Design of Experiment: Prelim Problems

May 2017

Please Do All Problems. Each of the 14 parts carries an equal weight of 10 points. Note: each scenario in question 1 counts as three parts.

For each test, state the null and alternative hypotheses in terms of the model parameters.

1. A researcher comes to you for advice about the analysis of an experiment she has conducted. She has used 5 treatments and she has 9 observations of some response variable Y for each treatment. She shows you the following ANOVA from a computer printout:

Source	d.f.	S.S.	F statistic	$P[> F]$
Treatments	4	100	8.33	0.0001
Error	40	120		

The researcher wants to know from you whether the analysis is correct, and if so what it means. For each of the three scenarios below answer the following questions:

(a) What is the appropriate model? Write a model equation, including any and all effects. Explain which effects are fixed and which are random, and state any distributional assumptions.

(b) Based on the model in (a), what is the corresponding ANOVA (include sources of variation, d.f., formulas for sums of squares, and the statistic for testing $H_0: \tau_1 = \tau_2 = \cdots = \tau_5$).

(c) Is the ANOVA from the printout above appropriate for testing $H_0: \tau_1 = \tau_2 = \cdots = \tau_5$?

Scenario I: 45 animals were used for the study and each treatment was applied to 9 animals selected at random.

Scenario II: 45 animals were used but they came from 9 different litters of size 5 each, and each treatment was assigned to one animal selected at random from each litter.

Scenario III: 15 animals were used, each treatment was assigned to 3 animals selected at random, and 3 observations were made on each animal.

2. Sugar cane is very sensitive to climate and crop management practices. A sugar yield experiment involved 4 management practices (MANAGE) at each of 10 locations (LOCATION) in Louisiana. There were 2 fields assigned to each management practice at each location. The yield of extracted sugar (tons per acre) was calculated for each field. Use the partial SAS output below to answer the following questions.

Dependent Variable: yield

Source	DF	Sum of Squares
Model		
Error		
Corrected	79	82.426

Source	DF	Sum Square	Mean Square	F Value	Pr > F	E(MS)
MANAGE						
LOCATION		51.012			-	
MANAGE*LOCATION		15.865				

MANAGE	yield LSMEAN
1	4.666
2	5.053
3	4.841
4	4.267

(a) State whether you would consider each factor to be fixed or random and explain your reasoning.

(b) State the statistical model and the corresponding assumptions. (Use unrestricted model).

(c) Complete the ANOVA table using the information above and summarize the results of the F tests.

(d) The researcher plans to compare each alternative management practice (MANAGE=2, 3, and 4) with the standard practice (MANAGE=1) and also compare the average of the alternative practices versus the standard. Please construct these confidence intervals with overall confidence level at least 90%.

(e) Construct confidence interval for the variance component for factor LOCATION with 95% confidence level. (Round d.f. to the nearest integer if it cannot be found in tables.)

Applied Statistics Preliminary Examination
Theory of Linear Models
May 2017

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 $\boldsymbol{A}$.
  - $\text{tr}(\boldsymbol{A})$ 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.
  - $\mathbb{E}(X)$ 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 $\boldsymbol{\mu}$ 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 matrix $\boldsymbol{A}$ is given in block form as

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{0} & \boldsymbol{A}_{22} \end{pmatrix}, \qquad \text{or} \qquad \boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{0} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix},$$

then $|\boldsymbol{A}| = |\boldsymbol{A}_{11}| \cdot |\boldsymbol{A}_{22}|$.

1. Let $\boldsymbol{A}$ be a positive definite symmetric matrix, and let $\boldsymbol{B}$ be formed from $\boldsymbol{A}$, both given in block form:

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} \boldsymbol{I} & -\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}, \quad \text{and note that} \quad (\boldsymbol{B}\boldsymbol{A})^{-1} = \begin{pmatrix} \boldsymbol{C}^{-1} & \boldsymbol{0} \\ -\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{C}^{-1} & \boldsymbol{A}_{22}^{-1} \end{pmatrix},$$

where $\boldsymbol{C} = \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$. In addition, consider a random vector $\boldsymbol{x}$ with (positive definite) variance-covariance matrix $\boldsymbol{\Sigma}$, given in block form along with its inverse as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \sigma^{11} & \boldsymbol{\sigma}^{12} \\ \boldsymbol{\sigma}^{21} & \boldsymbol{\Sigma}^{22} \end{pmatrix}, \quad \text{with } \sigma_{11} \text{ and } \sigma^{11} \text{ scalars.}$$

(a) Produce an expression for $\boldsymbol{A}^{-1}$ in terms of $\boldsymbol{C}^{-1}$ and the various block elements $\boldsymbol{A}_{ij}$ of $\boldsymbol{A}$.

(b) Show that $|\boldsymbol{A}| = |\boldsymbol{C}| \cdot |\boldsymbol{A}_{22}|$.

(c) Show that $|\boldsymbol{\Sigma}| \le \sigma_{11} \cdot |\boldsymbol{\Sigma}_{22}|$, with equality if and only if $\boldsymbol{\sigma}_{21} = \boldsymbol{0}$.

(d) Show that $\sigma^{11} = 1/\sigma_{11}$, if and only if $\boldsymbol{\sigma}_{21} = \boldsymbol{0}$.

2. Suppose that the true model relating $\boldsymbol{y}$ and the $n \times p$ matrix $\boldsymbol{X}$ is:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathrm{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}), \tag{$\dagger$}$$

where $\boldsymbol{X}_1$ is $n \times p_1$ and $\boldsymbol{X}_2$ is $n \times p_2$, with conformal $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, and $p_1 + p_2 = p$. Assume that $\boldsymbol{X}$, $\boldsymbol{X}_1$, and $\boldsymbol{X}_2$ are each of full rank, and define $\boldsymbol{G}_i = (\boldsymbol{X}_i^T\boldsymbol{X}_i)^{-1}$, for $i = 1, 2$.

(a) Construct a size α test of $H_0 : \boldsymbol{\beta}_2 = 0$.

(b) Fit the reduced model $\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}$ and compute the bias of the LSE $\hat{\boldsymbol{\beta}}_1$ of $\boldsymbol{\beta}_1$, given that ($\dagger$) is the true model. State conditions under which the bias of $\hat{\boldsymbol{\beta}}_1$ is zero.

(c) Compute the usual unbiased estimator of σ^2 under the reduced model in (b), and give its bias given that ($\dagger$) is the true model. State conditions under which the bias of $\hat{\boldsymbol{\beta}}_1$ is zero.

(d) Fit the original model ($\dagger$) under the simplifying assumption that the columns of $\boldsymbol{X}_1$ are orthogonal to the columns of $\boldsymbol{X}_2$, and compute the LSE of $\hat{\boldsymbol{\beta}}$ in terms of $\boldsymbol{y}$, $\boldsymbol{X}_i$, and $\boldsymbol{G}_i$, for $i = 1, 2$.

(e) For given $\boldsymbol{\lambda}_2 \neq \boldsymbol{0}$, construct a size α test of $H_0 : \boldsymbol{\lambda}_2^T\boldsymbol{\beta}_2 = 0$ under the model and conditions specified in (d).

3. Consider the one-way anova model $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, for $i = 1, 2, 3$ and $j = 1, \ldots, n_i$, with $\epsilon_{ij} \sim$ IID $N(0, \sigma^2)$ for all i and j, and note that the total sample size is $N = n_1 + n_2 + n_3$. The model can be written in vector and matrix form as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{y} = (y_{11}, \ldots, y_{1,n_1}, y_{21}, \ldots, y_{2,n_2}, y_{31}, \ldots, y_{3,n_3})^T, \qquad \boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2, \alpha_3)^T.$$

(a) Show that it is possible to find a generalized inverse $\boldsymbol{G}$ of $\boldsymbol{X}^T\boldsymbol{X}$ in the form of a diagonal matrix, with diagonal elements $\mathrm{diag}(\boldsymbol{G}) = (0, g_1, g_2, g_3)^T$, and determine the values of g_1, g_2, and g_3.

(b) Find the LSE of $\boldsymbol{\beta}$ corresponding to the generalized inverse $\boldsymbol{G}$ found in (a).

(c) Show that functions of the form $\eta = c(\mu + \alpha_i)$ are *estimable*, for any $c \in \mathbb{R}$ and any $i = 1, 2, 3$.

(d) Construct a test of H_0 below, and state the distribution of the test statistic under H_0:

$$H_0 : \frac{1}{5}(\mu + \alpha_1) = \frac{1}{10}(\mu + \alpha_2) = \frac{1}{15}(\mu + \alpha_3).$$

(e) Assuming that $\sigma^2 = 1/2$, compute an expression for the *power* of the test in (d) under the alternative hypothesis:

$$H_1 : \frac{1}{5}(\mu + \alpha_1) - \frac{1}{10}(\mu + \alpha_2) = 1, \quad \text{and} \quad \frac{1}{5}(\mu + \alpha_1) - \frac{1}{15}(\mu + \alpha_3) = 3.$$