Preliminary Examination:
LINEAR MODELS

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

1. Consider the model

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij};\ \ i = 1, 2;\ \ j = 1, 2, 3$$

with $E(\epsilon_{ij}) = 0$ and $\text{var}(\epsilon_{ij}) = \sigma^2\mathbf{I}$.

a. This model can be written in matrix/vector form as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon$, where $\boldsymbol{\beta} = (\mu, \tau_1, \tau_2)'$. Define $\mathbf{Y}$, $\mathbf{X}$ and $\boldsymbol{\varepsilon}$.

b. For the $\mathbf{X}$ in your answer to part (a), can we find the inverse of $\mathbf{X}'\mathbf{X}$? Why or why not? Explain.

c. Find an ordinary least square (OLS) estimator $\hat{\boldsymbol{\beta}}$. Note: You need to represent the solution explicitly as a function of Y_{ij}'s. You need to show the derivation. Claiming or stating results from courses won't be granted credits, but you can use the formula of the normal equation directly without deriving it.

d. For $\hat{\boldsymbol{\beta}}$ you've obtained in part (c), find $E(\hat{\boldsymbol{\beta}})$.

2. In many applications, there can be multiple response variables, in which case the linear regression model can be written as

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E},$$

where $\mathbf{Y} = (Y_{ij})_{1\le i\le n, 1\le j\le d}$ is a matrix in $\mathbb{R}^{n\times d}$, $\mathbf{X} = (x_{ij})_{1\le i\le n, 1\le j\le p}$ is a matrix in $\mathbb{R}^{n\times p}$, $\mathbf{B} = (\beta_{ij})_{1\le i\le p, 1\le j\le d}$ is a matrix in $\mathbb{R}^{p\times d}$, and $\mathbf{E} = (\epsilon_{ij})_{1\le i\le n, 1\le j\le d}$ is a matrix in $\mathbb{R}^{n\times d}$.

Assume that the errors ϵ_{ij}, $1 \le i \le n$, $1 \le j \le d$, are independent normal random variables with mean zero and variance σ^2. In this case, the least squares estimator is defined as

$$\hat{\mathbf{B}} = \underset{\mathbf{B}\in\mathbb{R}^{p\times d}}{\text{argmin}} \|\mathbf{Y} - \mathbf{X}\mathbf{B}\|^2,$$

where for a matrix $\mathbf{A} = (a_{ij})$, the norm is defined as $\|\mathbf{A}\| = (\sum_i \sum_j a_{ij}^2)^{1/2}$. Suppose you observe only the response matrix $\mathbf{Y}$ and the design matrix $\mathbf{X}$.

a. Derive the least square estimator $\hat{\mathbf{B}}$ and $\hat{\sigma}^2$.

b. Construct a test for the null hypothesis that the first column of $\mathbf{X}$ is an irrelevant variable. You need to derive the test statistic, its distribution under the null hypothesis and explain the decision rule (i.e., when to reject).

3. Consider a simple linear regression model,

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n,$$

where $\{\epsilon_i\}$ are iid $N(0, \sigma^2)$ random variables.

Let $\mathbf{Y}' = (Y_1, Y_2, \ldots, Y_n)$, $\boldsymbol{\beta}' = (\beta_0, \beta_1)$, and $\mathbf{X}' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix}$. Furthermore, for known constants $\{z_i : i = 1, \ldots, n\}$, define $\mathbf{Z}' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_n \end{bmatrix}$, and assume that $\mathbf{Z}'\mathbf{X}$ is a non-singular matrix.

a. Show that the so-called instrumental-variable estimator, $\tilde{\mathbf{b}} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Y}$, is an unbiased estimator of the vector $\boldsymbol{\beta}$.

b. Find the sampling distribution of $\tilde{\mathbf{b}} = (\tilde{b}_0, \tilde{b}_1)'$, including its mean vector and variance-covariance matrix. Furthermore, let $\hat{\mathbf{b}} = (\hat{b}_0, \hat{b}_1)'$ be the OLS estimator of $\boldsymbol{\beta}$. Prove that $var(\tilde{b}_1) \geq var(\hat{b}_1)$.

c. Let $\mathbf{P_Z}$ denote the projection matrix on the column space $\mathcal{C}(\mathbf{Z})$.

c-i. Show that $\mathbf{Y}'(\mathbf{I} - \mathbf{P_Z})\mathbf{Y}/\sigma^2$ has a chi-squared distribution. Give its degrees of freedom and the non-centrality parameter.

c-ii. What is the value of the non-centrality parameter when $\beta_1 = 0$? Justify your answer.

d. Let the residuals corresponding to the instrumental-variable estimator $\tilde{\mathbf{b}}$ be denoted by

$$\tilde{\mathbf{e}} = \mathbf{Y} - \mathbf{X}\tilde{\mathbf{b}}.$$

d-i. Derive the joint distribution of $\tilde{\mathbf{b}}$ and $\tilde{\mathbf{e}}$, and answer whether $\tilde{\mathbf{b}}$ and $\tilde{\mathbf{e}}$ are independently distributed.

d-ii. Now let $\mathbf{Q}$ denote the $n \times n$ matrix so that $\tilde{\mathbf{e}} = \mathbf{Q}\mathbf{Y}$. Show that the matrix $\mathbf{Q}$ is not symmetric. Is the matrix $\mathbf{Q}$ idempotent? Explain why or why not.

d-iii. Give a necessary and sufficient condition for the distribution of $\tilde{\mathbf{e}}'\tilde{\mathbf{e}}/\sigma^2$ to be a chi-squared distribution.