Preliminary Examination: Linear Models

10:00 am - 12:30 pm, Monday, May 4, 2026

Answer questions by showing all of your work.

1. Consider the linear model

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i, \qquad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2), \quad i = 1, \ldots, n.$$

Now consider the centered model

$$y_i = \alpha + \beta_1 (x_{i1} - \bar{x}_1) + \cdots + \beta_k (x_{ik} - \bar{x}_k) + \epsilon_i,$$

where

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}, \qquad j = 1, \ldots, k.$$

In matrix form, this model can be written as

$$\boldsymbol{y} = \begin{bmatrix} \mathbf{1} & \boldsymbol{X}_c \end{bmatrix} \begin{pmatrix} \alpha \\ \boldsymbol{\beta}_c \end{pmatrix} + \boldsymbol{\varepsilon},$$

where

$$\boldsymbol{X}_c = \begin{pmatrix} x_{11} - \bar{x}_1 & \cdots & x_{1k} - \bar{x}_k \\ \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_1 & \cdots & x_{nk} - \bar{x}_k \end{pmatrix},$$

and

$$\boldsymbol{\beta}_c = (\beta_1, \ldots, \beta_k)^\top.$$

Assume that, for each $j = 1, \ldots, k$, the values $x_{1j}, \ldots, x_{nj}$ are not all identical, and that $\boldsymbol{X}_c$ has full column rank.

(a) Prove that the ordinary least squares estimators of α and $\boldsymbol{\beta}_c$ are

$$\hat{\alpha} = \bar{y} \qquad \text{and} \qquad \hat{\boldsymbol{\beta}}_c = (\boldsymbol{X}_c^\top \boldsymbol{X}_c)^{-1} \boldsymbol{X}_c^\top \boldsymbol{y}.$$

(b) Prove the following theorem. *Note: We proved similar results in class, but you may not simply cite them here. Your proof should include details comparable to those used in class.*

Let $\boldsymbol{y}$ be an n-dimensional random vector with

$$E(\boldsymbol{y}) = \boldsymbol{\mu}, \qquad \mathrm{Var}(\boldsymbol{y}) = \boldsymbol{\Sigma}.$$

If $\boldsymbol{A}$ is an $n \times n$ symmetric matrix, prove that

$$E(\boldsymbol{y}^T \boldsymbol{A} \boldsymbol{y}) = \boldsymbol{\mu}^T \boldsymbol{A} \boldsymbol{\mu} + tr(\boldsymbol{A}\boldsymbol{\Sigma}).$$

(c) Using the result in part (b), find an unbiased estimator of σ^2.

(d) Let $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ denote the ordinary least squares estimators from the original (non-centered) model. Find the distribution of

$$\frac{\text{SSR}}{\sigma^2},$$

where $\text{SSR} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ and $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}$. If you use any theorem, clearly state the theorem and verify all conditions required for applying it.

2. Consider the following linear model:

$$Y_i = \frac{i}{n}\beta + \epsilon_i, \qquad i = 1, 2, \ldots, n, \tag{1}$$

where the errors ϵ_i follow the time-series model

$$\epsilon_i = \rho\epsilon_{i-1} + \sqrt{1-\rho^2}\, e_i, \qquad i = 2, 3, \ldots, \tag{2}$$

with $\rho \in (-1, 1)$ and

$$\epsilon_1 = \sqrt{1-\rho^2}\, e_1.$$

Assume that $\{e_i\}$ are iid random variables with

$$E(e_i) = 0, \qquad E(e_i^2) = \sigma^2,$$

and that e_i is independent of ϵ_k for all $k < i$.

Thus, model (1) may be viewed as a linear time trend (without an intercept, for simplicity) plus time-series noise. In answering the following questions, you may use the algebraic facts below:

- $$\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}.$$
- $$\sum_{i=0}^n \rho^i = \frac{1-\rho^{n+1}}{1-\rho}, \qquad |\rho| \neq 1.$$
- $$\sum_{i=1}^\infty i\rho^i = \frac{\rho}{(1-\rho)^2}, \qquad |\rho| < 1.$$

- Let $\{a_i\}$ and $\{b_i\}$ be two sequences. Assume that $\{a_i\}$ is nonnegative and nondecreasing, and that

$$\lim_{n\to\infty} \frac{a_n^2}{\sum_{i=1}^n a_i^2} = 0.$$

Also assume that

$$\sum_{i=1}^{\infty} i|b_i| < \infty \qquad \text{and} \qquad \sum_{i=1}^{\infty} b_i \neq 0.$$

Then, as $n \to \infty$,

$$\sum_{1\le i<j\le n} a_i a_j b_{j-i} \approx \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{j=1}^{\infty} b_j\right).$$

(a) Find $\mathrm{Cov}(\epsilon_i, \epsilon_{i+k})$ for $k \ge 0$.

Hint: You may use mathematical induction by first considering simple cases such as $(i,k) = (1,0)$, $(2,0)$, $(1,1)$, and $(1,2)$ and then generalizing. Alternatively, you may first express ϵ_i in terms of $\{e_j\}$ and then derive the covariance using the independence of the variables $\{e_j\}$.

(b) Let $\hat\beta_1$ denote the ordinary least squares estimator of β based on the model in (1). Derive $\hat\beta_1$ and find

$$\lim_{n\to\infty} n\,\mathrm{Var}(\hat\beta_1).$$

(c) Since the equation (2) implies

$$\epsilon_i - \rho\epsilon_{i-1} = \sqrt{1-\rho^2}\, e_i,$$

we may define the transformed variables

$$Z_i = Y_i - \rho\epsilon_{i-1}$$

Suppose, hypothetically, that the variables $\{Z_i\}$ were observed. Let $\hat\beta_2$ denote the ordinary least squares estimator of β based on the model with the transformed data $\{Z_i\}$. Derive $\hat\beta_2$ and find

$$\lim_{n\to\infty} n\,\mathrm{Var}(\hat\beta_2).$$

Compare this limit with the one obtained in part (b).

(d) Another method for estimating β is maximum likelihood. Assume now that $\{e_i\}$ are iid *normally* distributed. Derive the likelihood function for the parameters β, σ^2, and ρ based on the observed data $Y_1, \ldots, Y_n$ under model (1).

Hint: To avoid inverting the covariance matrix, you may use the factorization

$$f(\epsilon_1,\ldots,\epsilon_n) = f(\epsilon_1)\, f(\epsilon_2 \mid \epsilon_1)\, f(\epsilon_3 \mid \epsilon_1,\epsilon_2) \cdots f(\epsilon_n \mid \epsilon_1,\ldots,\epsilon_{n-1}).$$