

**ROYAL CIVIL SERVICE COMMISSION
CIVIL SERVICE COMMON EXAMINATION (CSCE) 2008
EXAMINATION CATEGORY: TECHNICAL**

PAPER III: SUBJECT SPECIALIZATION for STATISTICS GROUP

Date : 20/12/08
Total Marks : 100
Examination Time: $2\frac{1}{2}$ Hours
Reading Time : 15 minutes

Read the following instructions carefully before answering the questions.

The subject specialization paper is set to test your theoretical knowledge of Statistics as well as analytical skills & application of it. Marks will be given based on the knowledge of Statistics as well as clarity and preciseness of the response.

The paper consists of two Sections:

Section A

Part a: 30 multiple-choice questions of one mark each (30 marks)

Part b: 4 short answer questions of 5 marks each (20 marks).

Section B

This section consists of two case studies and you are to attempt one question. (50 marks)

All answers are to be written in the answer sheet.

Paper III consists of 8 pages including this page.

Section A

Part a. 30 multiple-choice questions of one mark each (30 marks).

(In this part four choices (a,b,c & d) are provided against each question. Write the question number on the answer sheet with the corresponding answer choice. No need to copy the whole question on the answer sheet.)

1. In hypothesis testing, the hypothesis tentatively assumed to be true is
 - a) the alternative hypothesis
 - b) the null hypothesis
 - c) either the null or the alternative
 - d) none is correct
2. The error of rejecting a true hypothesis is
 - a) Type I error
 - b) Type II error
 - c) Same as β
 - d) Committed when not enough information is available
3. When the p-value is used for hypothesis testing, the null hypothesis is rejected if
 - a) p-value = α
 - b) p-value > α
 - c) p-value < α
 - d) $\alpha <$ p-value
4. The probability of making a Type II error is denoted by
 - a) α
 - b) β
 - c) $1 - \alpha$
 - d) $1 - \beta$
5. Which of the following is not the measure of central location?
 - a) Mean
 - b) Mode
 - c) Range
 - d) Median
6. Which of the following does not need to be known in order to compute the p-value?
 - a) Knowledge of whether the test is one tailed or two tailed
 - b) The value of the test statistic
 - c) The level of significance
 - d) None of the above alternatives
7. Which of the following type of sampling is most widely used?
 - a) Systematic Random Sampling
 - b) Cluster Sampling

- c) Stratified Random Sampling
 - d) Simple Random Sampling
8. Originated by Karl Pearson about 1990, the coefficient of correlation describes the strength of the relationship between two sets of interval-scaled or ratio-scaled variables. The coefficient of correlation can assume any value from;
- a) 0.00 to 1.00
 - b) -2.00 to 2.00
 - c) -1.00 to 0.00
 - d) -1.00 to 1.00
9. The normal probability distribution has the following major characteristics except;
- a) It is bell shaped
 - b) It is symmetrical about the mean
 - c) The distribution is not asymptotic
 - d) The location of a normal distribution is determined by the mean (μ)
10. When one or more of the experiment's outcomes are observed we call this an
- a) Event
 - b) Experiment
 - c) Outcome
 - d) Probability
11. Which of the following is a correct statement about a probability?
- a) It may range from 0 to 1
 - b) It may assume negative values
 - c) It may be greater than 1
 - d) All the above are correct
12. The value of the Standard Normal Distribution (z) when the significance level $\alpha = 0.05$ is;
- a) 1.64
 - b) 1.96
 - c) 1.16
 - d) 1.54
13. The median for the following data is;
60,000; 80,000; 65,000; 275,000; 70,000
- a) 60,000
 - b) 65,000
 - c) 70,000
 - d) 80,000

14. If p is the proportion of Bhutanese population that has confidence in the DPT government, then the maximum variation will be at

- a) $p = \frac{1}{2}$
- b) $p = \frac{1}{4}$
- a) $p = \frac{1}{8}$
- b) $p = \frac{1}{16}$

15. High precision for an estimator can be viewed in terms of

- a) High MSE
- b) High Variance
- c) Low MSE
- d) Low Variance

16. If $E(\hat{\theta}) = \theta$, then $\hat{\theta}$ is

- a) Efficient
- b) Consistent
- c) Sufficient
- d) Unbiased

17. If the population is heterogeneous, then the sample size needed to study it will be

- a) Equal sample size with that of homogenous population
- b) More sample size than that of homogenous population
- c) Less sample size than that of homogenous population
- d) None of the above

18. Which of the following is not the use of Chi-Square (χ^2) test?

- a) Test of independence
- b) Test of goodness of fit
- c) Test of Normality
- d) Test of homogeneity

19. The following equation

$$p = \frac{\sum p_t q_t}{\sum p_0 q_t} * 100 \text{ is;}$$

- a) Laspeyre's price index
- b) Paasche's price index
- c) Fisher's ideal index
- d) Value index

20. A time series is a collection of data recorded over a period of time – weekly, monthly, quarterly, or yearly. Which of the following is not a component of a time series?
- Cycle
 - Seasonality
 - Moving Average
 - Irregularity
21. The strength of the association between a set of independent variables X and a dependent variable Y is measured by the
- Coefficient of Correlation
 - Coefficient of determination
 - Standard error of estimate
 - All of the above
22. What test statistic is used for a global test of significance?
- z statistic
 - t statistic
 - Chi-square statistic
 - F statistic
23. What is the relationship between the coefficient of correlation and the coefficient of determination?
- They are unrelated
 - The coefficient of determination is the coefficient of correlation squared
 - The coefficient of determination is the square root of the coefficient of correlation
 - They are equal
24. A coefficient of correlation was computed to be -0.90. This result means:
- The relationship between two variables is weak
 - The relationship between two variables is strong and positive
 - The relationship between two variables is strong and negative
 - There is no relationship between the two variables
25. The non-parametric method used to test for differences between two dependent samples is
- The median test
 - Chi-square test
 - Wilcoxon signed-rank test
 - Kruskal-Kallis test
26. Spearman's coefficient of rank correlation is best applied when
- The data are measured with a nominal scale
 - The data are measured with an ordinal scale
 - The data are measured with an interval scale
 - The data are measured with a ratio scale

27. The degrees of freedom for a Chi-square test of a contingency table with 6 rows and 3 columns is

- a) 18
- b) 15
- c) 12
- d) 10

28. For a probability distribution the sum of the probabilities for all possible outcomes must equal

- a) 0.5
- b) 1.0
- c) 1.5
- d) 0.0

29. The binomial and poisson probability distributions are

- a) Continuous
- b) Either discrete or continuous
- c) Discrete
- d) Normal

30. Based on your assessment of the stock market, you state that the chances that stock prices will start to go down within two months are 50-50. This concept of probability based on your beliefs is called

- a) Classical probability
- b) Empirical probability
- c) Subjective probability
- d) Independence

Section A

Part b. Four short questions of five marks each (20 marks).

1. If you ask three strangers about their birthdays, what is the probability that;
 - a) All were born on Wednesday?
 - b) None were born on Saturday?
2. In a lottery game, three numbers are randomly selected from a number of balls numbered 1 through 50.
 - a) How many permutations are possible?
 - b) How many combinations are possible?
3. The arithmetic mean, the mode and the median of a group of 75 observations were calculated to be 27, 34 and 29 respectively. It was later discovered that one observation was wrongly read as 43 instead of the correct value 53. Examine to what extent the calculated values of the three averages will be affected by error.
4. Distinguish between the following terms
 - a) Parameter and statistic
 - b) Qualitative data and quantitative data
 - c) Sample and population
 - d) A discrete and a continuous variable

Section B

(Two case study questions are provided. Attempt one question. 50 marks are allotted for this question)

1. During the last few years, youth unemployment has emerged as a pressing issue. If you are assigned to work in the department of Labour and were tasked to design a survey to tackle unemployment, how will you proceed? Give detailed survey methodology including (a) Planning the survey; (b) Questionnaire design; (c) Sampling methodology and (d) Tabulation plans.

2. Town representatives of Paro are considering increasing the number of police in an effort to reduce crime. Before making a final decision the town representatives asked the officer-in-charge (OC) of police to survey other cities to determine the relationship

between the number of police and the number of crimes reported. The OC gathered the following sample information. The data provided is hypothetical.

Town	Police	Number of crimes
Thimphu	15	17
Phuntsholing	17	13
S. Jongkhar	25	5
Gelegphug	27	7
Mongar	17	7
Trashigang	12	21
Samtse	11	19
Bumthang	22	6

- If we want to estimate crimes on the basis of the number of police, which variable is the dependent variable and which is the independent variable.
- Draw a scatter diagram
- Compute the correlation coefficient “r”
- Interpret the result

$$\text{Hint: } r = \frac{\sum (x - \bar{x})(y - \bar{y})}{(n-1)s_x s_y} = \frac{n \sum xy - \sum x \sum y}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}}$$