

(2)

Show your calculations and provide suitable motivation for full credit.

The asterisk (*) indicates the transpose of a vector or matrix, \oplus is the symbol for the direct sum of two subspaces of a vector space, and \otimes is the symbol used for tensor products. For any $m \in \mathbb{N}$ write $\mathbf{1}_m = (1, \dots, 1)^* \in \mathbb{R}^m$, $P_m = \frac{1}{m} \mathbf{1}_m \mathbf{1}_m^*$ for the orthogonal projection onto $[\mathbf{1}_m]$, and P_m^\perp for the orthogonal projection onto $[\mathbf{1}_m]^\perp \cap \mathbb{R}^m$; I_m is the $m \times m$ identity matrix.

1. Let

$$P_0 = \frac{1}{6} \begin{pmatrix} 1 & -1 & 2 \\ -1 & 1 & -2 \\ 2 & -2 & 4 \end{pmatrix}$$

and $L_0 \subset \mathbb{R}^3$ the range of this transformation. Furthermore let $a = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$ and L be the subspace of \mathbb{R}^3 spanned by this range and the vector a . Consider the linear model $Y = \eta + E$, $\eta \in L$, $E \stackrel{d}{=} N_3(0, \sigma^2 I_3)$, for some unknown $\sigma^2 \in (0, \infty)$.

- Prove that P_0 is a projection and determine its range.
 - Determine the subspace L and the projection P onto L .
 - Determine a confidence region of level $1 - \alpha$ ($\alpha \in (0, 1)$) for $\eta_1 = P_1 \eta$, where P_1 is the projection onto $L_0^\perp \cap L = L_1$.
2. Consider the linear model $Y_{ij} = \alpha_i + (-1)^{j+1} \beta x + E_{ij}$, for $i = 1, 2$ and $j = 1, 2$, where the α_i and β are real parameters, and the error variables E_{ij} are i.i.d. $N(0, \sigma^2)$ for unknown $\sigma^2 \in (0, \infty)$, and where the number $x \neq 0$ is known.
- Express $Y = (Y_{11}, Y_{12}, Y_{21}, Y_{22})^*$ in terms of the parameter $\theta = (\alpha_1, \alpha_2, \beta)^*$ by means of a designs matrix X .
 - Compute the *ML* estimators $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}$ of $\alpha_1, \alpha_2, \beta$, respectively.
 - Compute the unbiased version $\bar{\sigma}^2$ of the *ML* estimator of σ^2 .
 - Determine the test statistic for testing the null hypothesis $H_0 : \beta = 0$, and give its distribution under H_0 .
3. Let be given the linear model $Y_{ijk} = \mu + \alpha_i + \beta_{ij} + \gamma x_i y_j z_k + E_{ijk}$ for $i = 1, \dots, I$, $j = 1, \dots, J$, and $k = 1, \dots, K$ ($I \times J \times K > 3 + I + IJ$). In this model the $\mu, \alpha_i, \beta_{ij}$, and γ are real parameters, and the E_{ijk} are i.i.d. $N(0, \sigma^2)$ for some unknown $\sigma^2 \in (0, \infty)$. We will, moreover, assume that $a \perp 1_I$, $B \perp [1_I] \otimes [1_J]$, and $z \perp 1_K$, where

$$a = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_I \end{pmatrix}, \quad B = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1J} \\ \vdots & & \vdots \\ \beta_{I1} & \cdots & \beta_{IJ} \end{pmatrix},$$

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_I \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_J \end{pmatrix}, \quad z = \begin{pmatrix} z_1 \\ \vdots \\ z_K \end{pmatrix}.$$

Using tensor product notation the model says that \mathbf{Y} (the vector of all Y_{ijk}) lies in a subspace $L \subset \mathbb{R}^I \otimes \mathbb{R}^J \otimes \mathbb{R}^K$.

- (a) Give a concise description of the model using tensor products.
- (b) The subspace L corresponding to the model can be written as $L = L_1 \oplus L_2 \oplus L_3 \oplus L_4$, where the L_j are mutually orthogonal subspaces. Prove this claim.
- (c) Suppose we want to test the null hypothesis $H_0 : \gamma = 0$. Compute the numerator of the LR test statistic for this problem.

Design of Experiments: Prelim Problems
August 2014

- Do All the Problems.
- For each test, state the null and alternative hypotheses in terms of the model parameters.
- All parts have equal weights.

1. A process engineer is testing the yield of a product manufactured on three machines. Each machine can be operated at two power settings. Furthermore, a machine has three stations on which the product is formed. An experiment is conducted in which each machine is tested at both power settings, and three observations on yield are taken from each station. The runs are made in random order.

- (a) Write a linear model for the data resulting from this experiment. Make sure to clearly define the terms as they relate to this design and state the assumptions.
- (b) Break down the total sum of squares into the sum of squares of the individual effects. Make sure to write each sum of squares explicitly.
- (c) Write the ANOVA table appropriate for your model and include the $E(MS)$ column.
- (d) Assume that the manufacturer has several machines but chooses three machines at random for this experiment. Write the changes, if any, you would make to your answers to parts (a) and (b).

2. Twenty-four comparable plots of land are used as experimental units to study strategies for controlling blight (disease) in potatoes. Under investigation are two fungicides (F1 versus F2) and two application times (Early versus Late). The response is yield of blight-free potatoes from each plot. The experiment is conducted in a CRD with 5 treatment groups. Group descriptions, sizes, and response summary statistics are as follows:

| Treatment Group: | No Fungicide | F1, Early | F1, Late | F2, Early | F2, Late |
|------------------|--------------|-----------|----------|-----------|----------|
| Group Size | 8 | 4 | 4 | 4 | 4 |
| Sample Mean | 15 | 26 | 20 | 32 | 30 |
| Sample Variance | 5 | 8 | 10 | 11 | 10 |

- (a) Write out a linear (cell) means model equation appropriate for analysis of this data. Clearly define your notation and specify any conditions on terms.
- (b) Compute the corresponding ANOVA table appropriate for your model.
- (c) Form a contrast representing the difference of the means of the two fungicides, and perform a one-sided test against the assertion that F2 is more effective than F1 ($\alpha = 0.05$). Remember to state H_0 and H_1 in terms of the treatment means defined in part (a).

- (d) Form a contrast representing interaction between fungicide type and application time, and perform a test for interaction using this contrast ($\alpha = 0.05$). Remember to state H_0 and H_1 in terms of the treatment means defined in part (a).
- (e) Form Fisher's 95% simultaneous two-sided confidence intervals for the pairwise differences in mean yield for each active treatment condition versus the "No Fungicide" condition. What do you conclude from these?
- (f) Suppose this experiment is to be conducted in a balanced CRD. Find the minimum number of plots per treatment group such that the 95% Fisher's confidence intervals for the differences of the means are narrower than the confidence intervals in part (e). What are the new widths? Show your work.