ROYAL CIVIL SERVICE COMMISSION
BHUTAN CIVIL SERVICE EXAMINATION (BCSE) 2024
EXAMINATION CATEGORY: <u>TECHNICAL</u>

PAPER III: SUBJECT SPECIALISATION PAPER FOR <u>STATISTICS</u>

Date   : October 5, 2024
Total Marks   : 100
Writing Time   : 150 minutes (2.5 hours)
Reading Time   : 15 minutes (prior to writing time)

GENERAL INSTRUCTIONS:

1. Write your Registration Number clearly and correctly on the Answer Booklet.

2. The first 15 minutes is to check the number of pages of Question Paper, printing errors, clarify doubts and to read the instructions. You are NOT permitted to write during this time.

3. This paper consists of **TWO SECTIONS**, namely SECTION A & SECTION B:

    - **SECTION A** has two parts: Part I - 30 Multiple Choice Questions

                   Part II - 4 Short Answer Questions

                   All questions under SECTION A are **COMPULSORY.**

    - **SECTION B** consists of two Case Studies. Choose only **ONE** case study and answer the questions of your choice.

4. All answers should be written on the Answer Booklet provided to you. Candidates are not allowed to write anything on the question paper. If required, ask for additional Answer Booklet.

5. All answers should be written with correct numbering of Section, Part and Question Number in the Answer Booklet provided to you. Note that any answer written without indicating the Section, Part and Question Number will NOT be evaluated and no marks will be awarded.

6. Begin each Section and Part on a fresh page of the Answer Booklet.

7. You are not permitted to tear off any sheet(s) of the Answer Booklet as well as the Question Paper.

8. Use of any other paper including paper for rough work is not permitted.

9. **You must hand over the Answer Booklet to the Invigilator before leaving the examination hall.**

10. This paper has **11 printed pages**, including this instruction page.

GOOD LUCK

SECTION A

PART I: Multiple Choice Questions (30 marks)

Choose the correct answer and write down the letter of your chosen answer in the Answer Booklet against the question number e.g. 31 (d). Each question carries ONE mark. Any double writing, smudgy answers or writing more than one choice shall not be evaluated.

1. Given a dataset with a mean of 100 and a variance of 100, what is the z-score for a value of 90?
   a) 1
   b) -2
   c) -1
   d) 2

2. What is the value of variance if the sample mean is 80 and the coefficient of variation (CV) is 5%?
   a) 4
   b) 5
   c) 16
   d) 25

3. In a simple linear regression model, which of the following is true if the slope β_1 is zero?
   a) The predictor variable is perfectly correlated with the response variable.
   b) The response variable has no variation.
   c) The predictor variable has no impact on the response variable.
   d) The intercept β_0 must also be zero.

4. Suppose the average income of civil servants in Bhutan is Nu. 30,000 with a standard deviation of Nu 3,000. If all civil servants receive a 10% raise in the coming year, what will be the new mean income and standard deviation?
   a) Nu 30,000 and Nu 3,000
   b) Nu 33,000 and Nu 3,000
   c) Nu 30,000 and Nu 3,300
   d) Nu 33,000 and Nu 3,300

5. If X is normally distributed with mean 8 and standard deviation 4, what is $P(5 \leq X \leq 10)$?

   a) 0.2266
   b) 0.4649
   c) 0.6915
   d) 0.9181

6. In a hypothesis test for the difference between two means, what does a 95% confidence interval that includes zero imply about the null hypothesis?
   a) There is strong evidence against the null hypothesis.
   b) There is weak evidence against the null hypothesis.
   c) The null hypothesis should be rejected.
   d) The null hypothesis should not be rejected.

7. In a multiple regression analysis, which of the following scenarios is most likely to lead to multicollinearity?
   a) High correlation between the dependent variable and independent variables
   b) High correlation among the independent variables
   c) Inclusion of a categorical variable with many levels
   d) Use of too many dummy variables

Suppose the following data were collected from household survey. Use the information in the table to answer Questions 8 and 9.

$$let\ Y = \begin{cases} 1, & if\ the\ household\ at\ least\ one\ smart\ phone \\ 0, & otherwise \end{cases}$$

Primary Sampling Unit	Household serial number	y	Base Weight (W)
1	1	1	120
1	2	1	120
2	1	1	40
2	2	1	40
3	1	0	50
3	2	1	50
4	1	1	80
4	2	0	80
4	3	1	80

8. What is the estimated total households from the survey?
   a) 9
   b) 660
   c) 15
   d) 530

9. What is the estimated proportion of household having at least one smart phone?
   a) 77.8%
   b) 79.5%
   c) 80.3%
   d) 81.6%

10. If the CPI for the current year is 120 and the CPI for the previous year was 115, what is the inflation rate?
    a) 4.35%
    b) 5.00%
    c) 4.17%
    d) 5.22%

11. The height distribution of 80 staff members at the National Statistics Bureau follow a normal distribution with mean 165 cm and standard deviation 25 cm. If 15 staff have their height between 160 cm and 170cm, the number of staff whose height is less than 160 cm is?
    a) ≈ 24
    b) ≈ 29
    c) ≈ 31
    d) ≈ 34

The summary output of the regression model studying effect of gender (two factors: male and female) on test scores is presented below. Use the output to answer **Question 12, 13 and 14**.

Coefficients:				
	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	76.464	1.295	59.045	< 2e-16 ***
GenderMale	-6.120	1.831	-3.342	0.00118 **

12. What does the intercept represent in this regression model?
    a) The average score for males
    b) The average score for females
    c) The difference between male and female scores
    d) The overall average score across all participants

13. What is the predicted score for a male participant according to this model?
    a) 76.464
    b) 70.344
    c) 82.584
    d) Cannot be determined from the given information

14. If another factor, such as age, were added to this model, under what condition might the coefficient for '*GenderMale*' change significantly, and what would that imply?
    a) If age is correlated with gender, indicating a potential confounding effect.
    b) If age is independent of gender, indicating no impact on the gender coefficient.
    c) If the sample size increases, leading to more accurate estimates.
    d) If the variance of the residuals increases, indicating a poorer fit of the model.

15. Which of the following statements is not the property of Bernoulli distribution?
    a) The experiment consists of n repeated trials.
    b) Each trials results in an outcome that may be classified under two mutually exclusive categories as a success or as a failure.
    c) The repeated trials are independent.
    d) The probability of success denoted by p, changes from trail to trail.

16. In the context of time series forecasting, what is the "Box-Cox transformation" primarily used for?
    a) Handling missing data
    b) Stabilizing variance
    c) Identifying seasonality
    d) Reducing autocorrelation

17. In a two-tailed hypothesis test, if the p-value is 0.04, what is the probability of observing a test statistic as extreme or more extreme than the observed one in the opposite direction?
    a) 0.04
    b) 0.08
    c) 0.02
    d) 0.96

18. Suppose a continuous random variable X has a PDF given by $f(x)=3x^2$ for $x \in [0,1]$. What is the expected value E[X]?
    a) 1/2
    b) 3/4
    c) 2/3
    d) 3/5

19. Suppose you have two different samples from the same population. The first sample has a size of 25, and the second sample has a size of 100. If the standard deviation of the population is known, how does the standard error of the mean for the first sample compare to that of the second sample?
    a) The first sample's standard error is twice as large.
    b) The first sample's standard error is four times as large.
    c) The first sample's standard error is half as large.
    d) The first sample's standard error is equal to that of the second sample.

20. A dataset has a mean of 50 and standard deviation of 7. If every data point is reduced by 5, what are the new mean and standard deviation?
    a) 45 and 7
    b) 50 and 7
    c) 45 and 2
    d) 55 and 2

21. In the context of experimental design, what is a "confounding variable"?
    a) A variable that is deliberately manipulated by the researcher.
    b) A variable that is held constant throughout the experiment to isolate the effect of the treatment.
    c) A variable that influences both the independent and dependent variables, potentially distorting the observed effect.
    d) A variable that is randomly assigned to subjects to ensure unbiased results.

22. For a binomial distribution with n = 8 and p = 0.2, what is the standard deviation?
    a) 0.12
    b) 2.24
    c) 1.28
    d) 1.13

Use the boxplot below to answer **Question 23**.

Monthly Household Expenses
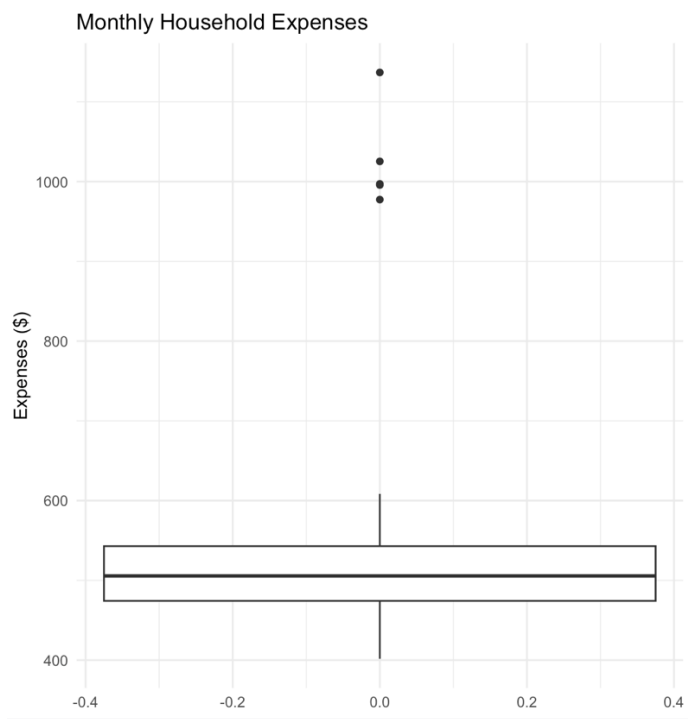
23. In the boxplot of monthly household expenses, what can be inferred about the skewness of the data?
    a) The data is symmetric.
    b) The data is positively skewed.
    c) The data is negatively skewed.
    d) The data is normally distributed.

24. A bag contains 25 balls marked 1 to 25. One ball is drawn at random. What is the probability that it is marked with a number of multiple 5 or 7?
    a) 5/25
    b) 3/25
    c) 6/25
    d) 8/25

25. In hypothesis testing, what is the main difference between a Type I error and a Type II error?
    a) Type I error is failing to reject a false null hypothesis; Type II error is rejecting a true null hypothesis.
    b) Type I error is rejecting a true null hypothesis; Type II error is failing to reject a false null hypothesis.
    c) Type I error occurs when the sample size is too large; Type II error occurs when the sample size is too small.
    d) Type I error occurs in one-tailed tests; Type II error occurs in two-tailed tests.

26. Which of the following is true about the posterior distribution in Bayesian statistics?
    a) It is the same as the prior distribution.
    b) It is the likelihood function alone.
    c) It is obtained by updating the prior distribution with the likelihood function.
    d) It is the prior distribution multiplied by the data.

27. What is the main purpose of using a control group in an experimental design?
    a) To increase the sample size of the experiment
    b) To ensure that the experimental procedure is standardized
    c) To provide a baseline for comparison and measure the effect of the treatment
    d) To reduce the cost of the experiment

28. In a multiple regression model, adding more independent variables to the model will always
    a) Increase the R^2 value
    b) Improve the model's predictive accuracy
    c) Reduce the variance of the error term
    d) Increase the multicollinearity in the model

29. The unit for variance is
    a) Unitless
    b) Original unit
    c) Original unit squared
    d) None of above

30. A city reports a CPI of 200 this year. If the CPI last year was 180, what is the percentage change in the purchasing power of the currency?
    a) 11.11%
    b) 10.00%
    c) 12.00%
    d) 8.89%

PART II – Short Answer Questions [20 marks]

This part has 4 Short Answer Questions. Answer ALL the questions. Each question carries 5 marks.

1. A company wants to estimate the average time (in hours) of its employees spend on training per month. A random sample of 40 employees show that the average training time is 8 hours with a standard deviation of 2 hours. Calculate a 95% confidence interval for the true average training time and interpret it. **(3 + 2 marks)**

2. A researcher wants to estimate the average monthly expenditure on healthcare with a margin of error of Nu 50 and a confidence level of 95%. From a preliminary study, the standard deviation of expenditure is estimated to be Nu 500. What sample size is needed? **(5 marks)**

3. A problem in Statistics is given to five students - Sonam, Karma, Penjor, Choki and Tashi. Their chances of solving it are 1/6, 1/5, 1/4, 1/3, 1/2, respectively. What is the probability that the problem will be solved? **(5 marks)**

4. Assume X_1, X_2, …….., X_n are independent and identically distributed (iid) random variables. And $\mathbb{E}(X_i) = \mu$ and Var $(X_i) = \sigma^2$. Write/derive the distribution of $\bar{X}$. **(5 marks)**

SECTION B: Case Study [50 marks]

Choose either CASE I OR CASE II from this section. Each case study carries 50 marks. Mark for each sub-question is indicated in the brackets.

Case I

A company wants to assess whether the proportion of customers who prefer their new product is different from the proportion of customers who preferred their previous product. They conducted a survey with a sample of 400 customers. Of these, 180 indicated a preference for the new product. The company previously reported that 40% of their customers preferred the old product.

You are required to perform hypothesis testing on the proportion of customers who prefer the new product. Answer the following sub-questions.

1. Formulate the hypothesis. Specify the significance level for this test. **(10 marks)**
2. Calculate test statistic and compare to its sampling distribution under the null hypothesis by calculating the p-value. What do you conclude at the level of significance decided? **(15 + 5 marks)**
3. Construct a 95% confidence interval for the proportion of customers who prefer the new product. Interpret the confidence interval in the context of the study. **(10 marks)**

4. Discuss how increasing the sample size would affect the test statistic and the confidence interval. Also explain why larger sample sizes generally provide more reliable results. **(5 + 5 marks)**

Case II

You are tasked with designing a survey to evaluate customer satisfaction for a newly launched mobile banking app by a financial institution. The goal of the survey is to gather feedback on user experience, identify areas for improvement, and assess overall satisfaction.

You may adopt the following structure.
1. Define the key objectives of the survey. **(5 marks)**
2. Design a sampling strategy for the survey. **(20 marks)**
    - Identify the target population for the survey.
    - Describe the sampling method you would use and justify your choice.
    - Explain how you would ensure the sample is representative.
3. Discuss the potential biases that might affect the survey results and suggest strategies to minimize them. **(5 marks)**
4. Data collection and analysis plan **(10 marks)**
    - Explain how you would collect the data from the survey.
    - Discuss data collection tools
    - Discuss the data entry and pre-processing steps required before performing statistical analysis.
5. Ethical consideration **(5 marks)**
6. Outline a overall work plan for survey administration. **(5 marks)**

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
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08691	.08534	.08379	.08226
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
-0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
-0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
-0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
-0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
-0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
-0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
-0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
-0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
-0.1	.46017	.45620	.45224	.44828	.44433	.44038	.43644	.43251	.42858	.42465
-0.0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414

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

TASHI DELEK