Preliminary Examination - Multivariate Models

1:00-3:30pm, Monday, May 3, 2021

Answer all questions and show all your work.

1. 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: $y_{ijk} = Purity_{ijk} - 93$)

		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
	y_{ijk}	-1	-3	0	4	-2	4	0	3	4	0	-1	2
		0	-4	1	0	-3	2	-2	2	0	2	2	1
Batch Totals	$y_{ij.}$	0	-9	-1	5	-4	6	-3	5	6	0	2	6
Supplier Totals	$y_{i..}$	-5				4				14			

a. State an appropriate statistical model for the data including model assumptions.

b. Complete the ANOVA table below.

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 factors A and B(A) are significant.

2. The pressure drop measured across an expansion value in a turbine is being studied. The design engineer considers important variables that influence pressure drop reading: 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 and C random. There are two replicates. The appropriate model for this design is

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

where τ_i, β_j, and γ_k are the effects of A, B, and C, respectively.

a. State appropriate model assumptions and complete the ANOVA table.

Source	**SS**	**df**	**MS**	**F**	**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			

b. Test the effects on A, AC, and ABC factors, respectively. You only need to give the test statistics, their associated sampling distribution, and related degrees of freedom under the null hypothesis, respectively.

c. Now assume that the factor A is fixed effect, but B and C are 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 statistics, their associated sampling distribution, and related degrees of freedom under H_0.

3. Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent $N_p\ (\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ random vectors with its pdf

$$f(\mathbf{x}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_0)\right\}$$

where $\boldsymbol{\mu}_0$ **is** a fixed $p \times 1$ vector and $\boldsymbol{\Sigma}$ is a $p \times p$ positive definite (p.d.) matrix.

Let $\overline{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i$.

a. Show that $\frac{1}{|\boldsymbol{\Sigma}|^b} \exp\left\{-\frac{1}{2} tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\right)\right\} \leq \frac{1}{|\mathbf{B}|^b} (2b)^{pb} \exp\{-bp\}$ for all p.d. $\boldsymbol{\Sigma}_{\mathrm{p}\times\mathrm{p}}$ with the equality holding only for $\boldsymbol{\Sigma} = \boldsymbol{B}/(2b)$.

b. Show that $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{X}_i - \boldsymbol{\mu}_0)(\boldsymbol{X}_i - \boldsymbol{\mu}_0)^T$ is the maximum likelihood estimator of **Σ.**

4. Let $\boldsymbol{X}$ be distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $|\boldsymbol{\Sigma}| > 0$. For $i = 1, \dots, p$, let $Z_i = \boldsymbol{e}_i^T(\boldsymbol{X} - \mu)/\sqrt{\lambda_i}$ where $(\lambda_1, \boldsymbol{e}_1), \dots, (\lambda_p, \boldsymbol{e}_p)$ are eigenvalue-eigenvector pairs of $\boldsymbol{\Sigma}$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$.

a. Show that $(Z_1, \dots, Z_p)$ is distributed from $N_p(\boldsymbol{0}, \boldsymbol{I}_p)$ where $\boldsymbol{I}_p$ is a p-dimensional identity matrix.

b. Show that $(\boldsymbol{X} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$ is distributed the chi-square distribution with p degrees of freedom.

5. There is a random sample with size n from bivariate normal population with mean vector $\boldsymbol{\mu} = \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}$ and var-cov matrix $\boldsymbol{\Sigma} = \begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{bmatrix}$. For the $n = 21$ pairs of observations, we observed that

$$\bar{\mathbf{x}} = \begin{bmatrix}.564\\ .603\end{bmatrix}, \boldsymbol{S} = \begin{bmatrix}.15 & .12\\ .12 & .15\end{bmatrix} \text{ and } \boldsymbol{S}^{-1} = \begin{bmatrix}18.52 & -14.81\\ -14.81 & 18.52\end{bmatrix}.$$

Find the simultaneous intervals, i.e., T^2- intervals, for μ_1 and for μ_2.