ROYAL CIVIL SERVICE COMMISSION
BHUTAN CIVIL SERVICE EXAMINATION (BCSE) 2025
EXAMINATION CATEGORY: TECHNICAL

PAPER II: GENERAL SUBJECT KNOWLEDGE PAPER FOR STATISTICS

Date	:	October 10, 2025
Total Marks	:	100
Writing Time	:	90 minutes (1.5 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 Parts: Part I & Part II**

   Part I consists of 70 multiple choice questions of 1 (one) mark each, and

   Part II consists of short answer questions for 30 marks.

4. All questions are COMPULSORY.
5. 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.
6. **All answers should be written with correct numbering of Part and Question Number in the Answer Booklet provided to you. Note that any answer written without indicating the correct Part and Question Number will NOT be evaluated and no marks will be awarded.**
7. Begin each Part in a fresh page of the Answer Booklet.
8. You are not permitted to tear off any sheet(s) of the Answer Booklet as well as the Question Paper.
9. Use of any other paper including paper for rough work is not permitted.
10. **You must hand over the Answer Booklet to the Invigilator before leaving the examination hall.**
11. This paper has **13 printed pages**, including this instruction page.

GOOD LUCK!

PART I
Multiple Choice Questions (70 marks)

Choose the correct answer and write down the letter of your chosen answer in the Answer Booklet against the question number e.g. 71 (a). Each question carries ONE mark. Any double writing, smudgy answers or writing more than one choice shall not be evaluated.

1. Which of the following best defines *high-quality data* in statistical practice?
   a) Data that is consistent with previous studies and expectations.
   b) Data that is large in volume, regardless of how it was collected.
   c) Data that is fit for purpose.
   d) Data that has minimal variation and outliers.

2. A company records the time (in seconds) taken by employees to complete a task. Which level of measurement does this represent?
   a) Nominal
   b) Ordinal
   c) Interval
   d) Ratio

3. Which of the following is NOT true about the population "mean"`?
   a) A fixed quantity
   b) A parameter
   c) A random quantity
   d) Unknown, but what we want to know

4. How can "Information is Power" be misused statistically?
   a) By using small sample sizes to misrepresent trends
   b) By ensuring transparent data analysis
   c) By always using median over mean
   d) By avoiding data visualization.

5. A researcher submits a formal request to the National Statistics Office for access to individual-level census data, including names and addresses, for a housing study. The office rejects the request because
   a) disclosing identifiable data would breach confidentiality obligations.
   b) the census was funded by government, and only officials can use the data.
   c) only aggregated data has legal standing in academic publications.
   d) the data may not yet be cleaned or processed for public use.

6. Which of the following cannot be used to calculate a meaningful average?
   a) Age
   b) Monthly income
   c) Blood group type
   d) Exam score

7. Suppose the scores of the 10 graduates in the Bhutan Civil Service Examination 2024 are independent and identically distributed, following a normal distribution with a mean of 55 and a standard deviation of 4. Then, the $\Pr(\bar{X} > 57)$ *is*
   a) approximately 5.7%
   b) approximately 30.9%
   c) approximately 69.2%
   d) approximately 94.3%

8. Which of the following is TRUE about the Bias in statistics?
   a) Bias is the difference between sample standard deviation and the population mean.
   b) Bias is the difference between the estimate and the estimand.
   c) Bias is the average estimation error across infinitely many random samples.
   d) Bias is the average deviation of an individual data point from the sample mean.

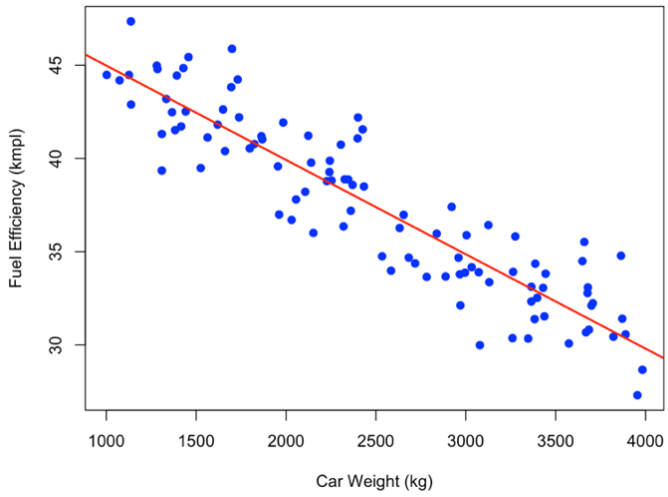
9. P-hacking is an example of
   a) ethical use of information.
   b) misusing statistical significance for false conclusions.
   c) always using $p < 0.0$.
   d) ignoring data entirely.

10. A researcher observes a strong correlation between ice cream sales and drowning deaths. However, a reviewer points out that both tend to rise during hotter months. What type of bias is most likely present in this analysis?
    a) Measurement error — drowning deaths may be inaccurately reported.
    b) Omitted variable bias — temperature influences both variables.
    c) Selection bias — the dataset may exclude colder months.
    d) Survivorship bias — only summer incidents are being analyzed.

11. You are preparing for a nationwide household Labour Force Survey. During a stakeholder consultation, Dzongkhag Planning Officers request data disaggregated by dzongkhag to support evidence-based planning and monitoring. What key factor must be ensured in the sample design to meet this request?
    a) Assign only open-ended questions to allow for diverse responses at the dzongkhag level.
    b) Design longer questionnaires specifically tailored for each dzongkhag.
    c) Include dzongkhag-level stratification in sample design.
    d) Use a two-phase sampling method to reduce the sample size while improving coverage.

12. If A and B are mutually exclusive events, then P(AB) is
    a) 0
    b) 0.5
    c) 1
    d) None of above

13. A scatter plot of 'Car Weight vs. Fuel Efficiency' shows a downward trend. A student concludes heavier cars are less efficient. What is wrong?



a) Correlation does not imply causation.
b) The plot does not account for differences in car types (e.g., SUVs vs. sedans).
c) The observed trend could be due to model year differences rather than weight alone.
d) The fuel efficiency axis might be measured differently (e.g., city vs. highway), affecting the interpretation.

14. In a survey of shoe sizes among staff at the National Statistics Bureau, the most common shoe size was 40. This value is best described as
a) mean.
b) median.
c) mode.
d) range.

15. Why is it important to pretest a survey questionnaire?
a) To determine the final survey cost and allocate budget accordingly
b) To detect unclear wording and ensure questions are interpreted as intended by respondents
c) To replace the need for interviewer training through practice during pretesting
d) To verify if respondents can complete the survey within the expected time limit

16. You want to compare two variables: household size and level of satisfaction with local services. What are their respective data types?
a) Ratio and Nominal
b) Interval and Ordinal
c) Ordinal and Nominal
d) Ratio and Ordinal

17. A policymaker questions the purpose behind government investment in national statistical systems. You responded by emphasizing that official statistics are not just technical outputs but play a vital role in democratic governance. What is the primary reason for compiling and disseminating official statistics?
    a) To collect data solely for academic use
    b) To ensure government ministries have access to tax records
    c) To uphold citizens' right to access public information
    d) To minimize the cost of maintaining administrative data systems

18. Which of the following is NOT TRUE about unbiasedness in statistical inference?
    a) An unbiased estimator always has the smallest possible variance.
    b) Unbiasedness is silent on how close the estimate is to the truth in any given sample.
    c) For a nonlinear function of a parameter, an unbiased estimator is typically not unbiased for that function.
    d) Unbiasedness does not guarantee accuracy in every sample, only on average across repeated samples.

19. Which of the following cannot directly interpreted from a boxplot?
    a) Median
    b) Presence of outliers
    c) Spread of middle 50%
    d) Mean

20. A policy analyst uses a 95% confidence interval for the difference in crime rates between two neighborhoods: [−2.1, 0.8]. What should they conclude?
    a) Crime is significantly higher in the first neighborhood.
    b) Crime is significantly lower in the first neighborhood.
    c) There is no significant difference.
    d) The sample size is too small.

21. Two dzongkhags have the same mean household size of 4, but dzongkhag A has a standard deviation of 0.5 and dzongkhag B has 2. What can be inferred?
    a) District A has more variation in family sizes.
    b) District B has more consistent household sizes.
    c) District A has more consistent household sizes.
    d) Both districts have identical distributions.

22. A histogram is constructed using very wide class intervals. What is the most likely consequence?
    a) The histogram may falsely suggest a uniform distribution.
    b) Extreme values will become more prominent.
    c) The shape will appear more symmetrical than it actually is.
    d) Important patterns or data variation may be obscured.

23. What is the main limitation of using administrative records as a data source for household surveys?
    a) They are always up-to-date.
    b) They may lack standardized definitions and have coverage gaps.
    c) They eliminate the need for sampling.
    d) They are more accurate than survey data.

24. A health ministry observes that areas with more hospitals also report higher numbers of illnesses. Which of the following interpretation best explains this pattern?
    a) Areas with more hospitals tend to have better reporting systems for illnesses.
    b) Higher population density may contribute to both more hospitals and more reported illnesses.
    c) Hospitals are likely placed in areas with greater demand for healthcare.
    d) Hospitals increase disease detection, inflating reported illness rates

25. Which of the following is NOT a component of data stewardship?
    a) Data acquisition
    b) Data de-identification
    c) Data storage
    d) Data marketing strategies

26. An e-commerce platform aims to recommend products based on customer behavior. Which method best supports this goal?
    a) Analyzing overall monthly revenue to adjust stock.
    b) Broadcasting identical promotions to all customers.
    c) Ranking items based on sporadic user reviews.
    d) Examining clickstreams and browsing history to suggest tailored products.

27. A policymaker noticed that higher literacy rates tend to coincide with lower poverty levels across districts. Which statistical concept is most appropriate to examine this relationship?
    a) Autocorrelation
    b) Correlation
    c) Extrapolation
    d) Regression

28. A company tests a new pricing strategy in Region A but not Region B. Sales rise in Region A. The CEO claims that the strategy worked. What is the key counterfactual question to assess causality?
    a) What would Region A's sales have been without the strategy?
    b) Did Region B's sales also rise?
    c) Is the strategy cost-effective?
    d) Are the regions statistically identical?

29. A farmer is evaluating different fertilizers. She records the number of plants that survive out of 100 trials. What is the random variable?
    a) Number of surviving plants
    b) Brand of fertilizer
    c) Type of crop
    d) Weather condition

30. Suppose the Gini index remains unchanged, but the poverty rate declines. What does this imply about income distribution?
    a) Income inequality has improved.
    b) The poor are benefiting equally with others.
    c) The gap between rich and poor has widened.
    d) The poor are earning more, but inequality persists.

31. A survey shows an increase in mobile phone ownership but a drop in school enrollment. Which statement is most accurate?
    a) Living standard is increasing due to digital access.
    b) Poverty is decreasing since gadgets are affordable.
    c) There may be inequality in dimensions of well-being.
    d) Unemployment is likely falling.

32. In order to build public trust and ensure the effective use of official statistics, why is it essential for statistical agencies to openly disclose their methods and procedures?
    a) To comply with international treaties that mandate transparency and data sharing
    b) To enable users to correctly interpret and assess the quality of the data provided
    c) To minimize the time and effort required from survey respondents during data collection
    d) To safeguard confidential information and prevent unauthorized access or leaks

33. In an increasingly interconnected world, why is it important for national statistical gencies to adopt internationally recognized methodologies when producing official statistics?
    a) To align with global standards for comparability and efficiency
    b) To simplify translation of reports for global audiences
    c) To reduce reliance on domestic data sources
    d) To delegate statistical work to international organizations

34. A study estimates average travel time to the nearest hospital as 75 minutes with a 95% confidence interval of (50, 100). The policy goal is to reduce time to under 60 minutes. What is the implication?
    a) The mean meets the goal.
    b) The interval suggests the goal is not certainly achieved.
    c) The sample overestimated true travel time.
    d) Travel time varies due to measurement error.

35. A researcher includes ZIP codes as a numeric variable in a regression model to predict income levels. What is the primary concern with this approach?
    a) ZIP codes are interval data and may not be linear.
    b) ZIP codes lack a natural zero, making interpretation difficult.
    c) ZIP codes are ordinal and should be analyzed as rankings.
    d) ZIP codes are categorical identifiers, not quantitative values.

36. If all values in a dataset are increased by 5, how does the standard deviation change?
    a) Increases
    b) Decreases
    c) Stays the same
    d) Becomes zero

37. The sum of deviations from the mean is always
    a) maximum.
    b) minimum.
    c) negative.
    d) zero.

38. A national statistics office is planning a nationwide Labour Force Survey but faces constraints in both budget and time. To ensure cost-effectiveness without compromising data quality, which of the following approaches should they adopt?
    a) Conduct a census covering all topics.
    b) Extend the questionnaire to reduce variance.
    c) Limit survey to large urban areas only.
    d) Use a probability-based household survey.

39. In the aftermath of a natural disaster, which technique would best support real-time emergency response?
    a) Downloading weather updates once a day
    b) Reviewing historical disaster records for trends
    c) Summarizing satellite images on a weekly basis
    d) Using live sensor data and satellite streams for immediate decision-making

40. If the consumer price index of a country is consistently rising, what economic phenomenon is most likely occurring?
    a) Deflation
    b) Depression
    c) Inflation
    d) Recession

41. According to the axioms of probability, the probability of any event is
    a) always greater than 1.
    b) always less than 0.
    c) equal to zero only.
    d) between 0 and 1 inclusive.

42. A survey manager wants to ensure that interviewers do not skip sensitive questions near the end of the questionnaire. What should be done to help reduce item non-response?
    a) Train interviewers to follow skip patterns.
    b) Leave sensitive items blank to avoid discomfort.
    c) Make all sensitive questions open-ended.
    d) Use only mail-in questionnaires for such items.

43. Sonam scored within the top 25% of the Preliminary Examinations 2024. This means Sonam's score is
    a) above the mean.
    b) above the median.
    c) above the first quartile.
    d) above the third quartile.

44. A data entry operator mistakenly enters "35" instead of "53" for a respondent's age. What is the most appropriate term for this error?
    a) Processing error
    b) Entry bias
    c) Interviewer bias
    d) Sampling error

45. The civil service statistics shows following age distribution of civil servants: Mean = 50, Median = 45, Mode = 40. What is the likely shape of the distribution?
    a) Symmetric
    b) Right-skewed
    c) Left-skewed
    d) Uniform

46. A statistician compares the mean monthly expenditure of two regions. Region A has a lower mean but much higher standard deviation. What does this suggest?
    a) Region A's spending amounts are generally closer to the average despite lower mean.
    b) Region B's expenditures are more consistently spread around a higher mean.
    c) The comparison is inconclusive without knowing the median values.
    d) Region A has higher inequality in spending.

47. A retailer observes a consistent increase in sales every November and December over the past five years. Which component of time series is most likely responsible for this pattern?
    a) Cyclical variation
    b) Irregular variation
    c) Random noise
    d) Seasonal variation

48. In a survey, the income of households is used to predict education attainment. Which statistical tool is most suitable for this?
    a) Chi-square test
    b) Correlation matrix
    c) Linear regression
    d) T-test for proportions

49. A program evaluation found an average reduction in stunting from 35% to 25%. What would help determine if this change is due to the program or random chance?
    a) Hypothesis testing
    b) Confidence level
    c) Experimental design
    d) Sampling fraction

50. In a survey, a respondent gives answers they believe the interviewer wants to hear. This is an example of
    a) frame error.
    b) measurement bias.
    c) sampling error.
    d) systematic variance.

51. Events A, B, and C are mutually exclusive if and only if
    a) all three events occur at the same time.
    b) at least two of the events occur simultaneously.
    c) the occurrence of any one event prevents all others from occurring.
    d) exactly one of the events occurs.

52. A coffee machine dispenses liquid with mean 200 ml and SD 5 ml. What proportion of cups contain between 190 ml and 210 ml?
    a) Approximately 34%
    b) Approximately 68%
    c) Approximately 95%
    d) Exactly 100%

53. In surveys, what does the process of "data cleaning" primarily involve?
    a) Discarding entries with partial or missing responses to ensure a clean dataset
    b) Reviewing and resolving anomalies, inconsistencies, and potential entry errors to improve data quality
    c) Filtering the dataset to a smaller, more reliable subset for quicker analysis
    d) Preparing hard copies of the survey forms in an organized format for archival and reporting purposes

54. The National Statistics Bureau wants to estimate the average monthly income of citizens using survey data. Which statistical concept is most appropriate for this task?
    a) Correlation analysis
    b) Hypothesis testing
    c) Time-series modeling
    d) Parameter estimation

55. When estimating the poverty rate for planning, the sample standard error is 1.2%. What does this standard error represent?
    a) The average amount by which the estimated poverty rate varies from the poverty rate across repeated samples.
    b) The percentage of poor households in the sample.
    c) The probability that poverty rate is exactly 1.2%.
    d) The amount of variability in poverty levels among households.

56. A dzongkhag education survey has a confidence interval of (52%, 66%) for primary school attendance. What happens if confidence level increases from 95% to 99%?
    a) Interval becomes narrower.
    b) Interval becomes wider.
    c) Mean estimate decreases.
    d) Sample size increases automatically.

57. An investor finds a correlation of r = 0.65 between Company X's stock price and gold prices. What should the investor avoid assuming?
    a) Further analysis is needed to understand the underlying connection.
    b) Gold prices can help predict Company X's stock movements.
    c) Investing in gold will automatically increase Company X's stock value.
    d) The relationship implies that Company X is a gold mining firm.

58. The Royal Bhutan Police evaluates whether installing speed cameras reduces traffic accidents. After analysis, they fail to reject the null, but accidents actually decreased due to cameras. This is an example of
    a) confirmation bias.
    b) measurement bias
    c) type I error.
    d) type II error.

59. A monitoring team estimates the average yield per acre is 850 kg with confidence interval (810, 890). A new fertilizer claims to increase yield by 50 kg. How would you interpret this in light of the confidence interval?
    a) New fertilizer effect is definitely visible.
    b) The yield is already high, so fertilizer won't help.
    c) The observed difference may not be statistically significant.
    d) We must lower the confidence level to test fertilizer.

60. A civil service analyst models the relationship between education level in years and income in Ngultrum (Nu). The R^2 of the simple linear regression is 0.45. What does this mean?
    a) 45% of variation in education is explained by income.
    b) 45% of variation in income is explained by education.
    c) Education and income have a 45% correlation.
    d) Income increases by Nu 45 for every year of education.

61. In a multiple regression model predicting student performance using study hours, teacher experience, and school facilities, the coefficient for study hours is found to be statistically insignificant. Which of the following is the most plausible explanation for this result?
    a) Study hours are irrelevant and add no unique information to the model.
    b) Study hours duplicate information already captured by another variable.
    c) The effect of study hours is masked by model misspecification.
    d) The data quality for study hours is poor and introduces noise into the model.

62. A factory has 0.5 defects per 100 items. In 10,000 items, what's the most likely number of defects?
    a) 0
    b) 25
    c) 50
    d) 100

63. A stadium tracks the number of fans entering through a gate per minute and uses a Poisson distribution to model the arrivals. Which of the following situations violates a key assumption?
    a) Fans enter randomly throughout the game .
    b) Fans come in large groups at specific times.
    c) The total average per hour remains stable.
    d) Fans enter independently of each other.

64. A national statistical agency wants to identify households vulnerable to poverty in urban areas. Which indicator is most appropriate?
    a) Access to mobile internet
    b) Crowding index or persons per room
    c) Number of children per household
    d) Total area of household land

65. In a study of rural households, owning a television was found to be positively correlated with household income. Which of the following statement best explains this finding?
    a) Higher-income households are more likely to afford televisions.
    b) Households with televisions earn more due to access to information.
    c) Television ownership directly results in increased household earnings.
    d) Watching television motivates people to pursue better-paying jobs.

66. A bank uses a normal distribution to model daily customer transaction amounts, with an average of Nu 5,000. To design an alert system for unusual spending, the bank needs to understand how transaction amounts are distributed. What is the probability that a randomly selected transaction exceeds Nu 5,000?
    a) 0%
    b) 25%
    c) 50%
    d) 75%

67. During a cabinet meeting, the unemployment rate is cited as "rising slightly". What does this indicate?
    a) Fewer people are joining the labor force.
    b) More people are retired.
    c) The population has increased.
    d) More people are actively seeking but not finding jobs.

68. A country with high Gross Domestic Product but low Human Development Index is under scrutiny. What could be a likely explanation?
    a) Economic growth is highly equitable.
    b) Investments in health and education are low.
    c) The population is mostly elderly.
    d) The unemployment rate is zero.

69. A teacher writes the word "DATA" on the board and asks students to rearrange all its letters to form different 4-letter sequences. How many unique arrangements are possible?
    a) 12
    b) 18
    c) 24
    d) 48

70. If the poverty rate decreases from 18% to 12% over five years, what can reasonably be inferred?
    a) Food prices have fallen significantly.
    b) Incomes have improved for all households.
    c) Living conditions improved for the poorest.
    d) Unemployment has reached zero

PART II
Short Answer Questions (30 marks)

Answer ALL short answer questions. Marks for each question are indicated in the brackets.

1. In a Likert-scale survey assessing employee engagement, the responses to a particular statement were:

   1.Strongly Disagree: 10%
   2.Disagree: 15%
   3.Neutral: 20%
   4.Agree: 40%
   5.Strongly Agree: 15%

   a) Interpret what this distribution suggests about employee perception. **[2 marks]**
   b) What could be the possible impact on survey results if the Neutral option is removed from the Likert scale? **[2 mark]**
   c) What type of central tendency measure would be most appropriate? **[1 Marks]**

2. Explain the importance of minimizing proxy respondents in household surveys. What are some challenges in achieving self-reported data from all adult members? **[2+3 marks]**

3. Explain why a strong metadata and dissemination system is essential for the effective use of survey microdata. **[5 marks]**

4. Do you think we can rely solely on administrative data instead of conducting surveys? Discuss with supporting arguments, highlighting both the advantages and limitations. **[5 marks]**

5. What are the key dimensions used to assess the quality of administrative data? Provide a brief description on each aspect. **[5 marks]**

6. State True of False and then support your answer with a brief explanation.

   a) In a two-tailed hypothesis test, doubling the p-value from a one-tailed test always yields the correct p-value for the two-tailed test. **[1 + 1.5 marks]**

   b) The Central Limit Theorem guarantees that the sampling distribution of the sample mean is normal for any sample size regardless of the population distribution. **[1 + 1.5 marks]**

TASHI DELEK