' \\
0 & aX_c'X_c
\end{pmatrix}
\]

where \( a = 1/(1 - \rho) \) and \( b = 1/[1 + (n - 1)\rho] \). (Hint: \( V^{-1} = a(I - b\rho J) \).)

d. Show that 
\[
X'V^{-1}Y = \begin{pmatrix}
bnY' \\
aX_c'Y
\end{pmatrix}
\]

e. Show that the GLS for \( \beta \) is given by 
\[
\hat{\beta} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \bar{Y} \\ (X_c'X_c)^{-1}X_c'Y \end{pmatrix},
\]

where \( \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i \).

2. Consider the cell means ANOVA model 
\[
Y_{ij} = \mu_i + \epsilon_{ij},
\]

for \( i = 1, 2, 3 \) and \( j = 1, 2, \ldots, n \), where \( \epsilon_{ij} \) are iid \( N(0, \sigma^2) \). The restriction 
\[
\mu_3 = \mu_1 - \mu_2
\]
is placed on the parameters. Define \( \beta = (\mu_1, \mu_2, \mu_3)' \).

a. Write this as a general linear model \( Y = X\beta + \epsilon \), and express the restriction in the form of \( A'\beta = \delta \).

b. Find the restricted least squares estimator, \( \hat{\beta}_R \). Express this estimator in terms of the treatment means, \( \bar{Y}_i \) for \( i = 1, 2, 3 \).

c. Define 
\[
Q(\beta) = (Y - X\beta)'(Y - X\beta)
\]

and let \( \hat{\beta} \) denote the unrestricted least squares estimators. How do \( Q(\hat{\beta}) \) and \( Q(\hat{\beta}_R) \) compare and why?

d. Find \( E[Q(\hat{\beta})] \) and \( var[Q(\hat{\beta})] \) (under the model without the restriction).

e. Consider testing \( H_0 : \mu_3 = \mu_1 - \mu_2 \). Give the \( F \) test statistic and its distribution when \( H_0 \) is true, and explain how this distribution will change under the alternative hypothesis \( H_a : \mu_3 - (\mu_1 - \mu_2) = \delta \neq 0 \).
3. Consider the general linear model $Y = X\beta + \varepsilon$, where $X$ is $n \times p$ with rank $r \leq p$, $\beta$ is $p \times 1$, and $\varepsilon \sim \mathcal{N}_n(0, V)$, where $V$ is known and nonsingular. Let $\hat{\beta} = (X'X)^{-1}X'Y$ denote an ordinary least squares estimator, and $(X'X)^{-1}$ denotes a generalized inverse of $X'X$. Define:

$$\hat{\sigma}^2 = (n - r)^{-1}Y'(I - P_X)Y,$$

where $P_X$ is the projection matrix onto the column space of $X$, $C(X)$. Suppose that $\lambda$ is a $p$-dimensional vector and $\lambda'\beta$ is estimable.

a. Suppose that $V = \sigma^2I$. Derive the sampling distribution of $\lambda'\hat{\beta}$.

b. Suppose that $VX = XQ$ for some matrix $Q$. Prove that $\lambda'\hat{\beta}$ and $(I - P_X)Y$ are independent.

c. Suppose that $V = \sigma^2(I + P_X)$, for some $\sigma^2 > 0$. Define

$$T = \frac{\lambda'\hat{\beta} - \lambda'\beta}{\sqrt{\hat{\sigma}^2\lambda'(X'X)^{-1}\lambda}}.$$

Find the constant $k$ such that $kT$ follows a $t$ distribution. What is its degrees of freedom? Is this a central $t$ distribution?

d. As in part (c), suppose that $V = \sigma^2(I + P_X)$. Use the results in part (c) to derive a $100(1 - \alpha)%$ confidence interval for $\lambda'\beta$.

e. As in part (c), suppose that $V = \sigma^2(I + P_X)$. Build a $100(1 - \alpha)%$ confidence interval for $\lambda'\beta$, based on the generalized least squares estimator and compare the two intervals in (d) and (e).