Preliminary Examination

Do all three problems. The asterisk * indicates the transpose of a vector or matrix, $\oplus$ is the symbol for direct sum (of two subspaces of a vector space), and $\otimes$ is the symbol for tensor product (of two vector spaces).

1. Let A be an $n \times q$-matrix of full rank $q \in \mathbb{N}$, and B an $n \times q$-matrix of full rank $r \in \mathbb{N}$, with $2 \leq q + r < n$. Assume that $A^*B = 0$. Introduce the subspaces $U = \{x = Ay : y \in \mathbb{R}^q\}$ and $V = \{x = Bz : z \in \mathbb{R}^q\}$ of $\mathbb{R}^n$.

   (a) Show that $U \perp V$.

   Consider the coordinate-free linear model $Y = \eta + E$, $\eta \in \mathbb{L} = U \oplus V$, where $E \stackrel{d}{=} N_n(0, \sigma^2 I_n)$, $0 < \sigma^2 < \infty$. The null hypotheses is given by $H_0 : \eta \in \mathbb{L}_0 = V$.

   (b) Determine $\mathbb{L}_1 = \mathbb{L}_0^{\perp} \cap \mathbb{L}$.

   (c) Express the orthogonal projections P, P_0, and P_1 on $\mathbb{L}$, $\mathbb{L}_0$, and $\mathbb{L}_1$ respectively in terms of the matrices A and B.

   (d) Compute the test statistic for testing $H_0 : \eta \in \mathbb{L}_0$, exploiting the results in 1(c).

   (e) What is the distribution of this test statistic under H_0?

2. Let the vectors $1_n = (1, \ldots, 1)^*, x = (x_1, \ldots, x_n)^*$, and $y = (y_1, \ldots, y_n)^*$ be mutually orthogonal, and write $z = (x_1 y_1, \ldots, x_n y_n)^*$. Consider the regression model

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{bmatrix} 1 & x_1 & x_1 y_1 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n y_n \end{bmatrix} \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} E_1 \\ \vdots \\ E_n \end{pmatrix} = X\theta + E,$$

   where the design matrix $X = [1_n \mid x \mid z]$ is of full rank (note that the vertical line $\mid$ indicates separation between columns), $\theta = (\alpha, \beta_1, \beta_2) \in \mathbb{R}^3$, and the $E_1, \ldots, E_n$ are iid $N(0, \sigma^2)$ for some $0 < \sigma^2 < \infty$.

   (a) Compute $(X^*X)^{-1}$.

   (b) Compute the maximum likelihood estimators of α, β_1, β_2 and σ^2.

   (c) Compute the test statistic for testing $H_0 : \beta_2 = 0$, and specify its distribution under the null hypothesis.

   (d) Find a confidence interval (of arbitrary level) for the parameter α.

Notational convention. In problems 3 and 4 below the following notation will be useful: for any $m \in \mathbb{N}$ write $1_m = (1, \ldots, 1)^* \in \mathbb{R}^m$, $P_m = \frac{1}{m} 1_m 1_m^*$ for the orthogonal projection onto $[1_m]$, and $P_m^\perp$ for the orthogonal projection onto $[1_m]^\perp \cap \mathbb{R}^m$.

3. Consider the model $Y_{ijk} = \mu + \gamma_{ij} + E_{ijk}$ ($i = 1, 2; j = 1, 2; k = 1, 2$), where μ and the γ_{ij} are real parameters, and the E_{ijk} are iid $N(0, \sigma^2)$ for some $0 < \sigma^2 < \infty$.

   (a) Express the vector

   $$Y = (Y_{111}, Y_{112}, Y_{121}, Y_{122}, Y_{211}, Y_{212}, Y_{221}, Y_{222})^*$$

   in terms of the parameter $\theta = (\mu, \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})^*$ by means of a design matrix X.

   (b) Is γ_{12} estimable? Motivate your answer.

   Henceforth, we will in addition assume that $\sum_{i=1}^{2} \sum_{j=1}^{2} \gamma_{ij} = 0$. Let us introduce the notation $\Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}$. Using tensor products the model says that Y lies in a subspace $\mathbb{L}$ of $\mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \mathbb{R}^2$.

   (c) Give a concise description of the model using tensor products.

   (d) The subspace $\mathbb{L}$ corresponding to the model can be written as $\mathbb{L} = \mathbb{L}_1 \oplus \mathbb{L}_2$, where $\mathbb{L}_1$ is associated with μ, and $\mathbb{L}_2$ with Γ. Determine $\mathbb{L}_1$ and $\mathbb{L}_2$.

   (e) Verify that $\mathbb{L}_1$ and $\mathbb{L}_2$ are orthogonal, and specify their dimensions.

   (f) Compute the orthogonal projections of Y onto $\mathbb{L}_1$ and $\mathbb{L}_2$.

   (g) Are the projections in 3(f) stochastically independent? Why?

Design of Experiments: Prelim Problems
May 2013

Please Do All the Problems

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

1. A consumer organization studied the effect of age of automobile owner on size of cash offer for a used car by utilizing 12 persons in each of three age groups (young, middle, elderly) who acted as the owner of a used car. A medium price, six-year-old car was selected at random for the experiment and the "owners" solicited cash offers for this car from 36 dealers selected at random from the dealers in the region. Randomization was used in assigning the dealers to the "owners." The following summary statistics are obtained for offers (hundreds of dollars):

Age Group	Mean
Young	21.50
Middle	27.75
Elderly	21.42
Overall	23.56

   SSE $= 82.17$

   a) Write a linear model for this experiment. Make sure to clearly define the terms as they relate to this design and state the assumptions.

   b) Write the ANOVA table and conduct the F-test of equality of factor level means. Use $\alpha = 0.01$.

   c) Conduct all pairwise comparisons by controlling the experimentwise error rate at the 0.05 level.

2. A medical study seeks to understand the effects of a new prescription drug and a nutritional supplement for treating a rare disorder, based on a single response variable. The study uses four treatment conditions: (1) placebo, (2) drug only, (3) supplement only, and (4) both drug and supplement. Human patients enter the study at different times. Once 4 patients have entered the study, the 4 treatments are randomly assigned to these patients (each patient receiving a different treatment). This process continues for each new block of 4 consecutive patients, until a total of 24 patients have entered the study. Let Y_{ij} denote the response of patient from j-th block who is assigned treatment condition i, for $i = 1, \cdots, 4$, and $j = 1, \cdots, 6$

Suppose

$$\bar{y}_{1\bullet} = 20, \quad \bar{y}_{2\bullet} = 15, \quad \bar{y}_{3\bullet} = 20, \quad \bar{y}_{4\bullet} = 5,$$

$$\sum_j (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})^2 = 79, \quad \sum_i \sum_j (y_{ij} - \bar{y}_{\bullet\bullet})^2 = 1531$$

and suppose that, within each treatment condition, Y_{ij} is normally distributed, with the same variance for all responses, regardless of treatment condition.

(a) Write a linear model equation appropriate for analyzing the responses Y_{ij} from this experiment, clearly state the assumptions and conditions for this model.

(b) Write the ANOVA table appropriate for your model including E(MS).

(c) Test whether there are any differences between the treatment conditions (Use $\alpha = 0.05$).

(d) Test whether the effects of the drug and supplement on the responses Y_{ij} could be additive, i.e. whether or not there appears to be interaction between the effects of the drug and supplement (Use $\alpha = 0.05$).

(e) Can you assert that the response mean is significantly smaller when the drug is given without the supplement versus when only a placebo is given? Test at level $\alpha = 0.05$.

3. Let Y_{ijk} denote the response from the split-plot that receives the j-th level of factor B, in the k-th whole plot that receives the i-th level of factor A. Often the following model is used:

$$Y_{ijk} = \mu + \tau_i + \eta_{k(i)} + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, \quad i = 1, \cdots, a, \quad j = 1, \cdots, b, \quad k = 1, \cdots, n$$

where

$$\eta_{k(i)} \overset{iid}{\sim} N(0, \sigma_\eta^2), \quad \text{independent of} \quad \varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$$

$$\sum_i \tau_i = \sum_j \beta_j = \sum_i (\tau\beta)_{ij} = \sum_j (\tau\beta)_{ij} = 0$$

(a) Derive the expected value and variance of $\bar{Y}_{ij\bullet} - \bar{Y}_{i'j\bullet}$ for $i \neq i'$, where $\bar{Y}_{ij\bullet} = \frac{1}{n} \sum_k Y_{ijk}$

(b) Derive the expected value of

$$\sum_{i=1}^a \sum_{k=1}^n (\bar{Y}_{i\bullet k} - \bar{Y}_{i\bullet\bullet})^2$$