Sample Questions for the PhD Preliminary Exam in Linear Models

Department of Mathematical Sciences
University of Cincinnati
January 2013

1. Suppose Y_1, Y_2, and Y_3 are measurements of the angles of a triangle subject to error. The information is given as a linear model $Y_i = \theta_i + \epsilon_i$, where θ_i's are the true angles, $i = 1, 2, 3$. Assume that $E(\epsilon_i) = 0$, and $Var(\epsilon_i) = \sigma^2$. Obtain the least squares estimates of θ_i (subject to the constraint $\sum_{i=1}^3 \theta_i = 180$).

2. (a) Let $\mathbf{x} = (X_1, \dots, X_k)^T \sim N_k(\mu, \mathbf{D})$, where μ is a $k \times 1$ vector and $\mathbf{D} = \text{diag}\{\sigma_1^2, \dots, \sigma_k^2\}$, $r(\mathbf{D}) = k$. Find the mean and variance of the random variable $U = \mathbf{x}^T \mathbf{D}^{-1} \mathbf{x}$.
   (b) Let $\mathbf{x} = (X_1, \dots, X_k)^T \sim N_k(\mu, \mathbf{\Sigma})$, where μ is a $k \times 1$ vector with $r(\mathbf{\Sigma}) = k$. What is the distribution of $U = (\mathbf{x} - \mu)^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \mu)$?
   (c) Suppose $\mathbf{A} = \mathbf{D}^{-1} - (\mathbf{D}^{-1}\mathbf{1}_k\mathbf{1}_k^T\mathbf{D}^{-1})/\mathbf{1}_k^T\mathbf{D}^{-1}\mathbf{1}_k$. Assume that $\mathbf{x} \sim N_k(\mu, \mathbf{D})$ distribution. Find the distribution of $\mathbf{x}^T \mathbf{A} \mathbf{x}$.

3. Consider the linear model $\mathbf{y} = X\beta + \epsilon$, $r(X) = r < p$ with $E(\epsilon) = 0$ and $Var(\epsilon) = \sigma^2 \mathbf{V}$, where $\mathbf{V}$ is a known p.d. matrix.
   (a) Find a least square solution of β.
   (b) Show that $\mathbf{c}^T\beta$ is estimable if and only if $\mathbf{c}^T(X^T\mathbf{V}^{-1}X)^{-}(X^T\mathbf{V}^{-1}X) = \mathbf{c}^T$ where A^{-} denotes a generalized inverse of A.

4. In the one-way ANOVA model, $Y_{ij} = \mu + \tau_i + \epsilon_{ij}$, $j = 1, \dots, n$, $i = 1, 2, 3$, with $N(0, \sigma^2)$ errors. τ_i's are fixed but unknown.
   (a) Consider a hypothesis $H_{01} : \mu + \tau_1 = 2(\mu + \tau_2) = 3(\mu + \tau_2)$. Is H_{01} testable?
   If yes, derive a test of H_{01}.
   (b) Is $H_{02} : \tau_2 = (\tau_1 + \tau_3)/2$ testable? If yes, derive a test of H_{02}.

5. Consider a two factor completely randomized design with Factor A at two levels and Factor B at three levels. Assume that two observations are to be made for each combination of the levels of A and B.
   Assume also that both factors A and B are random.
   Let y_{ijk} be the kth observation corresponding to the ith level of A and the jth level of B.
   Also let $Y = (y_{111}, y_{112}, \dots, y_{221}, y_{232})'$ be the vector of all observations.

(a) Write down the model for data y_{ijk} that includes interaction, and state all the assumptions.

(b) Find the mean and the covariance matrix of Y based on the model in (a).

(c) Find the EMS for factor A, B and A*B based on the model in (a). Show details.

6. Consider a paper manufacturer who is interested in three different pulp preparation methods (the methods differ in the amount of hardwood in the pulp mixture) and four different cooking temperatures for the pulp and who wishes to study the effect of these two factors on the tensile strength of the paper. Each replicate of a factorial experiment requires 12 observations, and the experimenter has decided to run three replicates. However, the pilot plant is only capable of making 12 runs per day, so the experimenter decides to run one replicate on each of the three days and to consider the days or replicates as blocks. On any day, he conducts the experiment as follows. A batch of pulp is produced by one of the three methods under study. Then this batch is divided into four samples, and each sample is cooked at one of the four temperatures. Then a second batch of pulp is made using another of the three methods. This second batch is also divided into four samples that are tested at the four temperatures. This process is then repeated, using a batch of pulp produced by the third methods. Assume that blocks is random effect and that Pulp preparation methods and Temperature are fixed effects. Use the following information:

Source of Variation	df	SS	MS	F	E(MS)
Block		77.55			
Preparation Method (A)		128.39			
Whole Plot error		36.28			
Temperature (B)		434.08			
A*B		75.17			
Subplot error		71.50			
Total		822.97			

(a) State an appropriate statistical model including model assumptions.

(b) Complete the ANOVA table above.

(c) Analyze the data and draw your conclusions. Use $\alpha = 0.05$.