MTL-766 : Major Exam Online

Time: 2:00-3:45 pm **Additional 15 minutes for uploading by 4:00 pm** **Max Marks 45**

Start each question on a new page and write your name and entry number on top of each sheet. Please ensure correct mapping of questions and page number in gradescope.

1. We wish to determine the equality of three variance-covariance matrices of normal distributed populations representing performance of students in 1st year, 2nd year, and third year in a program. Three random samples, each one comprising of 4 students, is taken from 1st, 2nd, 3rd year batch. Their performance is recorded in two subjects A and B. The sample data is as follows.

1st year		2nd year		3rd year	
A	B	A	B	A	B
20	42	30	25	22	33
29	33	20	38	19	41
32	45	18	27	42	35
25	40	36	36	17	27

Test the equality of three populations at 1% level of significance using the Box-M test. Show each step in the calculation clearly. [10]

2. Consider the two dimensional patterns

$$(2,1),\ (3,5),\ (4,3),\ (5,6),\ (6,7),\ (7,8).$$

(a) Compute the principal component(s). Transform the pattern data onto the principal component. Determine the variance s^2 of the transformed data.

(b) Depict the principal component graphically and depict the projection of the data on the principal component to explain the importance of PCA.

(c) Use the transformed data to test the hypothesis that the transformed data belongs to the normal distribution with variance $s^2 - 0.5$. Explain each of the step clearly. [10]

3. To test the "equal correlation" structure in the population, suppose we set the null hypothesis $H_0 : \rho = \rho_0 = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$ against the alternative hypothesis $H_1 : \rho \neq \rho_0$. The test statistics for large sample is given by

$$T = \frac{n-1}{1-\bar{r}^2}\Big((\sum\sum_{i<k} r_{ik} - \bar{r})^2 - \gamma \sum_{k=1}^{p} (\bar{r}_k - \bar{r})^2 \Big),$$

where $\bar{r}_k$ is the average of off diagonal elements in the k-th column and $\bar{r}$ is the overall mean of the off-diagonal elements in the sample correlation matrix. It is known that T is χ^2- distributed with degree of freedom $(p+1)(p-1)/2$.
Use this to test H_0 at 1% level of significance for the 150 samples with correlation matrix

$$R = \begin{pmatrix} 1 & .7501 & 0.6392 & 0.6363 \\ & 1 & 0.6925 & 0.7386 \\ & & 1 & 0.6625 \\ & & & 1 \end{pmatrix}.$$

The eigenvalues of R are 3.085, 0.382, 0.342, 0.217. Determine the first principal component for R, and analyze the correlations ρ_{Y_1,Z_1} and ρ_{Y_1,Z_2} from it. [10]

4. In a study involving ten attributes, two factors are extracted using sample correlation matrix. The factor loadings are as follows.

Variables	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Factor 1	0.20	-0.30	0.40	0.50	-0.30	0.20	0.40	0.40	0.50	0.20
Factor 2	0.20	-0.30	0.30	0.40	0.30	-0.20	-0.30	-0.70	-0.80	-0.90

Determine the variability explained by factor 1 and factor 2. What are the communality and specificity of the variables. Describe the total error in estimating the original sample data covariance matrix S by a two factor model. Explain each step clearly with justification. [10]

5. Use the Anderson-Darling test to check if the following data is normally distributed

$$6.31,\ 5.89,\ 4.50,\ 3.77,\ 4.25,\ 5.19,\ 5.79$$

Show each step in the calculation of the AD-test when applied to the given data. [5]