B. TECH. (COMPUTER SCIENCE & BUSINESS SYSTEMS) (CBCS - 2018 COURSE)
Computer Science & Business Systems
B.Tech. (CSBS) Sem - III :SUMMER : 2023
SUBJECT : COMPUTATIONAL STATISTICS

Day : Monday
Time : 02:30 PM-05:30 PM

Date : 15-05-2023
S-20448-2023
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) Illustrate the different types of control flow statements available in Python with flowchart. **(05)**

b) Discuss the following method associated with file object **(05)**
i) read () ii) readline () iii) readlines ()
iv) tell () v) seek () vi) write ()

OR

Q.1 a) Write python program to find the longest word in a text file. **(05)**

b) Bring out difference between multiple and multilevel inheritance. **(05)**

Q.2 Consider following dataset of "citywise social medial usage" perform following operation on it. **(10)**

Time	Pune			Mumbai			Delhi		
00:00-04:00	1333	1149	1593	2373	888	360	1789	1014	2333
04.00-08:00	1423	885	951	654	756	824	1750	1369	1753
08.00-12:00	1014	963	846	753	789	959	1959	1885	1951
2:00-16:00	6708	124	27	881	741	750	1789	1963	1846
16:00-20:00	8070	149	369	759	751	850	1654	1124	1027
20:00-00:00	3335	753	852	425	853	1051	1425	1149	1369

i) Using groupby function project city wise rate of app usage.
ii) Visualize given dataset using 4 possible charts of Matplotlib library.
iii) Find max usage rate in each city.
iv) Demonstrate transform () and apply () function.

OR

Q.2 a) Demonstrate how to split above given dataset of "citywise social media usage" by applying groupby function on any column. Find out mean, minimum and maximum values for that column. **(05)**

b) Give the significance of pivot table and crosstab functions and demonstrate them using suitable dataset. **(05)**

Q.3 Enlist multiple linear regression model assumptions. Using suitable dataset demonstrate now to validate model assumptions. **(10)**

OR

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

P.T.O.

Q.4 Write steps to design one way and two way ANOVA model and explain how to test the hypothesis. **(10)**

OR

Q.4 a) Calculate the covariance and correlation matrix for **(05)**

Height	172	180	183	190	160	188	179	156
Weight	83	95	77	89	72	99	86	53

b) What is discriminant analysis? With suitable example explain when and how to use linear discriminant analysis. **(05)**

Q.5 Compare principle component analysis (PCA) with factor analysis (FA) and stepwise apply PCA on following dataset to reduce the dependancy of 'Result' on 'Theory Marks' and 'Practical Marks'. **(10)**

Theory Marks	58	86	47	39	29	14	07	43
Practical Marks	20	18	33	43	49	23	29	33
Result	Pass	Fail	Pass	Pass	Pass	Fail	Fail	Pass

OR

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

Q.6 What is hierarchical clustering? With stepwise explaination apply K-means algorithm on following dataset to form three clusters. **(10)**
A_1 (2, 10), A_2 (2, 5), A_3 (8, 4), A_4 (5, 8), A_5 (7, 5), A_6 (6, 4), A_7 (1, 2), A_8 (4, 9)

OR

Q.6 a) Illustrate correlation and distance. **(05)**

b) Write note on cluster profiling. **(05)**

* * * * *