Applied Statistics Preliminary Examination
Theory of Linear Models
May 2023

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 $\boldsymbol{A}$.
  - $\text{tr}(\boldsymbol{A})$ 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. This question will review covariance and the associated correlation matrices, and demonstrate some techniques for constructing these from combinations of other covariance and correlation matrices, while ensuring the all-important requirement of positive definiteness. Recall that if $\boldsymbol{\Sigma} > 0$ is a (positive definite) covariance matrix, then the associated correlation matrix is $\boldsymbol{R} = \boldsymbol{V}^{-1}\boldsymbol{\Sigma}\boldsymbol{V}^{-1}$, where $\boldsymbol{V}$ is a diagonal matrix consisting of the diagonal entries of $\boldsymbol{\Sigma}$.

   (a) Show that a positive linear combination of positive definite matrices is again positive definite. That is, for positive scalars $\alpha_i > 0$ and positive definite matrices $\boldsymbol{A}_i > 0$, $i = 1, \dots, k$, show that
   $$\sum_{i=1}^{k} \alpha_i \boldsymbol{A}_i$$
   is positive definite.

   (b) Show that if $\boldsymbol{B}$ is any square matrix and $\boldsymbol{\Sigma}$ is positive definite, then $\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^T$ is positive definite.

   (c) If $\boldsymbol{A}_0, \boldsymbol{A}_1, \boldsymbol{A}_2$ are nonzero square matrices, $\boldsymbol{a}_1$ is a nonzero vector, and $\boldsymbol{\Sigma}$ is a positive definite covariance matrix, consider the matrix:
   $$\boldsymbol{B} = \boldsymbol{A}_0\boldsymbol{A}_0^T + \boldsymbol{A}_1(\boldsymbol{a}\boldsymbol{a}^T)\boldsymbol{A}_1^T + \boldsymbol{A}_2\boldsymbol{\Sigma}\boldsymbol{A}_2^T.$$
   Determine, with justification, whether or not $\boldsymbol{B}$ is positive definite.

   (d) Let R_0, R_1, R_2 be correlation matrices, and $\theta_1 \geq 0$ and $\theta_2 \geq 0$ be nonnegative scalars such that $\theta_1 + \theta_2 \leq 1$. Consider the matrix:
   $$\boldsymbol{C} = (1 - \theta_1 - \theta_2)R_0 + \theta_1 R_1 + \theta_2 R_2.$$
   Determine, with justification, whether or not $\boldsymbol{C}$ is a correlation matrix.

2. Consider the linear model
   $$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}, \qquad \boldsymbol{\epsilon} \sim (0, \sigma^2\boldsymbol{I}_n),$$
   where $\boldsymbol{X}$ is a full-rank $(n \times k)$ model matrix, and, as the notation suggests, the elements ϵ_i of the error vector $\boldsymbol{\epsilon}$ are uncorrelated with mean zero and a common variance of σ^2. Recall that $\hat{\boldsymbol{\mu}} = \boldsymbol{P}\boldsymbol{y} = \hat{\boldsymbol{y}}$ is the usual LSE of $\boldsymbol{\mu}$, where $\boldsymbol{P}$ denotes the projection matrix onto $\mathcal{C}(\boldsymbol{X})$. Define the quantity:
   $$r^* = \mathbb{E}\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|^2.$$
   (This is called the *risk* of the model in estimating $\boldsymbol{\mu}$.) We seek unbiased estimators of r^*. Throughout this Problem (except where indicated), assume that $\boldsymbol{\beta}$ is unknown, but σ^2 is known.

   (a) Show that $r^* = \|(\boldsymbol{I}_n - \boldsymbol{P})\boldsymbol{\mu}\|^2 + k\sigma^2$.

   (b) Show that $\mathbb{E}\|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2 = r^* + (n - 2k)\sigma^2$, and hence propose an unbiased estimator of r^*.

   (c) If $\boldsymbol{\epsilon} \sim \mathrm{N}(0, \sigma^2\boldsymbol{I}_n)$, find the distribution of $\|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2/\sigma^2$.

   (d) If σ^2 is unknown, propose an unbiased estimator of r^*.

3. Consider the following linear model for the observations $\{y_1, y_2, y_3\}$:

$$y_1 = \beta_1 + \beta_2 + \beta_3 + \epsilon_1,$$
$$y_2 = \beta_1 + \beta_3 + \epsilon_2,$$
$$y_3 = \beta_2 + \epsilon_3,$$

where $\{\epsilon_1, \epsilon_2, \epsilon_3\} \sim$ IID $\mathrm{N}(0, \sigma^2)$, and $\{\beta_1, \beta_2, \beta_3\}$ are unknown parameters to be estimated.

(a) Write out the model in the usual matrix form, $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Determine the rank of the design matrix $\boldsymbol{X}$, and find a generalized inverse of $\boldsymbol{X}^T\boldsymbol{X}$.

(b) Determine the form of all *estimable* functions $\eta = \boldsymbol{\lambda}^T\boldsymbol{\beta}$. Use this form to decide which of the following are estimable:

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

In particular, show that η_4 is estimable.

(c) Find the BLUE of η_4, and find one other linear unbiased estimator (LUE) of η_4 that is different from the BLUE.

(d) Construct a test of the null hypothesis:

$$H_0 : \eta_2 = 0, \qquad \text{and} \qquad \eta_4 = 3.$$

Clearly define the test statistic, and compute its distribution both under H_0 and under the alternative hypothesis, H_1.

Applied Statistics Preliminary Examination
Theory of Linear Models
August 2023

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 $\boldsymbol{A}$.
  - $\text{tr}(\boldsymbol{A})$ and $|\boldsymbol{A}|$: denotes the trace and determinant, respectively, of matrix $\boldsymbol{A}$.
  - $\boldsymbol{I}_n$: the $n \times n$ identity matrix.
  - $\boldsymbol{j}_n = (1, \dots, 1)^T$ 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 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ were proposed, where $\boldsymbol{\epsilon} \sim \mathrm{N}(0, \sigma^2 \boldsymbol{I}_9)$. 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 $\boldsymbol{A}$.