

1. Write a new solution to any part of a problem from Exam 1 for which you did not receive full credit.
2. Suppose we will conduct an experiment with a completely randomized design. We are interested in understanding the effects of two treatment factors  $A$  and  $B$  on the mean of the response variable. Each of these factors has three levels. We will assume that there is no interaction between the factors and will model the response variable as

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \quad i = 1, 2, 3 \quad j = 1, 2, 3 \quad k = 1, \dots, n_{ij}$$

where  $y_{ijk}$  denotes the response for the  $k^{\text{th}}$  experimental unit treated with level  $i$  of factor  $A$  and level  $j$  of factor  $B$ ;  $\mu, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3$  represent unknown parameters; the  $\varepsilon_{ijk}$  are i.i.d. random variables with mean 0 and unknown positive variance  $\sigma^2$ ; and  $n_{ij}$  denotes the number of experimental units treated with level  $i$  of factor  $A$  and level  $j$  of factor  $B$  ( $i = 1, 2, 3; j = 1, 2, 3$ ).

Based on available resources we are initially considering an experimental design where  $n_{11} = n_{13} = n_{23} = n_{32} = 5$  and  $n_{ij} = 0$  for all other combinations of  $i$  and  $j$ .

- (a) To understand the effects of factor  $A$  on the response, we are interested in estimating linear functions of the form  $\alpha_i - \alpha_{i'}$  for  $i < i'$ . List all such functions that would be estimable under the initially proposed design.
  - (b) To understand the effects of factor  $B$  on the response, we are interested in estimating linear functions of the form  $\beta_j - \beta_{j'}$  for  $j < j'$ . List all such functions that would be estimable under the initially proposed design.
  - (c) Assume that the initially planned observations ( $n_{11} = n_{13} = n_{23} = n_{32} = 5$ ) are essentially free but that any additional observation from an experimental unit treated with level  $i$  of factor  $A$  and level  $j$  of factor  $B$  will cost  $10 * i + 100 * j$  dollars. If we wish to be able to estimate all possible linear functions of the form  $\alpha_i - \alpha_{i'}$  for  $i < i'$  and  $\beta_j - \beta_{j'}$  for  $j < j'$  at minimal cost, how many additional observations are needed and at which levels of factors  $A$  and  $B$  should the observation(s) be taken?
3. In a first course on multiple regression, you learned about a test that is sometimes referred to as the *general linear test*. The general linear test statistic is given by

$$F = \frac{[SSE(R) - SSE(F)]/[df_R - df_F]}{SSE(F)/df_F}.$$

Here  $SSE(R)$  denotes the error or residual sum of squares for a reduced model,  $SSE(F)$  denotes the error or residual sum of squares for a full model,  $df_R$  denotes the degrees of freedom associated with  $SSE(R)$ , and  $df_F$  denotes the degrees of freedom associated with  $SSE(F)$ .

Assume the model  $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$  holds where  $\underline{\varepsilon}$  has a multivariate normal distribution with mean  $\underline{0}$  and dispersion  $\sigma^2 I$ . Assume, without loss of generality, that the reduced model is obtained from the full model by setting  $\beta_{m+1} = \cdots = \beta_K = 0$  in accordance with the null hypothesis  $H_0 : \beta_{m+1} = \cdots = \beta_K = 0$  for some  $m < K$ . Let  $X_1$  denote the matrix consisting of the first  $m$  columns of the design matrix  $X$ , let  $X_2$  denote the matrix consisting of the last  $K - m$  columns of  $X$ , let  $\underline{\beta}_1 = (\beta_1, \dots, \beta_m)'$ , and let  $\underline{\beta}_2 = (\beta_{m+1}, \dots, \beta_K)'$ . Finally suppose  $\text{Rank}(X) = K$ .

- (a) Prove that the distribution of  $F$  is noncentral  $F$  with numerator and denominator degrees of freedom  $df_R - df_F$  and  $df_F$ , respectively.
- (b) Determine the noncentrality parameter for  $F$  and show that it is 0 if and only if  $H_0 : \beta_{m+1} = \cdots = \beta_K = 0$  is true.
- (c) Determine a matrix  $L'$  such that  $L'\underline{\beta} = \underline{0}$  is equivalent to  $\beta_{m+1} = \cdots = \beta_K = 0$ .
- (d) Show that the test statistic

$$\frac{(L'\hat{\underline{\beta}})'(L'(X'X)^{-1}L)^{-1}L'\hat{\underline{\beta}}}{(K - m)\hat{\sigma}^2}$$

is equal to the general linear test statistic for testing  $H_0 : \beta_{m+1} = \cdots = \beta_K = 0$ , where  $\hat{\sigma}^2 = \frac{\underline{y}'(I - P_X)\underline{y}}{T - K}$ .