Preliminary Examination: LINEAR MODELS

Answer all questions and show all work.

1. Suppose that data $\{(x_{ij}, y_{ij}) : i = 1, \ldots, k, j = 1, \ldots, J\}$ can be modeled as having a common slope and possibly different intercepts using the linear model,

$$Y_{ij} = \beta_i + \gamma x_{ij} + \epsilon_{ij},$$

where $\{\epsilon_{ij}\}$ are independently and identically distributed $N(0, \sigma^2)$ random variables. Assume that no vector $(x_{i1}, \ldots, x_{iJ})$, for $i = 1, \ldots, k$, is proportional to the vector of 1s.

a. Determine the ordinary-least-squares estimator of $(\beta_1, \ldots, \beta_k, \gamma)'$.

b. Give an explicit expression for the size α likelihood-ratio test of the hypothesis,

$$H_0 : \beta_1 = \cdots = \beta_k = 0 \text{ versus } H_a : \text{ not } H_0$$

c. Compute the power of the test that you derived in part (b). (There are several ways of defining the non-centrality parameter for the test. Pick any one of these, and use it consistently in this part.) Show that the power is independent of γ.

2. Let $X_1, \ldots, X_n$ be iid random variables with a $N(\theta, \sigma^2)$ distribution conditional on θ and σ^2, and θ and σ^2 are unknown.

a. Give the classical test statistic for testing $\sigma^2 = \sigma_0^2$ against $\sigma^2 \neq \sigma_0^2$.

Suppose that the following assumptions are made concerning θ and σ^2: θ follows a $N(\theta_0, \tau^2)$ distribution with both θ_0^2 and τ^2 known, and $\sigma^2 = \sigma_0^2 > 0$.

b. The *(joint) marginal* distribution of $(X_1, \ldots, X_n)$ (i.e., unconditional on the parameters) is multivariate normal. Derive the mean and variance-covariance matrix for this distribution.

c. Find the distribution of the random variable

$$T(X_1, \ldots, X_n) = \frac{1}{\sigma_0^2} \sum_{i=1}^{n} (X_i - \bar{X})^2 + \frac{n(\bar{X} - \theta_0)^2}{n\tau^2 + \sigma_0^2}$$

under the marginal distribution of $(X_1, \ldots, X_n)$.

d. Let $(x_1, \ldots, x_n)$ denote the observed values of $(X_1, \ldots, X_n)$. The Bayesian prior predictive p-value to validate the specified model is defined as

$$p = P(T(X_1, \ldots, X_n) \geq T(x_1, \ldots, x_n))$$

under the marginal distribution of $(X_1, \ldots, X_n)$. Compare this Bayesian prior predictive p-value to the classical p-value for testing $\sigma^2 = \sigma_0^2$ against $\sigma^2 \neq \sigma_0^2$ in part (a).

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. Show that $\tilde{\mathbf{b}}$ and $\tilde{\mathbf{e}}$ are independently distributed and give the sampling distribution of $\tilde{\mathbf{e}}$.

d-ii. Now let $\mathbf{Q}$ denote the $n \times n$ matrix so that $\tilde{\mathbf{e}} = \mathbf{QY}$. 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.

4. Consider the 2-way random effects model with interactions:

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}, i = 1, 2, j = 1, 2, k = 1, \ldots, n_{ij},$$

with variance components σ_a^2, σ_b^2, σ_{ab}^2, and σ^2, respectively.

a. Assume $n_{11} = n_{12} = n_{21} = 2$, and $n_{22} = 3$. Find the Expected Mean Squares as functions of the variance components for the two factors.

b. Repeat part (a) with $n_{11} = n_{12} = n_{21} = n_{22} = 2$, Comment briefly on the results in parts (a) and (b).