Applied Statistics Preliminary Examination
Theory of Linear Models
May 2022

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 12 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 $\boldsymbol{A}$.
  - $\operatorname{tr}(\boldsymbol{A})$ 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.
  - $\mathbb{E}(\boldsymbol{x})$ and $\mathbb{V}(\boldsymbol{x})$: expectation and variance of random vector $\boldsymbol{x}$.
  - $\boldsymbol{x} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: 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 n-dimensional vector $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}_n)$, where $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{b}$ for an $(n \times k)$ full-rank matrix of constants $\boldsymbol{X}$ with $k < n$. Define the matrix $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$, and the quadratic forms $Q_1 = \boldsymbol{y}^T\boldsymbol{H}\boldsymbol{y}$ and $Q_2 = \boldsymbol{y}^T(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y}$.

   (a) Show that both of the matrices $\boldsymbol{H}$ and $\boldsymbol{I}_n - \boldsymbol{H}$ are symmetric and idempotent.

   (b) Find all the eigenvalues and ranks of each of $\boldsymbol{H}$ and $\boldsymbol{I}_n - \boldsymbol{H}$.

   (c) Find a constant c such that cQ_1 and cQ_2 have familiar distributions, and specify what these resulting distributions are.

   (d) Find the distribution of
   $$W = \frac{(n-k)Q_1}{kQ_2}.$$

2. Consider fitting the linear model where the full-rank $(n \times k)$ model matrix is decomposed as $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$, with $\boldsymbol{X}_1$ $(n \times k_1)$, $\boldsymbol{X}_2$ $(n \times k_2)$, $k = k_1 + k_2$, $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)$, and the normal spherical errors assumption, $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \boldsymbol{I}_n)$:
   $$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}.$$
   Suppose however that the vector of observations $\boldsymbol{y}$ comes from the reduced model $\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}$, and it is known that the column spaces of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are orthogonal, i.e., $\mathcal{C}(\boldsymbol{X}_1) \perp \mathcal{C}(\boldsymbol{X}_2)$. Thus, by fitting the model with model matrix $\boldsymbol{X}$ we are in fact *overfitting*. Let $\widehat{\boldsymbol{\beta}}^T = (\widehat{\boldsymbol{\beta}}_1^T, \widehat{\boldsymbol{\beta}}_2^T)$, denote the resulting LSE of $\boldsymbol{\beta}$, and $s^2 = (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})^T(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})/(n-k)$ the usual estimator of σ^2.

   (a) Find $\mathbb{E}(\widehat{\boldsymbol{\beta}}_1)$ and $\mathbb{E}(\widehat{\boldsymbol{\beta}}_2)$. Is $\widehat{\boldsymbol{\beta}}_1$ unbiased for $\boldsymbol{\beta}_1$?

   (b) Compute $\mathbb{V}(\widehat{\boldsymbol{\beta}}_1)$, $\mathbb{V}(\widehat{\boldsymbol{\beta}}_2)$, and $\text{Cov}(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2)$.

   (c) If $\mathcal{P}_{\mathcal{C}(\boldsymbol{X})}$, $\mathcal{P}_{\mathcal{C}(\boldsymbol{X}_1)}$, and $\mathcal{P}_{\mathcal{C}(\boldsymbol{X}_2)}$ denote the respective projection matrices onto the column spaces $\mathcal{C}(\boldsymbol{X})$, $\mathcal{C}(\boldsymbol{X}_1)$, and $\mathcal{C}(\boldsymbol{X}_2)$, show that $\mathcal{P}_{\mathcal{C}(\boldsymbol{X})} = \mathcal{P}_{\mathcal{C}(\boldsymbol{X}_1)} + \mathcal{P}_{\mathcal{C}(\boldsymbol{X}_2)}$.

   (d) Show that even though the model is overfit, s^2 is still unbiased for σ^2.

3. Consider the linear model $y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_p x_{i,p} + \epsilon_i$, where $i = 1, \dots, n$, $\{\epsilon_i\} \sim$ IID $\mathrm{N}(0, \sigma^2)$, and the $x_{i,j}$ are elements of the (n by $p+1$) design matrix $\boldsymbol{X}$ (of rank $k < p+1 < n$). Let $\boldsymbol{y} = (y_1, \dots, y_n)^T$, $\overline{y} = \boldsymbol{j}_n^T \boldsymbol{y}/n$ be the sample mean of the y_i, and let $\widehat{\boldsymbol{\beta}} = \boldsymbol{G}\boldsymbol{X}^T\boldsymbol{y}$ denote the LSE of $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ corresponding to the generalized inverse $\boldsymbol{G}$ of $\boldsymbol{X}^T\boldsymbol{X}$. Additionally, let $R^2 = (\widehat{\boldsymbol{\beta}}^T \boldsymbol{X}^T \boldsymbol{y} - n\overline{y}^2)/(\boldsymbol{y}^T\boldsymbol{y} - n\overline{y}^2)$ be the usual *coefficient of determination*, and define the quadratic forms:

$$Q_1 = (\overline{y}\boldsymbol{j}_n - \boldsymbol{y})^T(\overline{y}\boldsymbol{j}_n - \boldsymbol{y}), \qquad Q_2 = (\overline{y}\boldsymbol{j}_n - \boldsymbol{X}\widehat{\boldsymbol{\beta}})^T(\overline{y}\boldsymbol{j}_n - \boldsymbol{X}\widehat{\boldsymbol{\beta}}), \qquad Q_3 = (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})^T(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}).$$

(a) Without using calculus, prove that $\boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = 0$. (Your argument should use the form of $\widehat{\boldsymbol{\beta}}$ and derive any properties of generalized inverses that are deemed necessary in the proof.)

(b) Prove that $Q_1 = Q_2 + Q_3$.

(c) If $\mathcal{C}(\boldsymbol{X})$ denotes the column space of $\boldsymbol{X}$, prove or disprove the statement: $R^2 = 1$ if and only if $\boldsymbol{y} \in \mathcal{C}(\boldsymbol{X})$.

(d) Assuming that $\eta = 3\beta_1 - \beta_0$ is *estimable*, show how to construct a test of $H_0 : \eta = 2$ vs. $H_1 : \eta \neq 2$. Clearly define the test statistic, its distribution under both H_0 and H_1, and the rejection rule.

Applied Statistics Preliminary Examination
Theory of Linear Models
August 2022

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 13 parts carries an equal weight of 10 points.
- Abbreviations/Acronyms.
  - IID (independent and identically distributed).
  - SSE (sum of squared errors). Also called *residual sum of squares*.
  - 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.
  - $\mathbb{E}(\boldsymbol{x})$ and $\mathbb{V}(\boldsymbol{x})$: expectation and variance of random vector $\boldsymbol{x}$.
  - $\boldsymbol{x} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: 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.
  - Note the inverse for the patterned matrix $\boldsymbol{A}$:

$$\boldsymbol{A} = \begin{bmatrix} 9 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{bmatrix} \qquad \Longrightarrow \qquad \boldsymbol{A}^{-1} = \frac{1}{3} \begin{bmatrix} 1 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}.$$

Problems

1. An experiment was conducted to compare the yields for three varieties of corn (1, 2, and 3). Three plants of each variety were grown and the yield recorded, for a total of 9 observations, $\boldsymbol{y} = (y_1, \dots, y_9)^T$, where $\{y_1, y_2, y_3\}$ correspond to variety 1, $\{y_4, y_5, y_6\}$ variety 2, etc. In order to compare the mean yields of each variety, three linear models of the form $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ were proposed, where $\boldsymbol{\epsilon} \sim \text{N}(0, \sigma^2 \boldsymbol{I}_9)$. Letting $\{\mu_1, \mu_2, \mu_3\}$ be the true mean yields for varieties 1, 2, and 3, respectively, and defining the 9-dimensional vectors $\boldsymbol{j}_9$ (vector of 9 ones), $\boldsymbol{x}_1 = (\boldsymbol{j}_3, 0_6)^T$, $\boldsymbol{x}_2 = (0_3, \boldsymbol{j}_3, 0_3)^T$, $\boldsymbol{x}_3 = (0_6, \boldsymbol{j}_3)^T$ the models are as follows.

   Model A (cell means): $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3]$ and $\boldsymbol{\beta} = (\mu_1, \mu_2, \mu_3)^T$, which corresponds to:

$$y_i = \begin{cases} \mu_1 + \epsilon_i, & i = 1, 2, 3 \\ \mu_2 + \epsilon_i, & i = 4, 5, 6 \\ \mu_3 + \epsilon_i, & i = 7, 8, 9 \end{cases}$$

   Model B (reference cell mean): $\boldsymbol{X} = [\boldsymbol{j}_9, \boldsymbol{x}_2, \boldsymbol{x}_3]$ and $\boldsymbol{\beta} = (\mu, \alpha_2, \alpha_3)^T$, which corresponds to:

$$y_i = \begin{cases} \mu + \epsilon_i, & i = 1, 2, 3 \\ \mu + \alpha_2 + \epsilon_i, & i = 4, 5, 6 \\ \mu + \alpha_3 + \epsilon_i, & i = 7, 8, 9 \end{cases}$$

   Model C (effects): $\boldsymbol{X} = [\boldsymbol{j}_9, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3]$ and $\boldsymbol{\beta} = (\mu, \beta_1, \beta_2, \beta_3)^T$, which corresponds to:

$$y_i = \begin{cases} \mu + \beta_1 + \epsilon_i, & i = 1, 2, 3 \\ \mu + \beta_2 + \epsilon_i, & i = 4, 5, 6 \\ \mu + \beta_3 + \epsilon_i, & i = 7, 8, 9 \end{cases}$$

   The main goal of this Problem is to make inference on the contrasts, $\eta_1 = \mu_1 - \mu_2$ and $\eta_2 = \mu_1 + \mu_2 - 2\mu_3$, and to determine if these inferences depend on which model is fitted. To this end, let $\widehat{\boldsymbol{\beta}}$ denote the LSE of $\boldsymbol{\beta}$ in each respective model, and $\overline{y}_1 = (y_1 + y_2 + y_3)/3$, $\overline{y}_2 = (y_4 + y_5 + y_6)/3$, and $\overline{y}_3 = (y_7 + y_8 + y_9)/3$, be the sample means corresponding to each of varieties 1, 2, and 3.

   (a) Explain the meaning of each element of the parameter vector $\boldsymbol{\beta}$ in each of Models A and B. What is the relationship between these elements in each respective Model? (E.g., μ_1 is the mean of variety 1 in Model A; what parameter(s) does it correspond to in Model B?)

   (b) Compute the BLUEs of η_1 and η_2. Do the results depend on which of Models A or B is used? Why? (Note the "possibly useful result" on page 1.)

   (c) Compute SSE, and verify that it is the same regardless of which of Models A or B is used.

   (d) Construct a $(1-\alpha)100\%$ confidence interval for η_1 in the context of Models A and B, and show that the intervals are identical.

   (e) Construct a level α test for $H_0 : \eta_2 = 0$ in the context of Models A and B, and show that the tests are identical.

2. This Problem continues the analysis of Problem 1, but in the context of fitting the over-parametrized Model C to the vector of 9 observations, $\boldsymbol{y}$.

   (a) Characterize *all* the estimable functions of the type $\boldsymbol{\lambda}^T\boldsymbol{\beta}$. Express the contrasts η_1 and η_2 in terms of the Model C parameters, and hence show that they are estimable.

   (b) Are the BLUEs of η_1 and η_2 the same as computed earlier in the context of Models A and B? Either compute them, or give a solid argument for your reasoning.

   (c) Show that SSE is the same as for Models A and B, by either computing it, or by giving a solid reason why this should be the case.

   (d) Construct a level α test for $H_0 : \eta_2 = 0$. Is the test the same as computed earlier in the context of Models A and B?

   (e) Is it possible to remove the rank-defficiency in Model C? Justify your answer, and if so, show explicitly how this can be done using one of the two methods available.

3. Let $\boldsymbol{A}$ be an $m \times n$ non-zero matrix of rank r. Let $\boldsymbol{B}$ and $\boldsymbol{K}$ be nonsingular matrices (of orders m and n, respectively) such that

$$\boldsymbol{A} = \boldsymbol{B} \begin{bmatrix} \boldsymbol{I}_r & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{K}.$$

Consider the matrix $\boldsymbol{G}$ defined as:

$$\boldsymbol{G} = \boldsymbol{K}^{-1} \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{U} \\ \boldsymbol{V} & \boldsymbol{W} \end{bmatrix} \boldsymbol{B}^{-1},$$

for some $r \times (m - r)$ matrix $\boldsymbol{U}$, $(n - r) \times r$ matrix $\boldsymbol{V}$, and $(n - r) \times (m - r)$ matrix $\boldsymbol{W}$.

   (a) Show that $\boldsymbol{G}$ is a generalized inverse of $\boldsymbol{A}$.

   (b) Show that if $\boldsymbol{H}$ is a generalized inverse of $\boldsymbol{A}$, then it must be of the same form as $\boldsymbol{G}$ for some choice of the matrices $\{\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}\}$.

   (c) Show that different choices for the matrices $\{\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}\}$ lead to distinct generalized inverses of $\boldsymbol{A}$.