

## Preliminary Examination: LINEAR MODELS

Answer all questions and show all work.

1. An investigator proposes an unusual design to study conditions under which bacteria flourish. The experiment consists of two phases. Each phase of the experiment takes one week. The two phases are run over consecutive weeks. Only 10 petrie dishes can be experimented on in a given week.

Phase 1: In a temperature-controlled environment, the experimenter will grow 10 petrie dishes of bacteria under the null condition, which is called condition 0. It takes one week to grow these bacteria. At the end of the week, she will measure the bacteria count in each dish, say  $W_{0t}$ ;  $t = 1, \dots, 10$ .

Phase 2: In the same temperature-controlled environment, the experimenter will grow 10 petrie dishes of bacteria. Each dish will be grown under its own unique condition (conditions 1,  $\dots$ , 10), leading to the data  $W_{i1}$ ;  $i = 1, \dots, 10$ .

The experimenter believes that the data are most appropriately analyzed on a transformed scale and wishes to work with  $Y_{it} = \log(W_{it})$ , where  $\log$  represents the natural logarithm. The model for the data is,

$$Y_{it} = \mu_i + \epsilon_{it}$$

for appropriate  $i$  and  $t$ , where  $\{\mu_i\}$  are fixed but unknown constants, and  $\{\epsilon_{it}\}$  are independent and identically distributed  $N(0, \sigma^2)$  random variables.

Later in the question, use  $\eta_i$  to denote the median of the distribution of  $Y_{it}$ . Use  $\mu_i^W$  to denote the mean of  $W_{it}$ .

- a. Under the experimenter's assumption, provide a formula for a 95% confidence interval for  $\mu_1 - \mu_2$ . (*Hint: You can rely on the estimate of  $\sigma^2$  from Phase 1.*)
- b. Suppose that the 10 treatments examined in Phase 2 of the experiment have a  $2 \times 5$  factorial structure. The level of Factor A is 1 for conditions 1 through 5, and it is 2 for conditions 6 through 10. The level of Factor B is  $1 + i(\text{mod}5)$  for condition  $i$ ;  $i = 1, \dots, 10$ . Under a *main-effects* model, but relying only on the estimate of  $\sigma^2$  from Phase 1, provide a formula for a 95% confidence interval for  $\mu_1 - \mu_6$ . (Points will be deducted for a very wide interval.)
- c. Describe a graphical summary that would help you decide whether the main-effects model is sufficient, or whether there is a need to expand the mode. Sketch pictures illustrating a situation where there is no need to expand the model and others illustrating a situation where this is a need to expand the mode. Your artwork need not

be perfect, but it should convey your point. In addition, describe a hypothesis test that would help you decide whether the main-effects model is adequate.

- d. From your answer in (a), transform it to obtain a confidence interval for the ratio of the median of  $W_{1t}$  to the median of  $W_{2t}$ . Can this resulted interval be interpreted as an approximate 95% confidence interval for the ratio of the mean of  $W_{1t}$  to the mean of  $W_{2t}$ ? Explain briefly.

2. Consider a very small normal mixed linear model with

$$\begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \mu + \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \end{pmatrix}$$

where as usual, the  $\epsilon_{ij}$ 's are iid  $N(0, \sigma^2)$  independent of  $\alpha_i$  that are iid  $N(0, \sigma_\alpha^2)$ . Two ANOVA sums of squares for this model can respectively be built from the random vectors

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \mathbf{Y} \text{ and } (1 \ 1 \ -1 \ -1) \mathbf{Y}$$

**Argue very carefully** that these two vectors (one is  $2 \times 1$  and one that is  $1 \times 1$ ) are independent.

3. a. Show that a least squares estimator of  $\theta$  in the general linear model

$$\mathbf{Y} = \mathbf{W}\theta + \varepsilon, \quad \varepsilon \sim (\mathbf{0}, \sigma^2 \mathbf{I})$$

is given by  $\hat{\theta} = (\mathbf{W}'\mathbf{W})^{-}\mathbf{W}'\mathbf{Y}$ , where  $(\mathbf{W}'\mathbf{W})^{-}$  is a generalized inverse of  $\mathbf{W}'\mathbf{W}$ . Here,  $\mathbf{Y}$  and  $\varepsilon$  are  $n \times 1$  random vectors of response and error variables, respectively;  $\theta$  is a  $p \times 1$  parameter vector;  $\mathbf{W}$  is an  $n \times p$  matrix.

- b. Now partition  $\mathbf{W}$  and  $\theta$  as  $\mathbf{W} = [\mathbf{X} \ \mathbf{Z}]$  and  $\theta = [\tau' \ \gamma']'$  to obtain the following general linear model:

$$\mathbf{Y} = \mathbf{X}\tau + \mathbf{Z}\gamma + \varepsilon, \quad \varepsilon \sim (\mathbf{0}, \sigma^2 \mathbf{I}).$$

This model could be used for a block design where  $\tau$  is a parameter vector of the effects of  $v$  treatments,  $\gamma$  is a parameter vector of the effects of  $b$  blocks, and  $X$  is the  $n \times v$  design matrix of rank  $v$  which has element 1 in town  $u$  column  $i$  if the  $u$ th observations is on treatment  $i$ ;  $Z$  is the  $n \times b$  design matrix of rank  $b$  which has element 1 in row  $u$  column  $j$  if the  $u$ th observation is on block  $j$ .

- b-i. Obtain an expression for the least squares estimator,  $\hat{\tau}$ , for  $\tau$  in the form  $\mathbf{C}\hat{\tau} = \mathbf{Q}$ , where the expressions for  $\mathbf{C}$  and  $\mathbf{Q}$  are in terms of  $\mathbf{X}$  and  $\mathbf{Z}$ .

- b-ii. Prove that  $\mathbf{1}_v$  is an eigenvector of  $\mathbf{C}$  and that the rank of  $\mathbf{C}$  much be less than  $v$ .

- c. Let  $\mathbf{C}^-$  be the Moore-Penrose generalized inverse of  $\mathbf{C}$ . Let  $\mathbf{h}'_i \tau$ ,  $i = 1, \dots, g$ , be a set of  $g$  contrasts in the treatment effects  $\tau$ , and let  $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_g]'$ .

- c-i. Give a formula for the average variance of contrast estimators  $\mathbf{h}'_1 \hat{\boldsymbol{\tau}}, \mathbf{h}'_2 \hat{\boldsymbol{\tau}}, \dots, \mathbf{h}'_g \hat{\boldsymbol{\tau}}$  in terms of  $\mathbf{H}$  and  $\mathbf{C}^-$ .
- c-ii. Suppose that  $\mathbf{C}$  has rank  $v - 1$ . Let  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_v$  be a set of orthonormal eigenvectors for  $\mathbf{C}$  corresponding to eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_{v-1}, 0$ . Write  $\mathbf{C}^-$  in terms of its spectral decomposition and, hence, give a formula for the average variance of the least squares estimators of the contrasts  $\mathbf{h}'_1 \hat{\boldsymbol{\tau}}, \mathbf{h}'_2 \hat{\boldsymbol{\tau}}, \dots, \mathbf{h}'_g \hat{\boldsymbol{\tau}}$  in terms of  $\mathbf{H}$  and the eigenvalues and eigenvectors of  $\mathbf{C}$ .
- c-iii. The objective in designing a blocked experiment is often to minimize the average variance of the least squares estimators of a set of contrasts  $\mathbf{H}\hat{\boldsymbol{\tau}}$ . Thus, it is useful to have a lower bound for their average variance. For  $i = 1, \dots, v-1$ , set  $u_i^2 = \lambda_i^{-1} \mathbf{x}'_i \mathbf{H}' \mathbf{H} \mathbf{x}_i$  and  $\mathbf{v}_i^2 = \lambda_i$ . Use the Cauchy-Schwarz inequality to obtain a lower bound for the average variance of the above contrast estimators, in the class of design for which  $\text{tr}(\mathbf{C}) = t$  where  $t$  is some positive constant. (*Hint: This bound is a function of some or all of the eigenvalues and eigenvectors of  $\mathbf{C}$ ,  $t$  and  $\mathbf{H}$ .*)
- c-iv. Suppose that  $\mathbf{C} = a\mathbf{I}_v - bv^{-1}\mathbf{J}_v$ . Show that, if the rows of  $\mathbf{H}$  are a set of  $g = v - 1$  orthonormal contrasts, these form a set of  $v - 1$  orthonormal eigenvectors of  $\mathbf{C}$ . Obtain the corresponding eigenvalues.
- c-v. Using any of the information that you've gained so that, and  $\mathbf{C}$  and  $\mathbf{H}$  was defined in part (iv), prove that the average variance of the contrast estimators  $\mathbf{H}\hat{\boldsymbol{\tau}}$  is a minimum over all block designs of the same size and with fixed  $\text{tr}(\mathbf{C}) = t$ .
4. The pressure drop measured across an expansion valve in a turbine is being studied. The design engineer considers the important variables that influence pressure drop reading to be gas temperature on the inlet side (A), operator (B), and the specific pressure gauge used by operator (C). These three factors are arranged in a factorial design, with A and B fixed, C random. There are two replicates. The appropriate model for this design is

$$Y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \epsilon_{ijkl}$$

where  $\tau_i$  is the effect of A,  $\beta_j$  is B, and  $\gamma_k$  is C.

- State model assumptions and complete the ANOVA table.
- Test the effects on A, AC, ABC factors, respectively. You only need to give the test statistics, and their associated distribution and degrees of freedom under  $H_0$ , respectively.
- Now assume that A is fixed effect, but B and C random (the model form is unchanged). Test the gas temperature effect, i.e.,  $H_0 : \tau_i = 0$  for all  $i$ . You only need to give the test statistic, and the associated distribution and degrees of freedom under  $H_0$ .

Source	SS	df	$MS$	$E(MS)$
A	1023.36	2	?	?
B	423.82	3	?	?
C	7.19	2	?	?
AB	1211.97	?	?	?
AC	137.89	?	?	?
BC	209.47	?	?	?
ABC	166.11	?	?	?
Error	770.50	?	?	?
Total	3950.32	71		