ROYAL CIVIL SERVICE COMMISSION
CIVIL SERVICE COMMON EXAMINATION (CSCE) 2009
EXAMINATION CATEGORY: TECHNICAL

PAPER III: SUBJECT SPECIALIZATION for STATISTICS GROUP

Date : 8/11/09
Total Marks : 100
Examination Time: $2\frac{1}{2}$ Hours
Reading Time : 5 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 9 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. If you are doing an experiment, when should you decide on the research hypothesis?

a) Before collecting the data
b) After collecting the data
c) After analyzing the data
d) It doesn't matter

2. What is one of the distinctions between a population parameter and a sample statistic?

a) A population parameter is only based on conceptual measurements, but a sample statistic is based on a combination of real and conceptual measurements
b) A sample statistic changes each time you try to measure it, but a population parameter remains fixed
c) A population parameter changes each time you measure it, but a sample statistic remains fixed across samples
d) The true value of a sample statistic can never be known but the true value of a population parameter can be known

3. Kuensel printed a survey in its daily issue and asked readers to fill it out and send it in. Over 500 readers did so. This type of sample is called

a) A cluster sample
b) A self-selected sample.
c) A stratified sample.
d) A simple random sample.

4. Which of the following variables is not categorical?

a) Tree species
b) Hair colour of a person
c) Choice on a test item: true or false
d) Height of a person

5. Which one of theses statistics is unaffected by outliers?

a) Mean
b) Standard deviation
c) Interquartile range
d) Range

6. What is the effect of an outlier on the on the value of a correlation coefficient?
a) An outlier will always decrease a correlation coefficient
b) An outlier will always increase a correlation coefficient
c) An outlier might either decrease or increase a correlation coefficient, depending on whether it is in relation to the other points
d) An outlier will have no effect on a correlation coefficient

7. One use of a regression line is

a) To determine if any x-values are outliers
b) To determine if any y-values are outliers
c) To determine if a change in x causes a change in y
d) To estimate the change in y for a one-unit change in x

8. Which of the following would indicate that a data set in not bell shaped?
a) The range is equal to 5 standard deviations
b) The mean is much smaller than the median
c) The range is larger than the interquartile range
d) There are no outliers

9. A list of 5 pulse rate is: 69, 65, 81, 72, 93. What is the median for this list?
a) 76
b) 72
c) 69
d) 81

10. Pick the choice that best completes the following sentence. If a relationship between two variables is called statistically significant, it means the investigators think the variables are

a) Related in the population represented by the sample
b) Not related in the population represented by the sample
c) Related in the sample due to chance alone
d) Very important

11. Suppose that the distribution of a set of scores has a mean of 35 and a standard deviation of 10. If 5 is added to each score, what will be the mean and the standard deviation of the of the distribution of new scores?
a) Mean=40 and standard deviation=15
b) Mean=40 and standard deviation=10

c) Mean=35 and standard deviation=10
d) Mean=35 and standard deviation=15

12. A member of parliament wants to know what the voters of his constituency think of proposed legislation on local election of Gups. He mails a questionnaire on the subject to an SRS of 500 voters in his constituency. 310 questionnaires have been returned, of which 200 support the legislation. The population is

a) The 500 voters receiving the questionnaire
b) The voters in his constituency
c) The 200 voters supporting the legislation
d) The 310 voters who returned the questionnaire

13. A poll conducted by Bhutan Times reported that 60 percent approved of the DPT governments work till date. The true percentage of Bhutanese citizens who approve of the government work is a

a) Population
b) Sample
c) Parameter
d) Statistic

14. A set of data has a mean that is much larger than a median. Which of the following statement is most consistent with this information?

a) The data set probably has a few low outliers
b) The histogram of the data set is skewed right
c) The histogram of the data set is skewed left
d) The histogram of the data is symmetric

15. Which of the following are true statements?

I. If bias is present in a sampling procedure, it can be overcome by dramatically increasing the sample size
II. There is no such thing as a “bad” sample
III. Sampling techniques that use probability techniques effectively eliminate bias

a) I only
b) II only
c) III only
d) None of the statements are true

16. Which of the following are true statements?

I. Voluntary response sample often under represent people with strong opinions

II. Convenience samples often lead to under coverage bias
III. Questionnaires with non neutral wording are likely to have response bias

a) I and II
b) I and III
c) II and III
d) I, II and III

17. A researcher plans to study the opinion of the general public on effectiveness of hospital service delivery. He obtains a simple random sample of 60 people as they leave the hospital on one Saturday afternoon. All of them agree to participate in the survey. Which of the following are true statements?

I. Proper use of chance as evidenced by the simple random sample makes this a well designed survey
II. The high response rate makes this a well-designed survey
III. Selection bias makes this a poorly designed survey

a) I only
b) II only
c) III only
d) I and II

18. A recent online poll posted on Bhutan times asked the question "Your ratings of the DPT government's performance". Of the more than 3500 respondents, 34.1% said very good. What does this show?

a) The survey is meaningless because of voluntary response bias.
b) No meaningful conclusion is possible without knowing something more about the characteristics of the respondents
c) The survey would have been more meaningful if the poll used a control group
d) The survey would have been more meaningful if they picked a random sample of the 3500 respondents

19. A magazine printed a survey in its monthly issue and asked readers to fill it out and send it in. Over 1000 readers did so. This type of sample is called

a) A cluster sample
b) A self-selected sample
c) A stratified sample
d) A simple random sample

20. The distribution of exam scores for CSCE candidates in 2008 was approximately normal with a mean of 538 and a standard deviation of 116. To calculate the probability that the average score of SRS of 20 candidates is 550 or higher we calculate the z-score as

a) $\dfrac{550-538}{116}$

b) $\dfrac{550-538}{116/\sqrt{20}}$

c) $\dfrac{550-538}{116/\sqrt{19}}$

d) None of the above

21. Is the mean height of the female graduates less than the male graduates? Let μ_F = the mean height of female graduates and μ_M = the mean height of male graduates. What are the appropriate null and alternative hypothesis?

a) $H_0: \mu_F - \mu_M = 0$ and $H_1: \mu_F - \mu_M \neq 0$

b) $H_0: \mu_F - \mu_M \neq 0$ and $H_1: \mu_F - \mu_M = 0$

c) $H_0: \mu_F - \mu_M \neq 0$ and $H_1: \mu_F - \mu_M < 0$

d) $H_0: \mu_F - \mu_M \neq 0$ and $H_1: \mu_F - \mu_M > 0$

22. There are 27 students in a class. Do you think the probability that two of them share the same birthday is about

a) $\dfrac{1}{1000}$

b) $\dfrac{1}{100}$

c) $\dfrac{1}{10}$

d) $\dfrac{1}{2}$

23. If X has a binomial distribution with parameters n and p, then:

a) $\mu_x = np$

b) $\mu_x = p$

c) $\mu_x = \sqrt{np}$

d) $\mu_x = \sqrt{p}$

24. If $\bar{x}$ is the mean of a random sample of size n from an infinite population with mean μ and standard deviation σ, then:

a) $\sigma_{\bar{x}} = \sigma$

b) $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

c) $\sigma_{\bar{x}} = \frac{\sigma}{n}$

d) None of the above

25. The formulae for the mean is given as $E(X) = \mu_x = \sum x_i p_i$. This formulae is used to find the mean of a

a) Continuous distribution
b) Discrete distribution
c) Both
d) None of the above

26. The following equation is provided: $\hat{y} = b_0 + b_1 x$. Here b_1 refers to

a) Dependent variable
b) Independent variable
c) Intercept
d) Slope

27. In regression, to measure the goodness-of-fit, it is better to use

a) r^2
b) Adjusted r^2
c) Both
d) None of the above

28. A set of possible values that a random variable can assume and their associated probabilities of occurrence are referred to as

a) Probability distribution
b) The expected return
c) The standard deviation
d) The coefficient of variation

29. A statistical measure of the variability of a distribution around its mean is referred to as

a) A probability distribution
b) The expected return
c) The standard deviation
d) Coefficient of variation

30. The ratio of the standard deviation of a distribution to the mean of that distribution is referred to as

a) A probability distribution
b) The expected return
c) The standard deviation
d) Coefficient of variation

Section A

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

1. Suppose a coin is flipped several times.
   a) What is the probability of two tails in two flips?
   b) What is the probability of three heads in three flips?

2. What does the term "Sample Space" mean? Give an example of a sample space with equiprobable outcomes.

3. A card is drawn at random from a standard deck of 52 cards. What is the probability of

   a) A queen of diamonds
   b) An ace
   c) Not an ace

4. If p=.75 in a Bernoulli process, what does q equal?

Section B

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

1.Integrity of organizations in providing service to users is emerging as a hot issue. If you are assigned to work in the Anti Corruption Commission and were tasked to design a survey to tackle integrity of service providers, how will you proceed? Give detailed survey methodology including (a) Planning the survey; (b) Questionnaire design; (c) Sampling methodology and (d) Tabulation and analysis plans.

2. The scores of ten students on both the mid term and final exam in a statistics class are recorded in the table below

Student #	1	2	3	4	5	6	7	8	9	10
Midterm	30	12	15	25	22	28	18	16	24	28
Final	27	13	10	17	19	25	20	12	22	26

   a) Draw a scatter plot
   b) Find the correlation (r) between midterm and final

c) Find the regression line for predicting finals scores from midterm. Draw this on your scatter plot.
d) Calculate the total variance in the final exam scores
e) Calculate the explained and unexplained variances in final exam scores. Use r^2

Hint: $r = \dfrac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$