

**Preliminary Examination:  
LINEAR MODELS**

Answer all questions and show all work.  
Q1, Q2, Q3, and Q4 are 20 points each.

1. For a linear model  $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$  with  $\sigma^2 > 0$  where  $\mathbf{X}$  is an  $n \times (p + 1)$  matrix for predictors and  $\mathbf{I}$  is an  $n \times n$  identity matrix. Here  $\mathbf{X}$  is full rank with  $r(\mathbf{X}) = p + 1$ . We are considering the following hypotheses:

$$H_0 : \boldsymbol{\beta} = \mathbf{0} \text{ vs } H_1 : \boldsymbol{\beta} \neq \mathbf{0}$$

- a. Find the maximum likelihood estimators for  $\boldsymbol{\beta}$  and  $\sigma^2$  under the alternative hypothesis  $H_1$ .
- b. Find the maximum likelihood estimator for  $\sigma^2$  under the null hypothesis  $H_0$ .
- c. Show that the likelihood ratio test for the above hypotheses can be based on the test statistic:

$$F = \frac{\hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{Y} / (p + 1)}{(\mathbf{Y}' \mathbf{Y} - \hat{\boldsymbol{\beta}}' \mathbf{X} \mathbf{Y}) / (n - p - 1)},$$

where  $\hat{\boldsymbol{\beta}}$  denotes the OLS estimator for  $\boldsymbol{\beta}$ .

2. Let  $\mathbf{X}_{k \times 1} \sim N_k(\boldsymbol{\Sigma}\boldsymbol{\theta}, \sigma^2\boldsymbol{\Sigma})$  with  $\sigma^2 > 0$  and  $\boldsymbol{\Sigma}$  is positive definite. Let

$$\mathbf{B} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{1}(\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{\Sigma}^{-1},$$

where  $\mathbf{1} = (1, \dots, 1)'$ .

- a. Show that (i)  $\mathbf{B}$  is symmetric; (ii)  $\text{rank}(\mathbf{B}) = k - 1$ ; (iii)  $\mathbf{B}\boldsymbol{\Sigma}$  is idempotent.
  - b. Define  $\mathbf{Y} = \mathbf{B}\mathbf{X}$ . Find the distribution of  $\mathbf{Y}$ .
  - c. Obtain the distribution of  $\mathbf{Y}'\boldsymbol{\Sigma}\mathbf{Y}$  when (i)  $\boldsymbol{\theta} = \mathbf{0}$ , and (ii)  $\boldsymbol{\theta} \neq \mathbf{0}$ , respectively.
3. Suppose you have  $n$  subjects, labeled  $i = 1, \dots, n$ . On each you have  $r_i$  replicates of a response  $Y$ , and also  $p$  covariates  $X_1, \dots, X_p$ . Consider two models. The first (referred to as Model 1) is a model for the original observations, in the form

$$Y_{ij} = \beta_0 + X_{i1}\beta_1 + \dots + X_{ip}\beta_p + \epsilon_{ij},$$

where  $\epsilon_{ij}$ 's are uncorrelated with mean zero and variance  $\sigma^2$ ;  $Y_{ij}$  denotes the  $j$ th observation for the  $i$ th subject, for  $j = 1, \dots, r_i$ , and  $i = 1, \dots, n$ .

The second model (referred to as Model 2) is a model for  $\bar{Y}_i$ , the within-subject averages, in the form

$$\bar{Y}_i = \beta_0 + X_{i1}\beta_1 + \dots + X_{ip}\beta_p + \epsilon_i,$$

where  $\bar{Y}_i = r_i^{-1} \sum_{j=1}^{r_i} Y_{ij}$  and  $\bar{\epsilon}_i = r_i^{-1} \sum_{j=1}^{r_i} \epsilon_{ij}$ .

- a. Let  $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$ . For Model 1, write down a formal statement of the *ordinary* least squares problem to which the ordinary least squares estimate  $\hat{\beta}_{OLS}$  is the solution. Similarly, for Model 2, write down the *weighted* least squares problem to which the weighted least squares estimate  $\hat{\beta}_{WLS}$  is the solution.
  - b. Show that  $\hat{\beta}_{WLS} = \hat{\beta}_{OLS}$ .
  - c. Let  $e_{ij}$  be the residual from estimating  $Y_{ij}$  under Model 1, and let  $\bar{e}_i$  be the residual from estimating  $\bar{Y}_i$  under Model 2. Provide a simple expression relating  $\bar{e}_i$  to  $e_{ij}$ , for  $j = 1, \dots, r_i$ .
  - d. Suppose that you wish to assess Model 1 for lack of fit. Provide a test statistic for doing so and state its distribution under the appropriate null hypothesis. Do the same for Model 2. Be sure to specify any assumptions that are needed.
4. Consider a company that buys raw material in batches from three different suppliers. The purity of this raw material varies considerably, which causes problems in manufacturing the finished product. We wish to determine whether the variability in purity is attributable to differences between the suppliers. Four batches of raw material are selected at random from each supplier, and three determinations of purity are made on each batch. The data, after coding by subtracting 93, are shown in the table below. Use  $\alpha = 0.05$ . Use your calculator and statistical tables attached.

Coded Purity Data (Code:  $Y_{ijk} = \text{Purity} - 93$ ) are given below. Here,  $i$  is the index for suppliers,  $j$  is the index for batches, and  $k$  is the index for replicates, for  $k = 1, 2, 3$ ,  $j = 1, \dots, 4$ , and  $i = 1, 2, 3$ .

	Supplier 1				Supplier 2				Supplier 3			
Batches	1	2	3	4	1	2	3	4	1	2	3	4
	1	-2	-2	1	1	0	-1	0	2	-2	1	3
	-1	-3	0	4	-2	4	0	3	4	0	-1	2
	0	-4	1	0	-3	2	-2	2	0	2	2	1
<b>Batch Totals</b> $Y_{ij}$	0	-9	-1	5	-4	6	-3	5	6	0	2	6
<b>Supplier Totals</b> $Y_{i..}$	-5				4				14			

- a. State an appropriate statistical model for the data including model assumptions.
- b. Complete the ANOVA table.

Source	df	SS	MS
Suppliers, A		15.06	
Batched(suppliers), B(A)			
Error, E		63.33	
Total		148.31	

- c. Find the expected mean squares (EMS) of A, B(A) and E based on your model assumptions.
- d. Find point estimates of variance components in this model.
- e. Test that the factor A and B(A) are significant, respectively.