

ROYAL CIVIL SERVICE COMMISSION
BHUTAN CIVIL SERVICE EXAMINATION (BCSE) 2020
EXAMINATION CATEGORY: TECHNICAL

PAPER II: GENERAL SUBJECT KNOWLEDGE PAPER FOR STATISTICS

Date	: February 26, 2021
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 the 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 10 short answer questions of 3 (three) marks each.
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 on 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 **16 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 is NOT true about statistics?
 - a) Statistics is a study of human disease, agriculture, evolution, ecology and biodiversity.
 - b) Statistics is the art of learning from data.
 - c) Statistics is a science of information.
 - d) Statistics is a collection of facts.

2. The population consists of the set of all measurements in which investigator is interested. The population is also called
 - a) A sample
 - b) The universe
 - c) The class
 - d) The group

3. Which of the following is a set of measurements obtained on some variable?
 - a) Quantitative data
 - b) A data set
 - c) Qualitative data
 - d) Information

4. A conclusion drawn about a population based on the information in a sample from the population is
 - a) Mean
 - b) Median
 - c) A statistical inference
 - d) A descriptive statistic

5. Which of the following divides the data into two equal halves? The same is also called the 50th percentile.
 - a) Mean
 - b) Median
 - c) Mode
 - d) Standard Deviation

6. The 75th percentile of the dataset is also called
 - a) First quartile
 - b) Second quartile
 - c) Third quartile
 - d) Fourth quartile

7. The mean, median and mode of the following set of observations (4, 4, 5, 7, 10) are
 - a) mean=5, median=6 and mode=4
 - b) mean=6, median=5 and mode=4
 - c) mean=6, median=5 and mode=5
 - d) mean=6, median=4 and mode=4

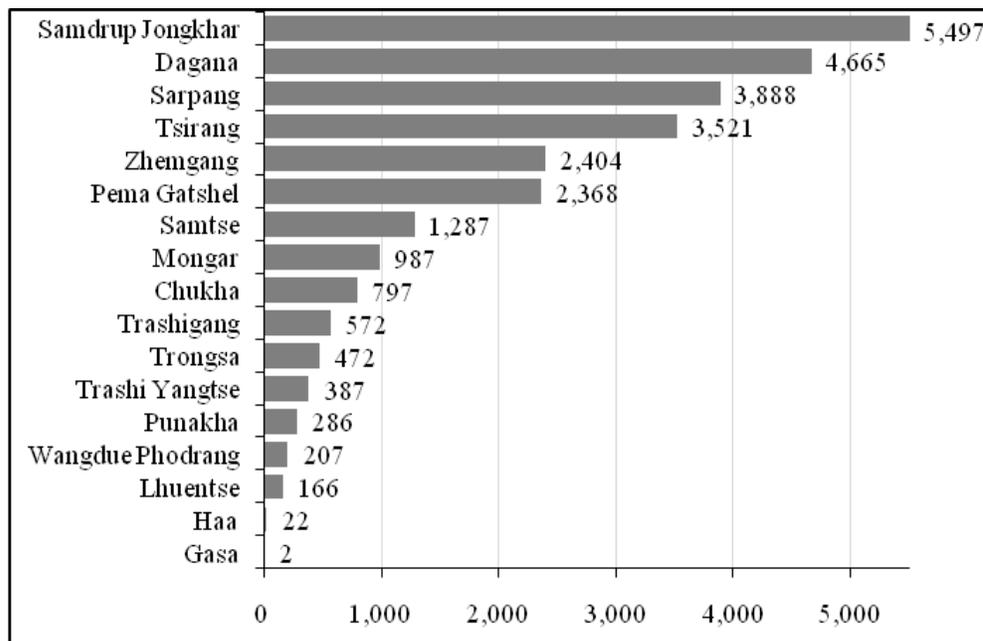
8. The mean, median and mode are the measures of
 - a) Central tendency
 - b) Dispersion
 - c) Logarithms
 - d) Proportions

9. Which of the following is resistant to extreme observations?
 - a) Mean
 - b) Median
 - c) Mode
 - d) Standard Deviation

10. The average squared deviation of the data points from their mean is
 - a) Range
 - b) Variance
 - c) Standard Deviation
 - d) Inter-quartile Range

11. A chart made of bars of different height and the height of each bar represents the frequency of values in the class. Which of the following fits the above description?
 - a) Histogram
 - b) Pie charts
 - c) Line charts
 - d) There is not enough information to answer the question

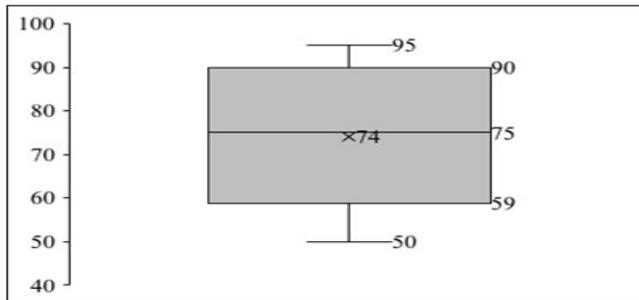
Use the following horizontal bar graph to answer **Questions 12, 13 and 14**. It shows the production of mandarin in Bhutan by 17 different dzongkhags as per 2019 annual agriculture survey. The productions were recorded in MT.



12. Which of the following is NOT true about mandarin production in Bhutan?
- About 20% of total mandarin production came from Samdrup Jongkhar dzongkhag in 2019.
 - About 13 dzongkhags contributed less than 10% of total mandarin production in 2019.
 - About 51% of total mandarin production came from the top three dzongkhags in 2019.
 - One of the southern dzongkhags was the highest mandarin producer in 2019.
13. What was the median mandarin production in 2019?
- Approximately 750 MT
 - Approximately 780 MT
 - Approximately 797 MT
 - Approximately 1619MT
14. Which of the following is NOT true about mandarin production in Bhutan?
- Punakha dzongkhag was in the 25th percentile
 - Chhukha dzongkhag was in the 50th percentile
 - SamdrupJongkhar dzongkhag was in the 55th percentile
 - Zhemgang dzongkhag was in the 75th percentile
15. Which of the following is a measure of the degree of asymmetry of a distribution?
- Standard deviation
 - Variance
 - Skewness
 - None of the above

16. Which of the following statement is TRUE about a symmetric distribution?
- The mean is to the right of the median and mode
 - The mean is to the left of the median and mode
 - Symmetric distribution with a single mode has mode=mean=median
 - None of the above
17. Assume that in a study we found that the mean is 15.85 and the standard deviation is 4.46. According to Chebyshev's theorem, at least three-quarters of the observations will lie within 2 standard deviations of the mean. Which of the following pair of points is TRUE?
- (4.46, 15.85)
 - (4.46, 24.77)
 - (6.93, 24.77)
 - (15.85, 24.77)
18. A box-and-whisker plot is another way of looking at a data set. Which of the following is NOT true about a box-and-whisker plot?
- We can determine the spread of the data set.
 - We can determine the skewness of the data set.
 - We can determine the central tendency of the data set.
 - We can't determine the existence of the outliers of the data set.
19. Which of the following would indicate that a dataset is NOT bell-shaped?
- The range is equal to 5 standard deviation.
 - The range is larger than the interquartile range.
 - The mean is much smaller than the median.
 - There are no outliers.
20. The geometric mean of 3, 6, x is 6. What is the value of x?
- 8
 - 10
 - 12
 - 16
21. A total of 25 students appeared for statistics exam and the average marks obtained was 75%. A total of 15 female students in the class obtained an average mark of 70%. What is the average mark of male students in the class?
- 63.0%
 - 79.5%
 - 79.6%
 - 82.5%

Use the following box-and-whisker plot to answer **Questions 22, 23 and 24**. It shows the distribution of marks obtained by 10 students in a class. Note that the percent is given on the vertical axis.



22. What is the range of marks obtained by students in the class?
 - a) 45
 - b) 50
 - c) 74
 - d) 95

23. What is the approximate median marks obtained by students?
 - a) 70
 - b) 74
 - c) 75
 - d) 95

24. The interquartile range (IQR) is the difference between third (Q3) and first (Q1) quartiles. What is the IQR of marks obtained by students in a class?
 - a) 31
 - b) 50
 - c) 74
 - d) 75

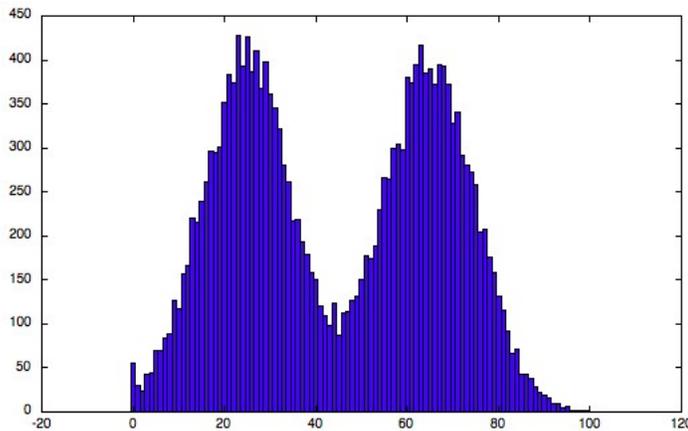
Use the following table to answer **Questions 25 and 26**. According to 2012 Bhutan Living Standard Survey (BLSS) of National Statistics Bureau, the mean monthly household food consumption expenditure by area in Ngultrum is provided in the table below:

	Urban	Rural
Mean	8,141	6,644
Standard Deviation (SD)	2,600	2,100

25. Which of the following is TRUE about the mean monthly household food consumption expenditure in Bhutan?
 - a) The monthly food consumption expenditure for urban households is more variable.
 - b) The monthly food consumption expenditure for rural households is more variable.
 - c) The monthly food consumption expenditure for urban and rural households has approximately equal variability.
 - d) None of the above.

26. The Co-efficient of Variation (CV) is the relative measure of variability. The variability in the monthly food consumption expenditure while accounting for differences in mean by area is approximately
- 30%
 - 31%
 - 32%
 - 33%

Use the following histogram to answer **Question 27**.



27. Which of the following is TRUE about the above histogram?
- The above histogram is unimodal.
 - The above histogram is bimodal.
 - The above is not a histogram.
 - None of the above.
28. Which of the following is NOT true about the standard normal distribution?
- The shape of the standard normal distribution has a smooth and bell-shaped curve.
 - The standard normal distribution was originally called the law of errors.
 - The standard normal distribution has mean=0 and standard deviation=1
 - The standard normal distribution has mean=1 and standard deviation=1
29. Which of the following is NOT true about a random normal variable?
- The probability that a normal random variable will be within 1 standard deviation from its median is approximately 68%.
 - The probability that a normal random variable will be within 1 standard deviation from its mean is approximately 68%.
 - The probability that a normal random variable will be within 2 standard deviation from its mean is approximately 95%.
 - The probability that a normal random variable will be within 3 standard deviation from its mean is approximately 99%.

30. A numerical measurement of the sample is called
- Quantitative data.
 - Time series data.
 - Statistics.
 - Population.
31. The probability distribution of all possible values that statistics may take when computed from random sample of the same size, drawn from a specified population is
- census of a population.
 - survey of a population.
 - sampling distribution of a population.
 - sampling distribution of a statistics.
32. Which of the following is NOT true about the Central Limit Theorem (CLT)?
- The distribution of the sample mean tends to a normal distribution regardless of the distribution of the population from which the random sample is drawn.
 - The theorem allows us to make probability statements about the possible range of values that same mean can take.
 - It also allows us to compute probabilities of how far away the sample mean may be from the population mean it estimates.
 - The theorem also provides that the rate at which the distribution approaches a normal distribution does not depend on the shape of the distribution of the parent population.
33. If an estimator is said to be unbiased, which of the following is TRUE?
- The expected value is equal to the population parameter.
 - There will be systematic deviation of the estimator from the parameter.
 - The distance between the expected value and the population parameter is large.
 - There is relatively large variance between the expected value and the population parameter.
34. The first quartile is called
- Lower quartile
 - Median
 - Upper quartile
 - Geometric mean
35. The second quartile is also known as
- Mean
 - Median
 - Mode
 - Standard Deviation (SD)

36. The concept of degrees of freedom (df) is important for understanding many statistical distributions. Which of the following is TRUE about the degrees of freedom?
- It is the total measurements divided by the total number of restrictions on the measurements.
 - It is the total number of measurements less the total number of restrictions on the measurements.
 - It is the total measurements divided by the total number of measurements itself.
 - None of the above.
37. A 95% confidence interval for the unknown population mean (μ) with known sample mean (\bar{x}) of 122 and the population standard deviation (σ) of 20 for a sample size (n) of 25 is
- (114.16,122.00)
 - (114.16,120.89)
 - (114.16,129.89)
 - (122.00,129.89)
38. Which one of these statistics is unaffected by outliers?
- Mean
 - Range
 - Standard deviation
 - Interquartile range
39. The level of significance of a statistical hypothesis test is denoted by
- α
 - β
 - $(1 - \alpha)$
 - $(1 - \beta)$
40. In statistical hypothesis testing, the level of significance to reject or fail to reject the null hypothesis is set at
- 1%
 - 5%
 - 10%
 - 15%
41. Which of the following is NOT true about p-value in hypothesis testing?
- The p-value is the level of significance that goes without value of the test statistics.
 - The p-value is the smallest level of significance α at which a null hypothesis may be rejected using the obtained level of the test statistics.
 - The p-value is the probability of obtaining a value of the test statistics as extreme as, or more extreme than the actual value obtained when the null hypothesis is true.
 - The p-value is the probability of obtaining a value of the test statistics as extreme as, or more extreme than the actual value obtained when the alternative hypothesis is true.

42. The mean of a set of 35 numbers is 55. What is the sum of the 35 numbers?
- 35
 - 1225
 - 1925
 - 3025
43. Which of the following steps is NOT true while calculating standard deviation?
- Subtract the mean from each term in the distribution to obtain a set of differences and square each sum difference and add the squares.
 - Divide the sum of the squared differences by the number of terms in the distribution.
 - Find the square root of the quotient just obtained; the positive square root is the standard deviation.
 - Find the square root of the quotient just obtained; the negative square root is the standard deviation.
44. The standard deviation of the sample mean is also called
- Standard deviation.
 - Squared standard deviation.
 - Standard error.
 - Standard deviation of the population mean.
45. Which of the following is TRUE about the distribution affects of standard deviation and the variance?
- If a constant is multiplied to each data point, it does not affect the variance or the standard deviation of the distribution.
 - If each data point is divided by a constant, it does not affect the variance or the standard deviation of the distribution.
 - If a constant is added or subtracted to each data point, it does not affect the variance or the standard deviation of the distribution.
 - If each data point is squared by itself, it does not affect the variance or the standard deviation of the distribution.
46. Failing to reject the null hypothesis when it is true is
- Alpha
 - Beta
 - Type I error
 - Type II error
47. All of the following are categorial variable EXCEPT
- Study major (Statistics, Mathematics, Physics, Engineering).
 - Color (White, Black, Gray, Red).
 - Performance rating (Poor, Good, Very good, Outstanding).
 - Gender (Male, Female).

48. Correlation provides association between two variables and it measures both the strength and direction of the linear relationship. If there is a positive relationship between two variables, what type of slope is expected in a scatterplot?
- Fairly a straight line
 - An upward slope
 - A downward slope
 - None of the above
49. Assume that Druk Pizza in Thimphu has a mean delivery time of 30 minutes and a standard deviation of 5 minutes. Using the standard normal distribution rule, what can you say about the delivery time of Druk Pizza in Thimphu?
- 95% are between 20-40 minutes
 - 95% are between 22-35 minutes
 - 95% are between 15-45 minutes
 - None of the above
50. If two variables share linear relationship, the graph will be a
- Curved
 - Hyperbola
 - Straight line
 - None of the above
51. The values of correlation between two variables will be
- More than 0
 - More than -1
 - Less than -1
 - Between -1 and +1
52. If there is a strong linear relationship between two variables, which of the following is a reasonable conclusion?
- The value of the correlation coefficient will be close to 0
 - The value of the correlation coefficient will be close to 1
 - The value of the correlation coefficient will be greater than 1
 - It is impossible to quantify their relationship
53. Suppose there is a correlation of $r=0.9$ between the number of hours per day a student study and GPA score. Which of the following is a reasonable conclusion?
- 90% of students who study receive high grades.
 - 90 of students who receive high grades study a lot.
 - 90% of the variation in GPA scores can be explained by variation in number of study hours.
 - 81% of the variation in GPA scores can be explained by variation in number of study hours per day.

54. Increasing the sample size in the survey can
- reduce the residual error.
 - reduce the inferential error.
 - reduce the sampling error.
 - None of the above.
55. Data collected through survey, census and experiment approach is called
- Administrative data.
 - Primary data.
 - Secondary data.
 - None of the above.
56. Which of the following is TRUE about the bar chart?
- It is a simple tool to visualize the relative or absolute frequencies of observed data.
 - It is a simple tool to visualize the average value of groups of data.
 - It is a simple tool to visualize the relationship of average value of one variable plotted against a second variable.
 - None of the above.
57. The weighted average of 63, 47 and 98 with respective weights of 0.144, 0.305 and 0.551 is
- 208.00
 - 77.41
 - 69.33
 - 25.80
58. Any well-defined procedure that produces an observable outcome that could not be perfectly predicted in advance is called
- A survey
 - A census
 - A random experiment
 - Administrative data source
59. Suppose you interviewed 600 students who appeared for 2020 PE exam of the Royal Civil Service Commission. It was observed that 38 students found the exam was easy. What is the binomial proportion (binomial fraction)?
- 0.05%
 - 0.06%
 - 6.30%
 - 6.50%

60. If a relationship between height and weight of a student is found to be statistically significant, it means
- variables are related in the sample due to chance alone.
 - variables are not-related in the population represented by the sample.
 - variables are related in the population represented by the sample.
 - variables are very important.
61. Simpson's Paradox occurs when
- no baseline risk is given, so it is not known whether or not a high relative risk has practical importance.
 - a confounding variable rather than the explanatory variable is responsible for a change in the response variable.
 - the direction of the relationship between two variables changes when the categories of a confounding variable are considered.
 - the results of a test are statistically significant but are really due to chance.
62. What is the standard deviation (SD) of the observations 75, 79, 82, 84?
- 3.39
 - 3.92
 - 11.5
 - 15.33
63. Which of the following is TRUE about Poisson distribution?
- Distribution has greater mean than variance.
 - Distribution has equal mean and variance.
 - Distribution has unequal mean and variance.
 - None of the above.
64. The heights of students were recorded and found to be normally distributed with mean (\bar{x}) 175cm and standard deviation (σ) 10cm. How much a student with 190 cm tall is away from the mean in terms of standard deviation?
- 1.0 SD
 - 1.5 SD
 - 2.0 SD
 - 2.5 SD
65. Sampling frame is important for conducting any survey. Which of the following is TRUE about sampling frame?
- The sampling frame is a smaller collection of units selected from the population.
 - The sampling frame is used to study about a system that is too large and costly to study in its entirety.
 - The sampling frame tells you how to gain access to the population units.
 - The sampling frame is defined as any number computed from your sample data.

66. In a survey, a pilot study is conducted before the real survey is run. Why is a pilot study important?
- A pilot study is important to help you identify respondents before the real survey is run.
 - A pilot study is important to help you identify problems and fix them before the real survey is run.
 - A pilot study is important to help you build good relationship between enumerators and respondents before the real survey is run.
 - A pilot study is important to help respondents and enumerators know what kind of questions are in the survey before the real survey is run.
67. The basis for any sampling method is the random sampling or simple random sampling. Which of the following is TRUE about the simple random sampling?
- In simple random sampling, each population unit has an equal probability of being chosen.
 - In simple random sampling, each population unit is chosen independently.
 - In simple random sampling, each population unit is chosen with/without replacement approach based on the objective of the survey.
 - All of the above.
68. A random sampling obtained by choosing a sample separately from each of the segments or groups of the population is called a
- Stratified random sampling.
 - Systematic random sampling.
 - Circular random sampling.
 - Convenience random sampling.
69. National Statistics Bureau has randomly chosen a sample of 500 respondents in a survey. About 300 people chose to respond and 200 did not respond to that survey. Which of the following is NOT true about the 200 people who did not respond to the survey?
- The 200 people who did not respond to the survey is called non-respondent.
 - The 200 people who did not respond to the survey are hard-core refusers and they should be penalized.
 - The 200 people who did not respond to the survey could be call-back and gather information.
 - The 200 people who did not respond to the survey will cause significant bias in the results of the survey.
70. Which of the following is NOT true about non-parametric tests?
- Non-parametric tests do not require assumptions about the distributions of the populations of interest.
 - Non-parametric tests require assumptions about the distribution of the populations of interest.
 - Non-parametric tests do not deal with particular parameters of the population.
 - Non-parametric tests are often called distribution-free methods.

PART II

Short Answer Questions [30 marks]

Answer ALL 10 short answer questions. Each question carries 3 marks.

Use the table below to answer Questions 1 to 10. It shows that the National Statistics Bureau has conducted a survey on salary and bonus paid in Ngultrum per month to Chief Executive Officers (CEOs) of 21 different private firms in 2020. The information gathered were sorted and ranked by gender as shown in the table below:

Private Firm	Consolidated salary and bonus	Gender
1	40,000	Female
2	45,000	Female
3	50,000	Female
4	55,000	Male
5	60,000	Male
6	65,000	Female
7	70,000	Female
8	75,000	Female
9	80,000	Male
10	85,000	Female
11	90,000	Female
12	95,000	Male
13	100,000	Male
14	105,000	Male
15	110,000	Female
16	115,000	Male
17	120,000	Male
18	125,000	Female
19	130,000	Female
20	135,000	Male
21	140,000	Male

1. What is the median salary paid to CEO?
2. What is the salary and bonus difference between CEO in 75th percentile and CEO in 25th percentile?
3. What is the standard deviation of salary and bonus of CEOs?
4. What is the range of salary and bonus for CEOs?

5. Is there any outlier in the data distribution? Suggest a way to detect outliers in the data distribution.
6. What is the mean salary and bonus for male and female respectively?
7. Whose salary and bonus are more variable and why?
8. When conducting a hypothesis test, we begin by specifying null and alternative hypothesis. State null and alternative hypothesis to test difference in salary and bonus between male and female CEOs?
9. What level of significance do we use to test the hypothesis and how do you relate to p-value?
10. The results of the two-sample t-test to test whether or not there is a significant difference in salary and bonus between male and female CEOs is provided below. What can you say about the results and justify your statements?

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Female	11	80454.55	9353.039	31020.52	59614.68	101294.4
Male	10	100500	9142.392	28910.78	79818.47	121181.5
combined	21	90000	6770.032	31024.18	75877.96	104122
diff		-20045.45	13125.27		-47516.95	7426.044

```

diff = mean(Female) - mean(Male)                                t = -1.5272
Ho: diff = 0                                                    degrees of freedom = 19

Ha: diff < 0                                                    Ha: diff != 0                                                    Ha: diff > 0
Pr(T < t) = 0.0716                                               Pr(|T| > |t|) = 0.1432                                           Pr(T > t) = 0.9284
    
```

TASHI DELEK