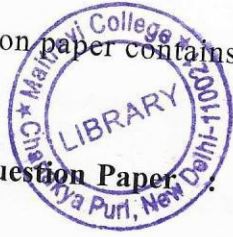


[This question paper contains 12 printed pages.]



21/12/2023
Evening
Your Roll No.....
G

Sr. No. of Question Paper : 5728

Unique Paper Code : 62273506

Name of the Paper : Data Analysis (Skill Enhancement Course)

Name of the Course : B.A. Programme

Semester : V

Duration : 3 Hours

Maximum Marks : 65

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. All questions carry equal marks (13 marks each).
3. Attempt any five questions.
4. Use of simple calculator is allowed.
5. Answers may be written either in English or Hindi; but the same medium should be used throughout the paper.

छात्रों के लिए निर्देश

1. इस प्रश्न-पत्र के मिलते ही ऊपर दिए गए निर्धारित स्थान पर अपना अनुक्रमांक लिखिए।
2. सभी प्रश्नों के अंक समान हैं। (प्रत्येक 13 अंक)
3. किन्हीं पाँच प्रश्नों का उत्तर दीजिए।
4. साधारण कैलकुलेटर के उपयोग की अनुमति है।
5. इस प्रश्न-पत्र का उत्तर अंग्रेजी या हिंदी किसी एक भाषा में दीजिए, लेकिन सभी उत्तरों का माध्यम एक ही होना चाहिए।

P.T.O.

1. (a) Distinguish between simple random sampling and systematic sampling?

(b) A study of selected Kickstarter Projects showed that overall majorities were successful, and some were non-Successful. In an effort to identify project types that influence success, selected project was subdivided into project categories given below. The results are as follows :

Outcome			
Category	Successful	Non Successful	Total
Film & Video	17,024	26,857	43,881
Games	5,105	10,075	15,180
Music	19,275	17,050	36,325
Technology	2,556	8,770	11,326
Total	43,960	62,752	106,712

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

(ii) Which type of percentage—row, column, or total - do you think is the most informative for these data? Explain. (5,8)

(क) सरल यादृच्छिक नमूनाकरण और व्यवस्थित नमूनाकरण के बीच अंतर स्पष्ट कीजिए?

(ख) चयनित किक स्टार्टर प्रोजेक्ट्स के एक अध्ययन से पता चला है कि समग्र बहुमत सफल थे, और कुछ असफल रहे थे। सफलता को प्रभावित करने वाले परियोजना प्रकारों की पहचान करने के प्रयास में, चयनित परियोजना को नीचे दी गई परियोजना श्रेणियों में विभाजित किया गया था। परिणाम क्षेत्र निम्नानुसार हैं :

परिणाम			
श्रेणी	सफल	असफल	कुल
फिल्म और वीडियो	17,024	26,857	43,881
गेम्स	5,105	10,075	15,180
संगीत	19,275	17,050	36,325
टेक्नोलॉजी	2,556	8,770	11,326
कुल	43,960	62,752	106,712

(i) कुल प्रतिशत, पंक्ति प्रतिशत और कॉलम प्रतिशत के आधार पर एक आकस्मिक तालिका तैयार कीजिए।

(ii) इन आंकड़ों के लिए किस प्रकार का प्रतिशत- पंक्ति, कॉलम, या कुल-सबसे अधिक जानकारी देने वाला है? स्पष्ट कीजिए।

2. (a) Construct a stem-and-leaf display on the basis of the following data from a sample of midterm exam scores in finance :

54, 69, 45, 98, 67, 93, 53, 74, 88, 62, 33, 21, 9, 88

- (b) The following data represents the complaints received by automaker.

Automaker	Number
Honda	169
Chrysler	439
Ford	440
General Motors	551
Nissan	467
Toyota	332
Others	516

(i) Construct a bar chart and a pie chart for the complaints received by automaker.

(ii) What percentage of complaints is received from Honda, Ford and Nissan individually? (4,9)

(क) वित्त में मध्यावधि परीक्षा के स्कोर के नमूने से निम्नलिखित डेटा के आधार पर एक स्टेम-एंड-लीफ डिस्प्ले तैयार कीजिए :

54, 69, 45, 98, 67, 93, 53, 74, 88, 62, 33, 21, 9, 88

(ख) निम्नलिखित डेटा ऑटो निर्माता द्वारा प्राप्त शिकायतों को दर्शाता है।

ऑटो निर्माता	संख्या
होंडा	169
क्रायस्लर	439
फोर्ड	440
जनरल मोटर्स	551
निसान	467
टोयोटा	332
अन्य	516

(i) ऑटो निर्माता द्वारा प्राप्त शिकायतों के लिए एक बार चार्ट और पाई चार्ट तैयार कीजिए।

(ii) होंडा, फोर्ड और निसान को अलग-अलग कितने प्रतिशत शिकायतें प्राप्त हुई हैं?

3. (a) What do you mean by Kurtosis? Write the syntax for kurtosis in Excel/R.

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

7, 9, 4, 0, 3, 7, 12

- (i) Compute mean, median and mode.
- (ii) Compute range, variance, standard deviation and coefficient of variation.
- (iii) Compute Z scores. Are there any outliers?
- (iv) Describe shape of the data set.

(5,8)

(क) कुटोसिस से क्या तात्पर्य है? एक्सेल/आर में कुटोसिस के लिए सिंटैक्स लिखिए।

(ख) डेटा का निम्नलिखित सेट के नमूने से है $n = 7$:

7, 9, 4, 0, 3, 7, 12

- (i) माध्य, माध्यिका और मोड की गणना कीजिए।
- (ii) सीमा, विचरण, मानक विचलन और भिन्नता गुणांक की गणना कीजिए।
- (iii) Z स्कोर की गणना कीजिए। क्या कोई आउटलायर हैं?
- (iv) डेटा सेट के आकार का वर्णन कीजिए।

4. (a) Differentiate between a scatter plot and time series plot.

(b) The following set of data is from a sample of $n = 6$:

7, 4, 9, 7, 3, 12

- (i) Compute first quartile (Q1), third quartile (Q3) and interquartile range.
- (ii) List the five-number summary.
- (iii) Construct a boxplot and describe its shape. (4,9)

(क) स्कैटर प्लॉट और टाइम सीरीज प्लॉट के बीच अंतर स्पष्ट कीजिए।

(ख) डेटा का निम्नलिखित सेट के नमूने से है $n = 6$:

7, 4, 9, 7, 3, 12

- (i) प्रथम चतुर्थक (Q1), तृतीय चतुर्थक (Q3) और अंतर चतुर्थक श्रेणी की गणना कीजिए।
- (ii) पांच संख्याओं का सारांश लिखिए।
- (iii) बॉक्स प्लॉट बनाएं और इसके आकार का वर्णन कीजिए।

5. (a) What are the two major sources of collection of data?

(b) The following is a set of data from a sample of $n = 10$ items :

Temperature (X)	24	27	30	31	34	35	38	40	42	45
Units Consumed (Y)	80	82	84	101	110	115	140	142	156	157

- (i) Compute covariance.
- (ii) Compute coefficient of correlation
- (iii) How strong is the relationship between X and Y? Explain. (4,9)

(क) डेटा के संग्रह के दो प्रमुख स्रोत क्या हैं?

(ख) निम्नलिखित डेटा सेट के नमूने में से है $n = 10$ मर्दें:

तापमान (X)	24	27	30	31	34	35	38	40	42	45
खपत की गई यूनिट (Y)	80	82	84	101	110	115	140	142	156	157

(i) सहसंयोजक की गणना कीजिए।

(ii) सहसंबंध गुणांक की गणना कीजिए।

(iii) X और Y के बीच संबंध कितना मजबूत है? स्पष्ट कीजिए।

6. (a) What do you mean by Central Limit Theorem?

(b) (i) Differentiate between Null Hypotheses and Alternative Hypothesis.

(ii) Distinguish between 'one tail' test and 'two tail' test.

(iii) You are the manager of a fast-food restaurant. The business problem is to determine whether the population mean waiting time to place an order has changed in the past month from its previous value of 4.5 minutes. From past experience, you can assume that the population standard deviation is 1.2 minutes and the population waiting time is normally distributed. You select a sample of 25 orders during a one-hour period. The sample mean is 5.1 minutes. Find whether there is evidence that the population mean waiting time to place an order has changed in the past month from its previous population mean value of 4.5 minutes. (Use $\alpha = 0.05$) (2,11)

(क) केंद्रीय सीमा प्रमेय से आप क्या समझते हैं?

- (ख) (i) शून्य परिकल्पना और वैकल्पिक परिकल्पना के बीच अंतर स्पष्ट कीजिए।
- (ii) 'वनटेल' टेस्ट और 'टू टेल' टेस्ट के बीच अंतर कीजिए।
- (iii) आप एक फास्ट-फूड रेस्तरां के प्रबंधक हैं। व्यावसायिक समस्या यह निर्धारित करना है कि क्या ऑर्डर देने के लिए लोगों को प्रतीक्षा करनी पड़ती है जो समय पिछले महीने में 4.5 मिनट के अपने पिछले मूल्य से बदल गया है। पिछले अनुभव से, आप मान सकते हैं कि लोगों का मानक विचलन 1.2 मिनट है और लोगों का प्रतीक्षा समय सामान्य रूप से वितरित किया जाता है। आप एक घंटे की अवधि के दौरान 25 ऑर्डर का एक नमूना चुनें। नमूना माध्य 5.1 मिनट क्या है? पता लगाएं कि क्या इस बात के प्रमाण हैं कि ऑर्डर देने के लिए लोगों का औसत प्रतीक्षा समय पिछले महीने में 4.5 मिनट के अपने पिछले औसत मूल्य से बदल गया है। (Use $\alpha = 0.05$)

7. (a) Define Type I error and Type II error.

(b) If $\bar{X} = 66$, $\sigma = 6$, and $n = 64$, Construct a 99% confidence interval estimate for the population mean, μ .

(c) Determine the critical values of t in each of the following circumstances :

(i) $1-\alpha = 0.95$, $n = 10$

(ii) $1-\alpha = 0.99$, $n = 10$

(iii) $1-\alpha = 0.95$, $n = 32$

(iv) $1-\alpha = 0.95$, $n = 65$

(3,6,4)

(क) टाइप I त्रुटि और टाइप II त्रुटि निर्धारित कीजिए।

(ख) यदि $\bar{X} = 66$, $\sigma = 6$, और $n = 64$ जनसंख्या के लिए 99% आत्मविश्वास अंतराल अनुमान का तैयार कीजिए।

(ग) निम्नलिखित परिस्थितियों में से प्रत्येक के महत्वपूर्ण मान निर्धारित कीजिए :

(i) $1-\alpha = 0.95$, $n = 10$

(ii) $1-\alpha = 0.99$, $n = 10$

(iii) $1-\alpha = 0.95$, $n = 32$

(iv) $1-\alpha = 0.95$, $n = 65$

8. Write short notes on **any two** :

(6.5×2=13)

- (a) Pareto Chart
- (b) Confidence Interval
- (c) Geometric Mean
- (d) Skewness

किन्हीं दो पर संक्षिप्त नोट्स लिखिए :

- (क) पेटोचार्ट
- (ख) आत्मविश्वास अंतराल
- (ग) ज्यामितीय माध्य
- (घ) वैषम्य

TABLE D.2
Percentage Points of
the *t* Distribution

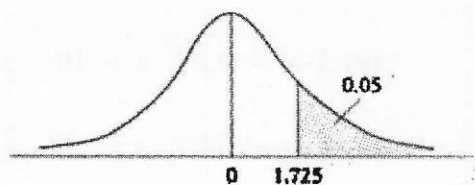
Source: From E. S. Pearson and
H. O. Hartley, eds., *Biometrika
Tables for Statisticians*, vol. 1,
3d ed., table 12, Cambridge
University Press, New York,
1966. Reproduced by
permission of the editors and
translators of *Biometrika*.

Example

$$\Pr(t > 2.086) = 0.025$$

$$\Pr(t > 1.725) = 0.05 \quad \text{for } df = 20$$

$$\Pr(|t| > 1.725) = 0.10$$



Pr df	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.05	0.01 0.02	0.005 0.010	0.001 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

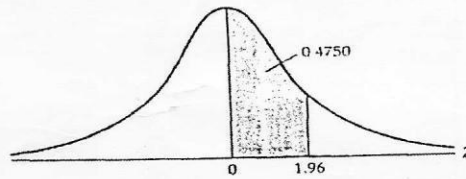
Note: The smaller probability shown at the head of each column is the area in one tail; the larger probability is the area in both tails.

TABLE A-1a AREAS UNDER THE STANDARDIZED NORMAL DISTRIBUTION.

Example

$$\Pr(0 \leq Z \leq 1.96) = 0.4750$$

$$\Pr(Z \geq 1.96) = 0.5 - 0.4750 = 0.025$$



Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Note: This table gives the area in a right-hand tail of the distribution (i.e. $Z > 0$). But since the normal distribution is symmetrical about $Z = 0$, the area in the left-hand tail is the same as the area in the corresponding right-hand tail. For example, $P(-1.96 \leq Z \leq 0) = 0.4750$. Therefore, $P(-1.96 \leq Z \leq 1.96) = 2(0.4750) = 0.95$.

P.T.O.

⇒ Entry is area A under the standard normal curve from $-\infty$ to $z(A)$



STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

Selected Percentiles

cumulative probability A:	0.90	0.95	0.975	0.98	0.99	0.995	0.999
$z(A)$:	1.282	1.645	1.960	2.054	2.32	2.576	3.090

(1000)