[This question paper contains 16 printed pages.]
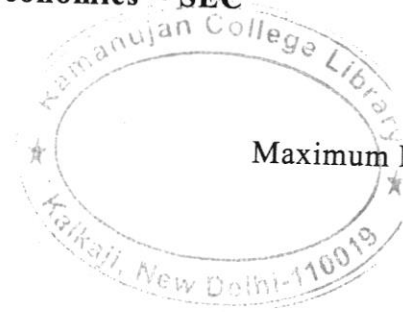
Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.

2. This question paper has **two** sections. Attempt any **TWO** questions from each section.

3. You do not require the use of R or Excel software to answer any question. Wherever asked, mention/discuss the command/function/syntax, as required in the question.

4. The questions in which R or Excel is not mentioned, the answers should be based on your own calculations.

5. Use of a simple non-programmable calculator is allowed.

6. Statistical tables are attached for your reference.

SECTION A

1. (a) The Principal at a college emailed a survey to a total of 300 students. The sample included 100 students randomly selected from each of the first year, second year, and third year of the college.

P.T.O.

(i) What type of sampling method was used?   (2)

(ii) Explain why the sampling method stated in (i) is the most efficient method.   (3)

(iii) How is the sampling method described above different from non-probability sampling? Explain.   (5)

(b) Differentiate between excel functions RAND() and RANDBETWEEN().

(3)

(c) Explain rep() command in R using an example.   (3)

2. (a) The following data represents the stress score and life satisfaction score collected from a sample of 10 participants :

Stress Score	Life satisfaction score
11	7
25	1
19	4
7	9
23	2
6	8
11	8
22	3
25	3
10	6

Using this data, answer the following questions :

(i) Compute sample correlation coefficient between the stress scores and life satisfaction scores. Comment on the correlation between the two.

(5)

(ii) Calculate the Z scores of stress scores. Are there any outliers?

(5)

(b) Explain data frame with the help of an example, in R. (3)

(c) Suppose 10 numbers are given in Excel from the cell A1 to A10. Explain the excel function to calculate mean of these 10 numbers. (3)

3. (a) Using examples, elaborate the methods used to organise categorical variables.

(5)

(b) Discuss the challenges faced in organizing and visualizing variables. (5)

(c) Explain the use of filter in Excel. (3)

(d) Write R commands for rolling a six-sided fair dice. (3)

SECTION B

4. (a) Answer the following questions :

(i) A coin is tossed at random 400 times and heads turn up 240 times. Can the coin be regarded as unbiased? Use 5% level of significance. (4)

(ii) How will the result in (i) be affected if the observed number of heads is 216? Use 5% level of significance. (2)

(b) Two random samples of sizes 8 and 11, drawn from two normal populations, are characterized as follows :

Population from which sample is drawn	Sample Size	Sum of Observations	Sum of squares of observations
I	8	9.6	61.52
II	11	16.5	73.26

You are to decide if the two populations can be taken to have the same variance. Which test function would you use? How is it distributed? Test the above hypothesis at 5% level of significance. (6)

(c) Explain the difference between getwd() and setwdQ commands in R. (4.5)

5. (a) A manufacturing company wants to estimate the average amount of purchase of its product in a month by the customers. If the standard deviation is Rs. 10, find the sample size if the maximum error is not to exceed Rs. 3 with a confidence level of 99%. How will the sample size change if the confidence level changes to 90%?

(6)

(b) The data for number of days present per week for a population of three employees, namely, Rishi, Radhika and Shanti of an organization is given below :

Employee	Number of Days
Rishi	5
Radhika	6
Shanti	4

Suppose you select samples of two administrative assistants *with* replacement from this population. Calculate the sampling distribution of sample mean. Compare it with the population mean number of days present. Are the two equal? Why or why not?

(6)

(c) Explain the use of the following R commands: read.csv and read.table.

(4.5)

6. (a) A study is conducted to see if wages of daily workers in North and South India differ from each other. Results for two-sample t-tests, assuming equal variances, for wages, are given below :

Two Sample t Test Assuming Equal Variances		
	North India	South India
Mean	12	10
Variance	16	9
Observations	4	5
Pooled Variance	12	
Hypothesized Mean Difference	0	
df	7	
t Stat	0.861	
P(T>=t) two tail	0.418	
P(T>=t) one tail	0.209	
T critical two tail at 1%	3.496	
T critical one tail at 1%	2.998	

(i) State the null and alternative hypotheses to test if the average daily wage in North India differs from average daily wage in South India.

(2)

(ii) At 1% level of significance, is there an evidence of difference in wages? Also test the hypothesis that average wages in North India are greater than those in South India.
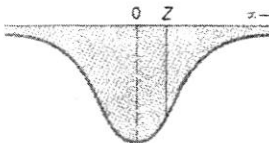
(6)

(iii) Test the hypothesis stated in part (ii) again at 5% and 10% level of

significance using p-value approach.   (2)

(b) Explain the use of Excel function : NORM.S.DIST   (3.5)

(c) Explain the use of the following R command using example: ls()   (3)

TABLE E.2
The Cumulative Standardized Normal Distribution

Entry represents area under the cumulative standardized normal distribution from − ∞ to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
Cumulative Probabilities										
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2482	0.2451
−0.7	0.2420	0.2388	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00103	0.00100
−3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
−3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
−3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
−3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
−3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
−3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
−3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
−3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
−3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
−4.0	0.0000316712									
−4.5	0.0000033977									
−5.0	0.0000002871									
−5.5	0.000000019									
−6.0	0.000000001									