

PSG COLLEGE OF ARTS & SCIENCE
(AUTONOMOUS)

BSc DEGREE EXAMINATION DECEMBER 2024
(Fourth Semester)

Branch – INFORMATION TECHNOLOGY

STATISTICS FOR INFORMATION TECHNOLOGY

Time: Three Hours

Maximum: 50 Marks

SECTION-A (5 Marks)

Answer ALL questions

ALL questions carry EQUAL marks

(5 x 1 = 5)

- 1 In systematic sampling, what does "n" represent?
 - (i)The size of the population
 - (ii)The number of sampling units to be selected
 - (iii)The interval size for selecting sampling units
 - (iv)The standard deviation of the population
- 2 What is the purpose of a test of significance in statistics?
 - (i)To summarize the data
 - (ii)To determine the population mean
 - (iii)To assess whether the observed difference between groups is due to chance
 - (iv)To calculate the standard deviation
- 3 What is the F-test used for?
 - (i)Comparing means of two independent groups
 - (ii)Comparing variances of two independent groups
 - (iii)Comparing proportions of two independent groups
 - (iv)Comparing means of more than two independent groups
- 4 In one-way ANOVA, if the null hypothesis is rejected, what can be concluded about the means of the groups?
 - (i)At least one group mean is different from the others
 - (ii)All group means are equal
 - (iii)No conclusion can be drawn about the group means
 - (iv)The means of two specific groups are different
- 5 When should the Mann-Whitney U test be used?
 - (i)To compare means of two independent groups with normal distributions
 - (ii)To compare medians of two independent groups with normal distributions
 - (iii)To compare means of two independent groups regardless of distribution
 - (iv)To compare variances of two independent groups regardless of distribution

Cont...

SECTION - B (15 Marks)Answer **ALL** Questions**ALL** Questions Carry **EQUAL** Marks

(5 x 3 = 15)

- 6 a State and prove sample mean $\bar{y}$ is an unbiased estimate of $\bar{Y}$.
OR
- b Prove that in stratified random sampling $V(\bar{y}_{st}) = \frac{1}{N^2} \sum_{i=1}^k N_i (N_i - n_i) \frac{S_i^2}{n_i}$.
- 7 a Discuss the role of the null hypothesis in hypothesis testing.
OR
- b Describe the procedure for difference of proportions in large samples.
- 8 a Describe the assumptions underlying the Student's t-test.
OR
- b A company claims that the mean delivery time for their online orders is 3 days. To test this claim, a random sample of 20 recent orders was selected, and the delivery times (in days) were recorded as follows:
3, 2, 4, 3, 3, 3, 2, 3, 4, 3, 4, 3, 3, 2, 3, 3, 3, 4, 2, 3
Perform a one-sample t-test at a significance level of 0.05 to determine whether there is evidence to support the company's claim. [At $\alpha = 0.05$ (two-tailed test) and $df = 19$, the critical t-value is 2.093]
- 9 a Explain the concept of one-way ANOVA.
OR
- b Describe the applications of the chi-square test.
- 10 a Explain the concept of a run test in statistical analysis.
OR
- b A student claims that they have improved their score in a math test after attending extra tutoring sessions. The student took two tests, before and after the tutoring sessions. The scores (out of 100) are as follows:
Before tutoring: 65, 70, 60, 68, 72
After tutoring: 75, 78, 70, 80, 82
Using the sign test at a significance level of 0.05, determine if there is a significant improvement in the student's scores after attending the tutoring sessions.

SECTION - C (30 Marks)Answer **ALL** questions**ALL** questions carry **EQUAL** Marks

(5 x 6 = 30)

- 11 a Prove that under SRSWOR $v(\bar{y}) = (1-f) \frac{s^2}{n}$.
OR
- b Describe the sampling and non-sampling errors.
- 12 a A random sample of 500 apples was taken from a large consignment and 60 were found to be bad. Obtain the 98% confidence limits for the percentage number of bad apples in the consignment.
OR
- b The mean yield of wheat from a district A was 210 pounds with S.D. 10 pounds per acre from a sample of 100 plots. In another district the mean yield was 220 pounds with S.D. 12 pounds from a sample of 150 plots. Assuming that the S.D. 12 pounds from a sample of 150 plots. Assuming that the S.D of yield of in the entire state was 11 pounds, test whether there is any significant difference between the mean yield of crops in the two districts.

Cont...

- 13 a To assess whether there's a significant difference in the average test scores between two groups of students, Group A and Group B, a random sample of 25 students from each group was selected, and their test scores were recorded as follows:

Group A: 75, 82, 68, 90, 88, 72, 79, 85, 91, 78, 83, 76, 84, 81, 87, 80,
77, 73, 89, 86, 74, 79, 84, 82, 75

Group B: 78, 85, 69, 92, 87, 75, 80, 86, 90, 79, 84, 77, 85, 82, 88, 81, 78,
76, 89, 83, 79, 85, 90, 88, 79

Perform a test of significance at a significance level of 0.05 to determine if there is a significant difference in the average test scores between the two groups. [At $\alpha = 0.05$ (two-tailed test) and $df = 48$, The critical t-value is 2.013]

OR

- b A company produces two types of batteries, A and B. The company wants to determine if there is a significant difference in the lifespans of the batteries produced by two different manufacturing processes. A random sample of 10 batteries of type A and 12 batteries of type B was selected, and their lifespans (in hours) were recorded as follows:

Type A: 200, 210, 220, 190, 215, 225, 205, 195, 200, 190

Type B: 220, 225, 235, 230, 215, 240, 225, 230, 235, 245, 220, 225

Perform an F-test at a significance level of 0.05 to determine if there is a significant difference in the lifespans between the two types of batteries.

[Given $\alpha = 0.05$ and $df_A = 9$, $df_B = 11$, the critical F-value = 3.49]

- 14 a To investigate whether there's a significant difference in the average scores among three different study groups (Group A, Group B, and Group C), 15 students are randomly assigned to each group, and their test scores after completing a study program are recorded as follows:

Group A: 75, 80, 85, 78, 82, 79, 83, 77, 81, 76, 84, 79, 80, 82, 76

Group B: 78, 82, 85, 80, 83, 79, 81, 77, 79, 75, 84, 78, 80, 82, 77

Group C: 80, 84, 87, 82, 85, 81, 83, 79, 82, 78, 86, 80, 81, 83, 77

Perform a one-way ANOVA to determine if there is a significant difference in the average test scores among the three study groups at a significance level of 0.05. [Given $\alpha = 0.05$ and $df_B = 2$, $df_W = 42$, the critical F-value is approximately $F(0.05, 2, 42) = 3.29$]

OR

- b On the basis of information given below about the treatment of 200 patients suffering from a disease, state whether the new treatment is comparatively superior to the conventional treatment.

	Favorable	Not Favorable	Total
New	60	30	90
Conventional	40	70	110

[Tabulated $\chi^2_{0.05} = 3.841$ for 1 d.f.]

- 15 a Two groups of students, Group A and Group B, were taught using different teaching methods. The exam scores (out of 100) for each group are as follows:

Method A: 78, 82, 85, 79, 83, 80

Method B: 75, 78, 83, 85, 79, 81

Using the Mann-Whitney U test at a significance level of 0.05, determine if there is a significant difference in the exam scores between the two teaching methods.

OR

- b Three groups of students (Groups A, B, and C) were assessed on their performance in a math test. The scores of five students in each group are as follows:

Group A: 72, 78, 80, 85, 88

Group B: 75, 80, 82, 86, 90

Group C: 70, 75, 78, 82, 85

Using the Kruskal-Wallis test at a significance level of 0.05, determine if there is a significant difference in the test scores among the three groups.

[Value of χ^2 at 5% level of significance for 2 degrees of freedom is 5.99]

Z-Z-Z END

PSG COLLEGE OF ARTS & SCIENCE
(AUTONOMOUS)

BSc DEGREE EXAMINATION DECEMBER 2024
(First Semester)

Branch – **INFORMATION TECHNOLOGY**

STATISTICS FOR INFORMATION TECHNOLOGY-I

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	The value which divides the frequency distribution in to two equal parts is --- a. Mean b. Median c. Mode d. Range	K1	CO1
	2	If the unit in which the data is expressed is cm, the unit of coefficient of range is --- a. Sq.Cm b.cm c. metre d. No unit	K2	CO1
2	3	If the Karl Pearson's coefficient of skewness is 1 for a distribution whose mean is 4 and mode is 3.5, what is the standard deviation of the distribution? a. 0.4 b. 0.5 c. 1.5 d. 1	K1	CO2
	4	What is the range of the coefficient of correlation? a. $r \geq 0$ b. $r \leq 1$ c. $r \leq -1$ d. $-1 \leq r \leq 1$	K2	CO3
3	5	How many terms are there in the binomial expansion $(x + y)^n$? a. $n+1$ b. n c. $n-1$ d. $2n$	K1	CO3
	6	If the values of the independent variable are not equally spaced, which of the following methods is used for interpolation? a. Newton's forward difference method b. Newton's backward difference method c. Lagrange's method d. Binomial method	K2	CO3
4	7	What is minimized in the method of least squares? a. The difference between actual value and estimated value b. The sum of differences between actual and estimated values c. The sum of squares of the differences between actual and estimated values d. The sum of squares of actual values.	K1	CO4
	8	Which method is used to calculate the trend values in Ratio to trend method? a. Method of moving averages of period 2 b. Method of moving averages of period 3 c. Method of semi averages d. Method of least squares	K2	CO4
5	9	Which of the following Excel function is used to find the arithmetic mean of a number of items? a. Mean() b. Arithmetic mean() c. Sum() d. Average()	K1	CO5
	10	STDEVP(x) provides a. Sample standard deviation b. Population standard deviation c. Mean squared deviation d. Root mean squared deviation	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.	Represent the data in a pie diagram	K4	CO1																						
		<table><tr><td>Year</td><td>2010</td><td>2011</td><td>2012</td><td>2013</td><td>2014</td></tr><tr><td>Production in Branch A</td><td>1200</td><td>1600</td><td>1600</td><td>1300</td><td>1500</td></tr></table>			Year	2010	2011	2012	2013	2014	Production in Branch A	1200	1600	1600	1300	1500										
		Year			2010	2011	2012	2013	2014																	
	Production in Branch A	1200			1600	1600	1300	1500																		
(OR)																										
11.b.	Calculate the Median.																									
	<table><tr><td>Height(cm)</td><td>145-150</td><td>150-155</td><td>155-160</td><td>160-165</td><td>165-170</td><td>170-175</td></tr><tr><td>Frequency</td><td>2</td><td>5</td><td>10</td><td>8</td><td>4</td><td>1</td></tr></table>	Height(cm)	145-150	150-155	155-160	160-165	165-170	170-175	Frequency	2	5	10	8	4	1											
	Height(cm)	145-150	150-155	155-160	160-165	165-170	170-175																			
Frequency	2	5	10	8	4	1																				
2	12.a.	Calculate the Karl Pearson's coefficient of skewness and interpret your result for the following information 25 15 23 40 27 25 23 25 20	K5	CO2																						
		(OR)																								
	12.b.	Compute Spearman's rank correlation coefficient .																								
		<table><tr><td>X:</td><td>10</td><td>8</td><td>20</td><td>15</td><td>12</td><td>16</td><td>14</td></tr><tr><td>Y:</td><td>12</td><td>19</td><td>22</td><td>21</td><td>17</td><td>11</td><td>18</td></tr></table>			X:	10	8	20	15	12	16	14	Y:	12	19	22	21	17	11	18						
X:	10	8	20	15	12	16	14																			
Y:	12	19	22	21	17	11	18																			
3	13.a.	Find the coefficient of x^{11} in the expansion $\left(x^3 - \frac{2}{x^2}\right)^{12}$	K3	CO3																						
		(OR)																								
	13.b.	Find the value of Y at X=0 given the set of values (-2,5), (1,7), (3, 11) and (7,34) using Lagranges method.																								
4	14.a.	Calculate the three yearly moving averages:	K4	CO4																						
		<table><tr><td>Year</td><td>1983</td><td>1984</td><td>1985</td><td>1986</td><td>1987</td><td>1988</td><td>1989</td><td>1990</td><td>1991</td><td>1992</td></tr><tr><td>Value</td><td>21</td><td>22</td><td>23</td><td>25</td><td>24</td><td>22</td><td>25</td><td>26</td><td>27</td><td>26</td></tr></table>			Year	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	Value	21	22	23	25	24	22	25	26	27	26
		Year			1983	1984	1985	1986	1987	1988	1989	1990	1991	1992												
	Value	21			22	23	25	24	22	25	26	27	26													
(OR)																										
14.b.	Fit a straight line trend equation by the method of least squares and estimate the trend value for the year 1985.																									
	<table><tr><td>Year</td><td>1979</td><td>1980</td><td>1981</td><td>1982</td><td>1983</td></tr><tr><td>Sales in Rs</td><td>100</td><td>120</td><td>140</td><td>160</td><td>180</td></tr></table>	Year	1979	1980	1981	1982	1983	Sales in Rs	100	120	140	160	180													
	Year	1979	1980	1981	1982	1983																				
Sales in Rs	100	120	140	160	180																					
5	15.a.	Explain the steps of entering 11 observations of rainfall in an excel sheet and computing the mean using statistical function.	K3	CO5																						
		(OR)																								
	15.b.	Write the procedure to compute standard deviation of 13 observations in Excel.																								

Cont..

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	From the following quantity of a product sold in two cities on various days of a week, identify the city in which the prices are more stable. <table><tr><td>Day</td><td>Mon</td><td>Tue</td><td>Wed</td><td>Thu</td><td>Fri</td><td>Sat</td></tr><tr><td>City A</td><td>498</td><td>500</td><td>505</td><td>504</td><td>502</td><td>509</td></tr><tr><td>City B</td><td>500</td><td>505</td><td>502</td><td>498</td><td>496</td><td>505</td></tr></table>	Day	Mon	Tue	Wed	Thu	Fri	Sat	City A	498	500	505	504	502	509	City B	500	505	502	498	496	505	K4	CO1									
Day	Mon	Tue	Wed	Thu	Fri	Sat																												
City A	498	500	505	504	502	509																												
City B	500	505	502	498	496	505																												
2	17	Explain the procedure of determining two regression equations with a suitable example.	K3	CO2																														
3	18	Estimate the number of students who obtained marks 45 using Newton's forward interpolation method, <table><tr><td>Marks</td><td>30</td><td>40</td><td>50</td><td>60</td><td>70</td></tr><tr><td>No.of students</td><td>31</td><td>42</td><td>51</td><td>35</td><td>31</td></tr></table>	Marks	30	40	50	60	70	No.of students	31	42	51	35	31	K3	CO3																		
Marks	30	40	50	60	70																													
No.of students	31	42	51	35	31																													
4	19	Determine the seasonal indices using ratio to moving average method. <table><tr><td>Year</td><td>IQ</td><td>IIQ</td><td>IIIQ</td><td>IVQ</td></tr><tr><td>1990</td><td>40</td><td>35</td><td>38</td><td>40</td></tr><tr><td>1991</td><td>42</td><td>37</td><td>39</td><td>38</td></tr><tr><td>1992</td><td>41</td><td>35</td><td>38</td><td>42</td></tr><tr><td>1993</td><td>45</td><td>36</td><td>36</td><td>41</td></tr><tr><td>1994</td><td>44</td><td>38</td><td>38</td><td>42</td></tr></table>	Year	IQ	IIQ	IIIQ	IVQ	1990	40	35	38	40	1991	42	37	39	38	1992	41	35	38	42	1993	45	36	36	41	1994	44	38	38	42	K4	CO4
Year	IQ	IIQ	IIIQ	IVQ																														
1990	40	35	38	40																														
1991	42	37	39	38																														
1992	41	35	38	42																														
1993	45	36	36	41																														
1994	44	38	38	42																														
5	20	Write the procedure to compute Mode of 120 observations in Excel.	K3	CO5																														

Z-Z-Z END