

Unique Paper Code : 62273506
 Name of the Course : BA Prog.(CBCS)
 Semester : Semester V
 Paper : Data Analysis
 UPC : 62273506
 Duration : 3 hours
 Maximum Marks : 65 Marks

Instructions for Candidates

1. Write your roll No. on the top immediately on receipt of this question paper.
2. All questions carry equal marks (16.25 marks each).
3. Attempt any **four** questions.
4. Note: Answers may be written either in English or in Hindi; but the same medium should be used throughout the paper.
5. Use of simple calculator is allowed.

1. (i) The following table provides the sources of electricity used by households in a city.

Source of Electricity	(1) Percentage
Coal	39%
Hydro and Renewables	13%
Natural Gas	27%
Nuclear Pwer	19%
Other	2%

- (a) Construct a pie chart.
- (b) What percentage of power is derived from coal, nuclear power, or natural gas?
- (c) For above data, do you prefer using a pareto chart or a pie chart? Why?

10

(ii) “A coefficient estimate is statistically significant at the 5% significance level”. Explain.

3

(iii) Differentiate between the scatter plot and the time series plot.

3.25

1. (i) निम्नलिखित तालिका एक शहर में घरों द्वारा उपयोग किए जाने वाले बिजली के स्रोत प्रदान करती है।

Source of Electricity	(1) Percentage
Coal	39%
Hydro and Renewables	13%
Natural Gas	27%
Nuclear Pwer	19%
Other	2%

(अ) एक पाई चार्ट बनाएं।

(ब) कोयले, परमाणु ऊर्जा, या प्राकृतिक गैस से कितने प्रतिशत बिजली प्राप्त होती है?

(स) उपरोक्त डेटा के लिए, क्या आप पारेतो चार्ट या पाई चार्ट का उपयोग करना पसंद करते हैं? क्यों?

10

(ii) "एक गुणांक अनुमान 5% महत्व स्तर पर सांख्यिकीय रूप से महत्वपूर्ण है"। समझाना।

3

(iii) स्कैटर प्लॉट और टाइम सीरीज़ प्लॉट के बीच अंतर करें।

3.25

2. (i) A survey of 1,700 shoppers were asked, "Does social media influence your purchase decisions?" The results indicated that 20.3% of females responded that their purchase decisions are influenced by social media compared to 10.5% of males. The sample sizes of males and females were not provided. Suppose that the results were as shown in the following table:

Gender			
Influenced	Male	Female	Total
Yes	89	173	262
No	761	677	1438
Total	850	850	1700

(a) Construct contingency tables based on total percentages, row percentages, and column percentages.

(b) Which gender enjoys more doing shopping?

10

(ii) What is the difference between simple random sampling and systematic random sampling?

6.25

2. (i) 1,700 दुकानदारों के एक सर्वेक्षण में पूछा गया, "क्या सोशल मीडिया आपके खरीद निर्णयों को प्रभावित करता है?" परिणामों ने संकेत दिया कि 20.3% महिलाओं ने जवाब दिया कि उनके खरीद निर्णय 10.5% पुरुषों की तुलना में सोशल मीडिया से प्रभावित हैं। पुरुषों और महिलाओं के नमूना आकार प्रदान नहीं किए गए थे। मान लीजिए कि परिणाम निम्न तालिका में दिखाए गए थे:

Gender			
Influenced	Male	Female	Total
Yes	89	173	262
No	761	677	1438
Total	850	850	1700

(अ) कुल प्रतिशत, पंक्ति प्रतिशत और कॉलम प्रतिशत के आधार पर आकस्मिक तालिकाओं का निर्माण करें।

(ब) किस वर्ग (महिला एवं पुरुष) को खरीदारी करने में अधिक आनंद आता है?

10

(ii) साधारण यादृच्छिक प्रतिचयन और व्यवस्थित यादृच्छिक प्रतिचयन में क्या अंतर है?

6.25

3. (i) The following is a set of data from a sample of $n = 11$ items:

X	Y
7	21
5	15
8	24
3	9
6	18
10	30
12	36
4	12
9	27
15	45
18	54

(a) Compute the covariance.

(b) Compute the coefficient of correlation.

(c) How strong is the relationship between X and Y ? Explain.

10

(ii) "Each average has its own special features and it is difficult to say which one is best". Explain and illustrate.

6.25

3. (i) निम्नलिखित $n = 11$ वस्तुओं के नमूने से डेटा का एक सेट है:

X	Y
7	21
5	15
8	24
3	9
6	18
10	30
12	36
4	12
9	27
15	45
18	54

(अ) सहप्रसरण की गणना करें।

(ब) सहसंबंध के गुणांक की गणना करें।

(स) X और Y के बीच संबंध कितना मजबूत है? समझाइये । 10

(ii) "प्रत्येक औसत की अपनी विशेष विशेषताएं होती हैं और यह कहना मुश्किल है कि कौन सा सबसे अच्छा है"। उदाहरण सहित समझाइये । 6.25

4. (i) The following table provides the data about the cost of electricity during July 2015 for a random sample of 50 one-bedroom apartments.

96	171	202	178	147	102	153	197	127
82	157	185	90	116	172	111	148	213
130	165	141	149	206	175	123	128	144
168	109	167	95	163	150	154	130	143
187	166	139	149	108	119	183	151	114
135	191	137	129	158				

(a) Construct a frequency distribution and a percentage distribution that have class intervals with the upper class boundaries 99, 119, and so on.

(b) Construct a cumulative percentage distribution. 10

(ii) Explain the problems of using a t-test as a hypothesis testing approach.

6.25

4. (i) निम्न तालिका जुलाई 2015 के दौरान 50 एक-बेडरूम अपार्टमेंट के यादृच्छिक नमूने के लिए बिजली की लागत के बारे में डेटा प्रदान करती है।

96	171	202	178	147	102	153	197	127
82	157	185	90	116	172	111	148	213
130	165	141	149	206	175	123	128	144
168	109	167	95	163	150	154	130	143
187	166	139	149	108	119	183	151	114
135	191	137	129	158				

(अ) एक आवृत्ति वितरण और एक प्रतिशत वितरण का निर्माण करें जिसमें उच्च वर्ग की सीमाओं के साथ वर्ग अंतराल 99, 119, और इसी तरह है।

(ब) संचयी प्रतिशत वितरण का निर्माण करें।

10

(ii) परीक्षण परिकल्पना के t परीक्षण दृष्टिकोण समस्या की व्याख्या करें। 6.25

5. (i) The following set of data is from a sample of $n = 7$:

12	7	4	9	0	7	3
----	---	---	---	---	---	---

(a) Compute the mean, median, and mode using either R/Excel .

(b) Compute the range, variance, standard deviation, and coefficient of variation using either Excel/R.

(c) Compute the Z scores. Are there any outliers?

(d) Describe the shape of the data set.

10

(ii) How would you interpret the sign and magnitude of a calculated 'r'? Consider in particular the values $r=0$, $r=-1$ and $r=+1$. **6.25**

5. (i) डेटा का निम्नलिखित सेट $n = 7$ के नमूने से है:

12	7	4	9	0	7	3
----	---	---	---	---	---	---

(अ) R / Excel का उपयोग करके माध्य, माध्यिका और बहुलक की गणना करें।

(ब) R / Excel का उपयोग करके सीमा, भिन्नता, मानक विचलन, और भिन्नता के गुणांक की गणना करें।

(स) Z स्कोर की गणना करें। क्या कोई आउटलेयर हैं?

(ड) डेटा सेट के आकार का वर्णन करें। **10**

(ii) आप परिकलित 'r' के चिह्न और परिमाण की व्याख्या कैसे करेंगे? विशेष रूप से मानों $r=0$, $r=-1$ और $r=+1$ पर विचार करें। **6.25**

6. Write short notes on any two of the following:

(a) Boxplot

(b) Central Limit Theorem

(c) Primary and secondary data **16.25**

6. निम्नलिखित में से किन्हीं दो पर संक्षिप्त टिप्पणी लिखिए:

(ए) बॉक्सप्लॉट

(बी) केंद्रीय सीमा प्रमेय

(सी) प्राथमिक और माध्यमिक डेटा **16.25**