Applied Statistics Preliminary Examination
Theory of Linear Models
May 2021

Instructions:

- Do all 3 Problems. Neither calculators nor electronic devices of any kind are allowed. Show all your work, clearly stating any theorem or fact that you use. Each of the 14 parts carries an equal weight of 10 points, except that 3(b), 3(c), and 3(d) are worth 12 points each.
- Abbreviations/Acronyms.
  - IID (independent and identically distributed).
  - LSE (least squares estimator); BLUE (best linear unbiased estimator). Sometimes the LSE may be designated OLS (ordinary least squares) estimator, in order to differentiate it from the GLS (generalized least squares) estimator.
- Notation.
  - $\boldsymbol{x}^T$ or $\boldsymbol{A}^T$: indicates transpose of vector $\boldsymbol{x}$ or matrix $\boldsymbol{A}$.
  - $\mathrm{tr}(\boldsymbol{A})$ and $|\boldsymbol{A}|$: denotes the trace and determinant, respectively, of matrix $\boldsymbol{A}$.
  - $\boldsymbol{I}_n$: the $n \times n$ identity matrix.
  - $\boldsymbol{j}_n = (1, \ldots, 1)^T$ is an n-vector of ones, and $\boldsymbol{J}_{m,n}$ is an $m \times n$ matrix of ones.
  - $\mathbb{E}(X)$ and $\mathbb{V}(X)$: expectation and variance of random variable X.
  - $\boldsymbol{x} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: the m-dimensional random vector $\boldsymbol{x}$ has a normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.
  - $X \sim t(n, \lambda)$: a t distribution with n degrees of freedom and noncentrality parameter λ. If $\lambda = 0$ we write simply: $X \sim t(n)$.
  - $X \sim F(n_1, n_2, \lambda)$: an F distribution with n_1 and n_2 numerator and denominator degrees of freedom respectively, and noncentrality parameter λ. If $\lambda = 0$ we write simply: $X \sim F(n_1, n_2)$.
- Possibly useful results.
  - If $\boldsymbol{x} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given in partitioned form as

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

    with $m_1 = \dim(\boldsymbol{x}_1)$, then the conditional distribution of $\boldsymbol{x}_1$ given $\boldsymbol{x}_2$ is

$$\boldsymbol{x}_1 | \boldsymbol{x}_2 \sim N_{m_1}\left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2),\ \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right).$$

Problems:

1. Let the scalar y and the p-dimensional vector $\boldsymbol{x}$ be jointly multivariate normally distributed as follows:

$$\begin{pmatrix} y \\ \boldsymbol{x} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu_y \\ \boldsymbol{\mu}_x \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \boldsymbol{\sigma}_{xy}^T \\ \boldsymbol{\sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{pmatrix} \right).$$

   (a) Show that the conditional mean of y given $\boldsymbol{x}$ takes on the form of the linear regression:

$$\mathbb{E}(y|\boldsymbol{x}) = a + \boldsymbol{\beta}^T(\boldsymbol{x} - \boldsymbol{b}),$$

   and find the value of the scalar a, the vector $\boldsymbol{b}$, and the coefficient vector $\boldsymbol{\beta}$.

   (b) Define the *mean squared error* (MSE) due to regression as $\sigma_\epsilon^2 := \mathbb{E}(y - \mathbb{E}(y|\boldsymbol{x}))^2$, and find an expression for it in terms of the elements of the covariance matrix of $(y, \boldsymbol{x})^T$.

   (c) If the *theoretical* R^2 is defined as the proportion of the variance of y that is accounted for by $\mathbb{E}(y|\boldsymbol{x})$, i.e., $\rho^2_{y|\boldsymbol{x}} := (\sigma_y^2 - \sigma_\epsilon^2)/\sigma_y^2$, show that

$$\rho^2_{y|\boldsymbol{x}} = \frac{1}{\sigma_y^2} \boldsymbol{\sigma}_{xy}^T \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\sigma}_{xy}.$$

   (d) Define the transformed regression vector $\boldsymbol{z} := \boldsymbol{R}\boldsymbol{x}$, where $\boldsymbol{R}$ is a $p \times p$ full-rank orthonormal matrix, i.e., $\boldsymbol{R}\boldsymbol{R}^T = \boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{I}_p$. Find the (joint) distribution of $(y, \boldsymbol{z})^T$.

   (e) Show that $\mathbb{E}(y - \mathbb{E}(y|\boldsymbol{x}))^2 = \mathbb{E}(y - \mathbb{E}(y|\boldsymbol{z}))^2$, and hence deduce that the theoretical R^2 is unchanged by the transformation in (d). That is, show that $\rho^2_{y|\boldsymbol{z}} = \rho^2_{y|\boldsymbol{x}}$.

2. Consider the vector of $n = 5$ observations $\boldsymbol{y} = (y_1, \ldots, y_5)^T$ from the linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where the full-rank matrix $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$ consists of the columns vectors $\boldsymbol{x}_1 = (x_{1,1}, \ldots, x_{5,1})^T$ and $\boldsymbol{x}_2 = (x_{1,2}, \ldots, x_{5,2})^T$, $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$, and $\boldsymbol{\epsilon} \sim \mathrm{N}(0, \sigma^2 \boldsymbol{I}_5)$. Define the following summary statistics:

$$s_{11} = \sum_{i=1}^5 x_{i,1}^2, \quad s_{12} = \sum_{i=1}^5 x_{i,1}x_{i,2}, \quad s_{22} = \sum_{i=1}^5 x_{i,2}^2, \quad t_1 = \sum_{i=1}^5 x_{i,1}y_i, \quad t_2 = \sum_{i=1}^5 x_{i,2}y_i.$$

   (a) Derive an expression for $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \widehat{\beta}_2)^T$, the LSE of $\boldsymbol{\beta}$, in terms of the above summary statistics.

   (b) Find an expression for the covariance matrix $\mathrm{var}(\widehat{\boldsymbol{\beta}})$ of the LSE, and hence calculate the correlation coefficient between $\widehat{\beta}_1$ and $\widehat{\beta}_2$.

   (c) Give an expression for the unbiased estimate of σ^2, and compute the coefficient of determination (R^2) for the fitted model.

   (d) Construct a level α test to determine if $\boldsymbol{x}_2$ is needed in the model that already has $\boldsymbol{x}_1$. That is, is it sufficient to have the model matrix consist of the single column vector $\boldsymbol{X} = \boldsymbol{x}_1$?

3. Consider the two-factor linear model $y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$, where $i = 1, 2, 3$ and $j = 1, 2, 3$, the ϵ_{ij} are IID $N(0, \sigma^2)$, but not all combinations (i, j) of factors A and B are observed. There are 3 cases to consider, each defined by the matrices below, where an asterisk $(*)$ in the (i, j) entry indicates that y_{ij} was observed. Thus, for example, in Case 1 the following vector of responses is observed: $\boldsymbol{y} = (y_{11}, y_{12}, y_{21}, y_{22}, y_{23}, y_{32}, y_{33},)^T$.

Case 1

*	*	
*	*	*
	*	*

Case 2

*	*	
	*	*
		*

Case 3

*	*	
*	*	
		*

Our aim in this Problem is to make inference on all *estimable* contrasts of factor A and factor B main effects, i.e., all differences of parameters of the form $\alpha_i - \alpha_{i'}$ for $i \neq i'$, and $\beta_j - \beta_{j'}$ for $j \neq j'$. Recall that in the generic linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, the linear combination $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is defined to be estimable if there exists a constant vector $\boldsymbol{a}$ such that $\mathbb{E}(\boldsymbol{a}^T\boldsymbol{y}) = \boldsymbol{\lambda}^T\boldsymbol{\beta}$. Also, denote by k the rank of the model matrix, i.e., $k = \mathrm{rk}(\boldsymbol{X})$.

(a) In Case 1, provide an easy argument to show that $3 \leq k \leq 6$. Explicitly find the value of k.

(b) In Case 1, and by finding an appropriate vector $\boldsymbol{a}$ in each case, show that all $\alpha_i - \alpha_{i'}$ and $\beta_j - \beta_{j'}$ contrasts are estimable.

(c) In Case 2, determine which $\alpha_i - \alpha_{i'}$ and $\beta_j - \beta_{j'}$ contrasts are estimable. For each contrast that is estimable, find an appropriate vector $\boldsymbol{a}$.

(d) In Case 3, determine which $\alpha_i - \alpha_{i'}$ and $\beta_j - \beta_{j'}$ contrasts are estimable. For each contrast that is estimable, find an appropriate vector $\boldsymbol{a}$.

(e) In Case 1, determine if the hypothesis stated below is *testable*, carefully justifying your answer. If it is testable, show how to construct a test statistic for it, and state the distribution of the test statistic under H_0:

$$H_0 : \alpha_1 - \alpha_2 = 1 \quad \text{and} \quad \alpha_1 - \alpha_3 = 2 \quad \text{and} \quad \alpha_2 - \alpha_3 = 3.$$

Applied Statistics Preliminary Examination
Theory of Linear Models
August 2021

Instructions:

- Do all 3 Problems. Neither calculators nor electronic devices of any kind are allowed. Show all your work, clearly stating any theorem or fact that you use. Each of the 14 parts carries an equal weight of 10 points.
- Abbreviations/Acronyms.
  - IID (independent and identically distributed).
  - LSE (least squares estimator); BLUE (best linear unbiased estimator). Sometimes the LSE may be designated OLS (ordinary least squares) estimator, in order to differentiate it from the GLS (generalized least squares) estimator.
- Notation.
  - $\boldsymbol{x}^T$ or $\boldsymbol{A}^T$: indicates transpose of vector $\boldsymbol{x}$ or matrix $\boldsymbol{A}$.
  - $\mathrm{tr}(\boldsymbol{A})$ and $|\boldsymbol{A}|$: denotes the trace and determinant, respectively, of matrix $\boldsymbol{A}$.
  - $\boldsymbol{I}_n$: the $n \times n$ identity matrix.
  - $\boldsymbol{j}_n = (1, \ldots, 1)^T$ is an n-vector of ones, and $\boldsymbol{J}_{m,n}$ is an $m \times n$ matrix of ones.
  - $\mathbb{E}(X)$ and $\mathbb{V}(X)$: expectation and variance of random variable X.
  - $\boldsymbol{x} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: the m-dimensional random vector $\boldsymbol{x}$ has a normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.
  - $X \sim t(n, \lambda)$: a t distribution with n degrees of freedom and noncentrality parameter λ. If $\lambda = 0$ we write simply: $X \sim t(n)$.
  - $X \sim F(n_1, n_2, \lambda)$: an F distribution with n_1 and n_2 numerator and denominator degrees of freedom respectively, and noncentrality parameter λ. If $\lambda = 0$ we write simply: $X \sim F(n_1, n_2)$.
- Possibly useful results.
  - If all the eigenvalues of $n \times n$ matrix $\boldsymbol{A}$ are less than 1 in absolute value, then

$$(\boldsymbol{I}_n - \boldsymbol{A})^{-1} = \boldsymbol{I}_n + \sum_{k=1}^{\infty} \boldsymbol{A}^k.$$

Problems

1. Consider the vector of observations $\boldsymbol{y} = (y_1, \dots, y_n)^T$ from the linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{X}$ is an $n \times k$ matrix of (full) rank k. Note that, with $\boldsymbol{x}_i = (x_{i,1}, \dots, x_{i,k})^T$ denoting the vector of covariates for the i-th case, $(y_i, \boldsymbol{x}_i)$, $i = 1, \dots, n$, we can write:

$$\boldsymbol{X}^T\boldsymbol{X} = \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i^T, \qquad \text{and} \qquad \boldsymbol{X}^T\boldsymbol{y} = \sum_{i=1}^{n} y_i\boldsymbol{x}_i.$$

Recall that if $\hat{\boldsymbol{\beta}}$ denotes the usual OLS estimator of $\boldsymbol{\beta}$, the i-th *residual* is $r_i = y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}$, and we denote by $P_{ii} = \boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i$ the i-th diagonal element of the *hat matrix*. The goal of this Problem is to obtain an expression for the difference $\hat{\boldsymbol{\beta}}^{(-i)} - \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}^{(-i)}$ denotes the OLS estimator of $\boldsymbol{\beta}$ with the i-th case omitted. (This is leads to the definition of measures of influence such as Cook's distance.)

   (a) Let vectors $\boldsymbol{u}$ and $\boldsymbol{b}$ be such that $|\boldsymbol{b}^T\boldsymbol{u}| < 1$, and assume that the eigenvalues of matrix $\boldsymbol{u}\boldsymbol{b}^T$ are less than 1 in absolute value. Using the "possibly useful results" on page 1, or otherwise, prove that:

$$(\boldsymbol{I}_n - \boldsymbol{u}\boldsymbol{b}^T)^{-1} = \boldsymbol{I}_n + \frac{1}{1 - \boldsymbol{b}^T\boldsymbol{u}}\boldsymbol{u}\boldsymbol{b}^T.$$

   (b) Using (a), or otherwise, prove the following special case of the Sherman-Morrison identity:

$$(\boldsymbol{A} - \boldsymbol{a}\boldsymbol{b}^T)^{-1} = \boldsymbol{A}^{-1} + \frac{1}{1 - \boldsymbol{b}^T\boldsymbol{A}^{-1}\boldsymbol{a}}\boldsymbol{A}^{-1}\boldsymbol{a}\boldsymbol{b}^T\boldsymbol{A}^{-1},$$

   where $\boldsymbol{A}$ is invertible and the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are such that $\boldsymbol{b}^T\boldsymbol{A}^{-1}\boldsymbol{a} \neq 1$. (Assume all the required conditions for the validity of the result in (a) also hold here.)

   (c) With $\boldsymbol{A} = \boldsymbol{X}^T\boldsymbol{X}$ and $\boldsymbol{c} = \boldsymbol{X}^T\boldsymbol{y}$, show that: $\hat{\boldsymbol{\beta}}^{(-i)} = (\boldsymbol{A} - \boldsymbol{x}_i\boldsymbol{x}_i^T)^{-1}(\boldsymbol{c} - y_i\boldsymbol{x}_i)$.

   (d) Use the above results to obtain a (simple) expression for $\hat{\boldsymbol{\beta}}^{(-i)} - \hat{\boldsymbol{\beta}}$ as a function of only: r_i, P_{ii}, $\boldsymbol{x}_i$, and the matrix $\boldsymbol{A}$ defined in (c).

2. In order to estimate the two parameters θ and ϕ, observations y_i, $i = 1, \dots, N$, are taken, each observation being additively contaminated by IID errors $\epsilon_i \sim \mathrm{N}(0, \sigma^2)$. A total of $N = n + m + m$ are observations, $y_i = \mu_i + \epsilon_i$, are thus collected according to the following scheme:

   - n observations having mean $\mu_i = \theta$, for $i = 1, \dots, n$;
   - m observations having mean $\mu_i = \theta - \phi$, for $i = n + 1, \dots, n + m$;
   - m observations having mean $\mu_i = \phi - \theta$, for $i = n + m + 1, \dots, N$.

   In order to answer the following questions, note that this can be set up as the standard linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, whence we denote by $\hat{\boldsymbol{\beta}} = (\hat{\theta}, \hat{\phi})^T$ the usual LSE of $\boldsymbol{\beta} = (\theta, \phi)^T$. To this end, define the following statistics:

$$S_1 = \sum_{i=1}^{n} y_i, \qquad S_2 = \sum_{i=n+1}^{n+m} y_i, \qquad S_3 = \sum_{i=n+m+1}^{N} y_i.$$

(a) Show that the covariance matrix of $\hat{\boldsymbol{\beta}}$ is:

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \frac{\sigma^2}{2mn} \qquad \begin{bmatrix} 2m & 2m \\ 2m & 2m+n \end{bmatrix}.$$

(b) Explicitly compute $\hat{\boldsymbol{\beta}}$.

(c) Explicitly compute s^2, the usual unbiased estimate of σ^2.

(d) Construct a 95% confidence interval for $\phi - \theta$.

(e) Determine a joint 95% confidence region for the two parameters in $\boldsymbol{\beta}$.

3. Consider the vector of 3 observations $\boldsymbol{y} = (y_1, y_2, y_3)^T$ from the linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, given in expanded form as:

$$\boldsymbol{y} = \begin{pmatrix} \beta_1 + \beta_2 + \beta_3 \\ \beta_1 + \beta_3 \\ \beta_2 \end{pmatrix} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N(0, \sigma^2 \boldsymbol{I}_3).$$

Define the following linear combinations of the parameters:

$$\eta_1 = \beta_1, \qquad \eta_2 = \beta_2, \qquad \eta_3 = \beta_3, \qquad \eta_4 = \beta_1 - 2\beta_2 + \beta_3, \qquad \eta_5 = \beta_1 - 3\beta_3.$$

In answering the following question parts, be sure to carry all calculations explicitly in order to yield results in as simple a form as possible.

(a) Show that only η_2 and η_4 are *estimable*.

(b) By working with the original rank-deficient parametrization, explicitly compute the BLUE of η_4, and determine its distribution.

(c) Reparametrize the model to full rank, i.e., construct a matrix $\boldsymbol{U}$ such that $\boldsymbol{\gamma} = \boldsymbol{U}\boldsymbol{\beta}$ is the new parameter vector in the full rank linear model $\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$. Then find the BLUE of η_4 in this context, and show that it is the same as in (b).

(d) Is it possible to find other *linear and unbiased* estimators of η_4 that are different from the BLUE? If it is possible, find one such estimator and compare its variance to that of the BLUE. If it is not possible, give a clear justification (proof) of why that is the case.

(e) Construct a level α two-sided test of the null hypothesis $H_0 : \eta_4 = 0$, and clearly state the rejection rule.