ROYAL CIVIL SERVICE COMMISSION
BHUTAN CIVIL SERVICE EXAMINATION (BCSE) 2011
EXAMINATION CATEGORY: <u>TECHNICAL</u>

PAPER III: SUBJECT SPECIALIZATION PAPER FOR: **STATISTICS**

Date	**: October 30, 2011**
Total Marks	**: 100**
Examination Time	**: 2 hours 30 minutes**
Reading Time	**: 15 Minutes**

INSTRUCTIONS

1. Write your Roll Number clearly on the answer booklet in the space provided.

2. The first 15 minutes is being provided to check the number of pages, printing errors, clarify doubts and to read the instructions. You are **NOT PERMITTED TO WRITE** during this time.

3. Use either **Blue** or **Black** ink pen or ball point pen for the written part and **Pencils** for the sketches and drawings.

4. All answers should be written on the Answer Booklet provided. Candidates are not allowed to write anything on the question paper.

5. This Question Booklet consists of **9** pages including this page. It is divided into two sections – namely SECTION A and SECTION B.

6. **SECTION A** consists of two parts. **Part I and Part II.**

   Part I consists of 30 multiple choice questions carrying one (1) mark each and is **compulsory**. The answer of your choice should be clearly written **in whole** along with the question and option number on your answer booklet.

   Part II consists of four (4) short answer questions of five (5) marks each and all questions are compulsory.
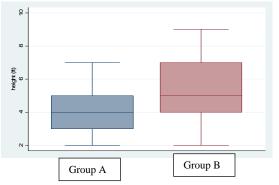
7. **SECTION B** consists of two **Case Studies**. Choose only **ONE** case study and answer the questions under your choice. Each case study carries fifty (50) marks in total.

Section A, Part I: 30 multiple choice questions of one mark each (30X1=30 marks)

[In this part, four answer choices (a, b, c, d) are provided for each question and you are to choose only one answer. Write the question number with the corresponding answer choice (either a, b, c or d) on the separate answer sheet].

1. A variable that can be ordered or ranked (e.g., excellent, good, fair) is called
   (a) Nominal
   (b) Ordinal
   (c) Continuous
   (d) Discrete

2. Which of the following measures of central tendency is affected by extreme values?
   (a) Mean
   (b) Median
   (c) Mode
   (d) All of the above

3. A distribution is left skewed when
   (a) Mean is equal to median
   (b) Mean is less than Median
   (c) Mean is greater than Median
   (d) None of the above

4. A parameter is a
   (a) Summary measure to describe a characteristic of the population
   (b) Summary measure to describe a characteristic of the sample
   (c) Totality of things under consideration
   (d) Portion of the population selected for analysis

5. Z score can be used to determine if a particular value is an outlier. If the mean is 5 and standard deviation is 3, what is the Z score for the value 20?
   (a) 1
   (b) 1.3
   (c) 3.4
   (d) 5

6. From the sample size of 10 children, $\sum x / n = 5$ and $\sum x^2 = 500$. The standard deviation of their age in years (using the maximum likelihood estimator) is
   (a) 25
   (b) 5
   (c) 10.5
   (d) 2.2

7. If P(X=0)=1/2 and P(X=1)=1/2. What is the expected value of X?
   (a) 0
   (b) 1/2
   (c) 1
   (d) 2/3

8. In a single throw of a single die, what is the probability of obtaining either a 3 or a 6?
   (a) 1/6
   (b) 1/3
   (c) 2/3
   (d) 1/2

9. A first step in data analysis is to plot the data.

   From the box plot, which of the following observations is false?
   (e) Median height of Group A is less than group B
   (f) Group A is more variable than group B
   (g) Lower quartiles in both groups are equal
   (h) Upper quartile in group A is less than group B

10. Which of the following is useful for visually identifying the outlier?
    (a) Normal (Q-Q) plot
    (b) Box plot
    (c) Histogram
    (d) All of the above

11. Shop A and shop B sell a same good with different prices but their standard deviations are exactly equal. You used a statistic- Coefficient of variation (C.V) and found out that shop A's C.V=10 and shop B's C.V=5. Which of the following statement is true?
    (a) Shop A and shop B have same variation
    (b) Shop A is more variable than shop B
    (c) Shop A is less variable than shop B
    (d) None of the above

12. Which of the following is an example of a continuous distribution?
    (a) Binomial
    (b) Exponential
    (c) Geometric
    (d) Poisson

13. An absent-minded person has 10 keys. One of the keys is for the main door. One night when he arrives his apartment, he selects a key at random. If it doesn't work, he replaces the key and selects at random from the 10 keys until he finally finds the right key. If X = the number of attempts the absent-minded person makes, what is the probability distribution of X?
    (a) Poisson distribution
    (b) Geometric distribution
    (c) Binomial distribution
    (d) Multinomial distribution

14. In a binomial experiment, there are only two outcomes, either "success" or "failure". If the probability of "success" is 0.6, what is the probability of "failure"?
    (a) 0.4
    (b) 0.6
    (c) 1
    (d) 0

15. A time series typically consists of the following components, except
    (a) Stable component
    (b) Trend component
    (c) Seasonal component
    (d) Non-random component

16. Which of the following is an unbiased estimator?
    (a) Expected value of the estimator is equal the parameter value
    (b) Variance of the estimator is small
    (c) Variance of the estimator is big
    (d) Mean squared error is small

17. Suppose that newspaper A reports the mean working hours of its employees is 10 hours with standard error of 1 hour while newspaper B reports the mean working hours of its employees is 10 hours with standard error of 2 hours. Which of the following is true?
    (a) Newspaper A's reporting is more precise
    (b) Newspaper B's reporting is more precise
    (c) Both newspapers are equally precise
    (d) None of the above

18. Which of the following is true about sampling error?
    (a) Selection bias is an example of a sampling error
    (b) Measurement error is an example of a sampling error
    (c) Sampling error is usually not reported in probabilistic term
    (d) The deviation between the sample estimate and true parameter value is the sampling error

19. A probability sample in which population units are partitioned, and then a probability sample of units is taken from each partition. This type of probability sampling is
    (a) Simple random sampling
    (b) Stratified random sampling
    (c) Systematic sampling
    (d) Quota sampling

20. If the covariance between two random variables X and Y is zero, then
    (a) X and Y are dependent
    (b) X and Y are independent
    (c) X and Y move in same direction
    (d) X and Y move in opposite direction

21. Which of the following is false about correlation coefficient?
    (a) Ranges between -1 to 1
    (b) Closer to 1, the stronger the positive linear relationship
    (c) Closer to -1, the stronger the negative linear relationship
    (d) Closer to 0, the weaker the causal effect

22. Which of the following is false about hypothesis tests?
    (a) Alternative hypothesis is the research hypothesis
    (b) Null hypothesis is the claim that needs to be proved
    (c) Alternative hypothesis is accepted when null hypothesis is rejected
    (d) Failing to reject null hypothesis does not mean that we accept it as true

23. A laboratory test shows that a person has a disease when infact he does not have it. In this case, the test
    (a) Makes Type I error
    (b) Makes Type II error
    (c) Increases the Power
    (d) Decreases the Power

24. If the significance level (α) of the test is fixed at 0.05 and P-value obtained is 0.03, what decision will you make?
    (a) Reject the null hypothesis
    (b) Not reject the null hypothesis
    (c) Increase the significance level
    (d) Decrease the significance level

25. Misuses of hypothesis tests are common in practice and should be avoided. All of the following are the common misuses, except
    (a) Equating statistical significance with practical significance
    (b) Hunting for significance
    (c) Ignoring significant results
    (d) Drawing conclusions from nonrandom and observational data

26. Which of the following is false about confidence interval?
    (a) Stated in terms of level of confidence
    (b) Provides a point estimate
    (c) Never 100% sure
    (d) Provides information about closeness of point estimate to unknown population parameter

27. Which of the following is an example of a non-parametric test?
    (a) Chi-Square test
    (b) Fisher's test
    (c) Student's t-test
    (d) Sign test

28. An investigator wants to test two samples, one of boys and one of girls. He wants to know whether the results for girls are significantly more variable than for boys. If the sample size for both boys and girls with their variances are given, which test statistic will be appropriate?
    (a) Chi-Square test
    (b) Fisher's test
    (c) Student's t-test
    (d) Sign test

29. In a simple linear regression ($\gamma_i = \beta_0 + \beta_1 x + \varepsilon_i$), where ε_i is called a random error. Usually all of the following are the assumptions made about the random error, except
    (a) Independent and identically distributed
    (b) Normally distributed
    (c) Expected value is zero
    (d) Known variance

30. Design of Experiment is an important field in statistics. An experiment was conducted to evaluate 3 treatments (A, B, C). A total of 10 experimental units with $n_A = 4$, $n_B = 2$, $n_C = 4$ are assigned to the 3 treatments. How many different arrangements are possible in assigning the 3 treatments to the 10 experimental units (Hint: use combination formula)
    (a) 3,150
    (b) 225
    (c) 12,600
    (d) 1,450

Section A, Part II: Four short answer questions (4X5=20 marks)

1. Discuss about the measures of central tendency. Give examples.

2. Suppose the expenditures of a person, denoted by X, are given by the following probability function.

X	750	800	950	1000
P(X=x)	0.3	0.2	0.3	0.2

    (a) Find the person's expected expenditures?
    (b) If the earnings were all 1000 units with equal probability of 0.25, what would have been the person's expected expenditures?

3. Describe the Central Limit Theorem and its importance.

4. Describe about simple random sampling and cluster sampling

Section B: From the two case studies, you are to attempt only **one**. (1X50=50 marks)

1. A researcher desires to study the average annual income of 30-year-old men in Bhutan. He draws a sample from the 30-year-old men in Thimphu to make his study.

    (a) Describe the sampled and target population.
    (b) What kind of sampling does the researcher uses?
    (c) Is the sampling biased? Explain.
    (d) What problems arise in drawing conclusion from the data he collects?
    (e) If you are an adviser to the researcher, how will you design the survey?
    (f) Assuming that the new design is representative of Bhutan; can he claim that average annual income is exactly equal to the true average annual income of the target population? Explain.
    (g) If the average income obtained is Nu. 80,000 and standard error is Nu. 2,000, construct a 95% confidence interval (assume the annual income has Gaussian/normal distribution).
    (h) Interpret the 95% confidence interval based on the values obtained in part g such that non-statistical persons can also understand.(If you could not work on part g, simply interpret the 95% CI without using the values)

2. You believe that those who are vegetarians tend to be younger people. To test this theory, you have obtained the following distribution of the number of vegetarians. An appropriate null hypothesis for a statistical test would be that equal numbers in all age groups are vegetarians. Use this theory to answer the questions. *(Chi-square distribution table is provided on page 9)*

Age group in years	15-19	20-24	25-29	30-34	35-39	40-44	45-49
No. of vegetarians	40	35	32	10	13	13	4

(a) Formulate the null and alternative hypotheses
(b) Set a level of significance
(c) If you decide to use the chi-square statistic as the test statistic, calculate the degrees of freedom
(d) If the null hypothesis is true, what is the expected number of vegetarians in each group (given that the total vegetarians =147)
(e) Compute the chi-square statistic
(f) Do you reject or fail to reject the null hypothesis based on value obtained in part e (if you could not work on part e, assume that the chi-square statistic obtained is less than the critical value)?
(g) What do you finally conclude?
(h) If you change (increase or decrease) the level of significance, will your conclusion differ?

Chi Square Distribution Table

d.f.	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$	$\chi^2_{.001}$
1	1.32	2.71	3.84	5.02	6.63	7.88	10.8
2	2.77	4.61	5.99	7.38	9.21	10.6	13.8
3	4.11	6.25	7.81	9.35	11.3	12.8	16.3
4	5.39	7.78	9.49	11.1	13.3	14.9	18.5
5	6.63	9.24	11.1	12.8	15.1	16.7	20.5
6	7.84	10.6	12.6	14.4	16.8	18.5	22.5
7	9.04	12.0	14.1	16.0	18.5	20.3	24.3
8	10.2	13.4	15.5	17.5	20.1	22.0	26.1
9	11.4	14.7	16.9	19.0	21.7	23.6	27.9
10	12.5	16.0	18.3	20.5	23.2	25.2	29.6
11	13.7	17.3	19.7	21.9	24.7	26.8	31.3
12	14.8	18.5	21.0	23.3	26.2	28.3	32.9
13	16.0	19.8	22.4	24.7	27.7	29.8	34.5
14	17.1	21.1	23.7	26.1	29.1	31.3	36.1
15	18.2	22.3	25.0	27.5	30.6	32.8	37.7
16	19.4	23.5	26.3	28.8	32.0	34.3	39.3
17	20.5	24.8	27.6	30.2	33.4	35.7	40.8
18	21.6	26.0	28.9	31.5	34.8	37.2	42.3
19	22.7	27.2	30.1	32.9	36.2	38.6	32.8
20	23.8	28.4	31.4	34.2	37.6	40.0	45.3
21	24.9	29.6	32.7	35.5	38.9	41.4	46.8
22	26.0	30.8	33.9	36.8	40.3	42.8	48.3
23	27.1	32.0	35.2	38.1	41.6	44.2	49.7
24	28.2	33.2	36.4	39.4	32.0	45.6	51.2
25	29.3	34.4	37.7	40.6	44.3	46.9	52.6
26	30.4	35.6	38.9	41.9	45.6	48.3	54.1
27	31.5	36.7	40.1	43.2	47.0	49.6	55.5
28	32.6	37.9	41.3	44.5	48.3	51.0	56.9
29	33.7	39.1	42.6	45.7	49.6	52.3	58.3
30	34.8	40.3	43.8	47.0	50.9	53.7	59.7
40	45.6	51.8	55.8	59.3	63.7	66.8	73.4
50	56.3	63.2	67.5	71.4	76.2	79.5	86.7
60	67.0	74.4	79.1	83.3	88.4	92.0	99.6
70	77.6	85.5	90.5	95.0	100	104	112
80	88.1	96.6	102	107	112	116	125
90	98.6	108	113	118	124	128	137
100	109	118	124	130	136	140	149

Table from Ronald J. Wonnacott and Thomas H. Wonnacott,
Statistics: Discovering Its Power, New York: John Wiley and Sons, 1982, p.352.