22DAU209N

PSG COLLEGE OF ARTS & SCIENCE
(AUTONOMOUS)

BSc DEGREE EXAMINATION DECEMBER 2025
(Second Semester)

Branch – **COMPUTER SCIENCE WITH DATA ANALYICS**

STATISTICAL DATA ANALYSIS

Time: Three Hours Maximum: 75 Marks

SECTION-A (10 Marks)
Answer **ALL** questions
ALL questions carry **EQUAL** marks (10 × 1 = 10)

Module No.	Question No.	Question	K Level	CO
1	1	Which of the following is a measure of central tendency? a) Range b) Standard Deviation c) Mean d) Coefficient of Variation	K1	CO1
	2	In statistics, qualitative data refers to a) Data that can be measured in numbers b) Data represented in graphs only c) Non-numeric data that describes categories or qualities d) Data used only for calculating standard deviation	K2	CO1
2	3	Which of the following correlation methods is used when the data is in ranks (ordinal form)? a) Karl Pearson's coefficient of correlation b) Spearman's rank correlation coefficient c) Scatter diagram method d) Regression analysis	K1	CO2
	4	Which of the following statements is true about regression analysis? a) It only shows the strength of association between two variables b) It determines the cause of variation in one variable due to another c) It is used only for qualitative data d) Regression equations are not used for prediction	K2	CO2
3	5	If P(A)=0.4, P(B)=0.5P(A) = 0.4, P(B) = 0.5, P(A∩B) =0.2 then the probability of AUB is a) 0.7 b) 0.6 c) 0.9 d) 1.0	K1	CO3
	6	In a Poisson distribution, the mean and variance are a) Equal b) Mean > Variance c) Variance > Mean d) Cannot be compared	K2	CO3
4	7	Which of the following is the first step in the general procedure for testing the significance of a hypothesis? a) Compute the test statistic b) Set up the null and alternative hypothesis c) Fix the level of significance d) Define the critical region	K1	CO4
	8	A t-test is generally used when a) Sample size is large (n>30) b) Population standard deviation is known c) Sample size is small and population standard deviation is unknown d) Both (a) and (b)	K2	CO4
5	9	The F-test is primarily used to a) Test equality of two population means b) Test equality of two population variances c) Test independence of attributes d) Test normality of data	K1	CO5
	10	The Chi-square test of independence is applied to data in the form of a) Raw scores b) Frequencies arranged in a contingency table c) Sample means d) Standard deviations	K2	CO5

Cont...

SECTION - B (35 Marks)

Answer **ALL** questions

ALL questions carry **EQUAL** Marks (5 × 7 = 35)

Module No.	Question No.	Question	K Level	CO
1	11.a.	Explain the uses of statistics in data analytics with suitable examples.	K2	CO1
		(OR)		
	11.b.	Draw a simple bar diagram for the strength of students in fou programmes in a college. Class: BA BBA BSc BCom No of Students: 600 400 1200 1500		
2	12.a.	Classify the scatter diagram indicating different types of correlation.	K4	CO2
		(OR)		
	12.b.	Distinguish between Correlation & Regression.		
3	13.a.	A can solve 80% of the problems, while B can solve 90% of problems in a statistics book. A problem is selected at random. What is the probability that at least one of them will solve it?	K3	CO3
		(OR)		
	13.b.	Explain the properties of Normal distribution.		
4	14.a.	Examine the level of significance for a test.	K4	CO4
		(OR)		
	14.b.	A car manufacturer claims that its new car gives a milage of at least 10 kms.per litre of petrol. A sample of 10 cars is taken, and their mileage recorded as follows : 11.2, 10.7, 11.3, 11.0, 10.8, 10.7, 10.6, 10.6, 10.7, 10.4 . Is there any statistical evidence to support the claim of the manufacturer about the milage of its car?		
5	15.a.	Write the testing procedure of one way ANOVA.	K3	CO5
		(OR)		
	15.b.	A chain of departmental stores opened three stores in Mumbai. The management wants to compare the sales of three stores over a six day long promotional period. Store A : 16, 17, 21, 18, 19, 29 Store B : 20, 20, 21, 22, 25, 28 Store C : 23, 24, 26, 27, 29, 30 Use Kruskal – wallis test to compare the equality of mean sales in all the three stores.		

SECTION -C (30 Marks)

Answer **ANY THREE** questions

ALL questions carry **EQUAL** Marks (3 × 10 = 30)

Module No.	Question No.	Question	K Level	CO
1	16	For the following data, Apply mean,median and mode Marks: 11-20 21-20 31-40 41-50 51-60 61-70 71-80 No of Students: 42 58 120 84 48 36 31	K3	CO1
2	17	The following data gives sales and net profit for some of the top auto makers during the quarter July- September 2006, Find out the correlation coefficient. Company : A, B, C,D,E,F,G,H SALES : 65, 22,24,10,5,16, 24, 34 PROFIT : 47, 22, 34.5, 3.5, 6.0, 9.0, 20.0, 32.0	K3	CO2
3	18	An investment consultant predicts that the odd against the price of a certain stock going up are 2:1 , and odds in favour of the price remaining the same are 1:3, What is the probability that the stock will go down?	K4	CO3
4	19	Examine the steps for conducting tests of significance for two means for small sample.	K4	CO4
5	20	The following table gives the number of good and bad parts produced by each of three shifts in a factory.	K4	CO5

Shifts	Good	Bad	Total
Day	900	130	1030
Evening	700	170	870
Night	400	200	600
Total	2000	500	2500

Is there any association between the shift and the equality of parts produced? Use Chi-Square test.

Z-Z-Z END