B. TECH. (COMPUTER SCIENCE & BUSINESS SYSTEMS) (CBCS - 2018 COURSE)

B.Tech. (CSBS) Sem - III : : SUMMER - 2022

SUBJECT : COMPUTATIONAL STATISTICS

Day : Thursday
Date : 2/6/2022

S-20448-2022

Time : 02:30 PM-05:30 PM
Max. Marks : 60

N. B :

1) All questions are **COMPULSORY**.
2) Figures to the right indicate **FULL** marks.
3) Use of non-programmable **CALCULATOR** is allowed.
4) Assume suitable data if necessary.

Q.1 **a)** Write python program to reverse a number and find the sum of reversed numbers all the digits in reversed number. **(05)**

b) Brief the process of data wrangling and enlist its significance for data analysis. **(05)**

OR

Q.1 Write python code to design employee management system using text file. **(10)**

Q.2 Consider following Dataframe and perform operations mentioned below it. **(10)**

PID	Shop	Raw Material 1	Raw Material 2	fuel	Labor	Transport	Damage less
P 101	S_1	727	256	89	1741	660	270
P 102	S_1	621	338	63	1366	827	292
P 103	S_1	412	336	85	1249	293	393
P 101	S_2	485	230	88	1650	583	453
P 102	S_2	579	227	51	1989	273	375
P 103	S_2	387	310	64	791	403	512

i) Add new column 'Total_Expense' to calculate total cost of product development
ii) Using groupby function find out maximum and minimum price per product.
iii) Apply transform function to find out maximum and minimum labor cost for shop 'S1'.
iv) Visualize given dataset using 5 charts of matplotlib library.

OR

Q.2 **a)** Using any dataframe give the process of pivot table creation. **(05)**

b) Describe the significance of data visualization using suitable data frame. **(05)**

Q.3 State and prove properties of multivariate normal distribution. **(10)**

OR

Q.3 What is multiple linear regression model? Consider the following student performance dataset and assume data values for 10 students. Demonstrate how to detect and handle following regression model pitfalls **(10)**

i) Multicollinearity
ii) Outliers
iii) Non-normality

PRN	Student Name	Sub1 theory	Sub1 practical	Sub 2 theory	Sub2 practical	Sub3 theory	Sub3 practical	Total	Percentage

P.T.O.

Q.4 Compare linear discriminant analysis (LDA) and principle component analysis (PCA). Which of the above technique is suitable for dimensionality reduction if data is normally distributed with equal variance. Justify your answer. **(10)**

OR

Q.4 Enlist the assumptions of multivariate regression model and with suitable example show how to validate model. **(10)**

Q.5 The company produces raw material length and width of raw material is given into the table. The 'Acceptance' column note the acceptance status of the user. Apply principle component analysis to reduce 'Acceptance' dependency on two columns. **(10)**

Length	12.95	12.53	13.57	13.16	12.58	12.16
Width	16.63	17.79	15.65	15.47	14.46	16.22
Acceptance	Accept	Accept	Accept	Accept	Reject	Reject

OR

Q.5 Describe factor analysis model and write the process to determine number of factors using factor analysis. **(10)**

Q.6 Enlist the different types of clustering algorithms with stepwise explanation apply K-means algorithm on following dataset to form two clusters for given data : **(10)**

Height	185	188	180	180	179	182	170	182
weight	72	77	70	88	68	72	56	72

OR

Q.6 **a)** Illustrate correlation and distance. **(10)**

b) Write note on Cluster Profiling.

* * * * *