

Semester End Examination Autumn 2018
Royal University of Bhutan
Paro College of Education
Paro: Bhutan

Module: MAT 408, Descriptive Statistics
Writing Time: Three Hours

Program: B.Ed (Secondary) **Level:** IV
Full Mark: 100

Instructions:

In this paper there are two sections, A and B. Do not write for the first 15 minutes. Use this time in reading questions. You will get three hours for answering the questions. Write the answers to all the questions in the answer sheets provided. Read the directions to each section and each question carefully before answering the questions. You are allowed to use scientific calculator, **fx-82** and **fx-100** beside other writing materials.

SECTION A
ONE Question-20 Marks

Directions: Each question under this section is followed by four alternative possible answers. Choose the correct answer and write it in the answer sheets provided.

Question 1

- a. The process of dealing with quantitative data analysis mainly involves the following statistical tools, EXCEPT
- A averages.
 - B dispersion.
 - C tabulation.
 - D correlation.
- b. Which formula is used for finding the mean of a continuous distribution using step-deviation method?

- A $\bar{X} = \frac{\sum X_i}{N}$
- B $\bar{X} = A + \frac{\sum f_i d_i}{N}$
- C $\bar{X} = \frac{\sum f_i X_i}{N}$
- D $\bar{X} = A + \frac{\sum f_i d_i}{N} \times c$

- c. What is the most relevant measure of dispersion when we have an open-ended distribution?
- A Range
 - B Mean deviation
 - C Quartile deviation
 - D Standard deviation
- d. If a distribution is normal, the value of coefficient of kurtosis is always equal to
- A 3.00.
 - B 2.00.
 - C 0.00.
 - D 1.00.
- e. The correlation coefficient value which indicates a very high negative correlation between two variables X and Y is
- A $r_{xy} = - 8.80$
 - B $r_{xy} = - 0.18$
 - C $r_{xy} = - 8.00$
 - D $r_{xy} = - 0.89$
- f. What formula gives the correlation coefficient between the two variables X and Y?
- A $r = \sqrt{b_{xy} / b_{yx}}$
 - B $r = \sqrt{b_{yx} / b_{xy}}$
 - C $r = b_{yx} . b_{xy}$
 - D $r = b_{xy} . b_{yx}$
- g. What is NOT the example of secondary data?
- A Civil census documents
 - B Research field diaries
 - C Police crime records
 - D School official documents
- h. The measure of dispersion which involves the minimal use of a given data set is
- A mean deviation.
 - B standard deviation.
 - C quartile deviation.
 - D inter-quartile range.

- i. If Spearman's ranking, $R = 0.77$, $n = 15$, what is the value of ΣD^2 ?
- A 46.38
B 46.00
C 46.05
D 8.250
- j. The central moment of a random variable X , which gives us the value of variance is calculated using the formula
- A $\frac{\sum f_i(X_i - \bar{X})}{N}$
B $\frac{\sum f_i(X_i - \bar{X})^2}{N}$
C $\frac{\sum f_i(X_i - \bar{X})^3}{N}$
D $\frac{\sum f_i(X_i - \bar{X})^4}{N}$
- k. Two concepts involve in determining the width of a continuous frequency distribution are
- A class mark and class boundary.
B range and assumed number of classes.
C true class limit and exact number of classes.
D upper limit and lower limit of class intervals.
- l. The regression coefficient for the line of regression Y on X is given by the formula
- A $\frac{r\sigma_x}{\sigma_y}$
B $\frac{r\sigma_y}{\sigma_y}$
C $\frac{r\sigma_y}{\sigma_x}$
D $\frac{r\sigma_x}{\sigma_x}$
- m. The intersecting point of the 'less than' ogive and the 'more than' ogive in general gives the value of
- A mean.
B mode.
C median.
D geometric mean.

- n. Given $r_{xy} = 0.38$, $\text{Cov}(x,y) = 10.2$, $\sigma_x^2 = 16$, what is the value of σ_y ?
- A 10.2
 - B 6.80
 - C 0.38
 - D 36.0
- o. Which graph can be used as an alternative to the frequency distribution table for organizing data?
- A Pie-graph
 - B Histogram
 - C Stem-and-leaf plot
 - D Box-and-whisker plot
- p. The square root of the variance of a data set gives the value of
- A mean deviation.
 - B quartile deviation.
 - C standard deviation.
 - D coefficient of variation.
- q. Which measure of central tendency is sensitive to all the scores in a distribution?
- A Mean
 - B Median
 - C Mode
 - D Quartiles
- r. In order to represent a data set in a box-and-whisker plot, we need to find the following, EXCEPT the
- A median.
 - B sixth decile.
 - C first quartile.
 - D third quartile.
- s. Given that $\Sigma d^2 = 30$, and $n = 10$, what is the value of R ?
- A 0.820
 - B 0.800
 - C 0.082
 - D 8.200
- t. The data showing the changes over a period of time can be best represented using a
- A bar graph.
 - B pie-chart.
 - C line graph.
 - D histogram.

SECTION B
FIVE Questions – 80 Marks

Instructions: There are seven questions in this section. Attempt any **FIVE** questions. Each question carries 16 marks. The intended mark for each sub-question is given in the brackets.

Question 2

- a. What you understand by ‘primary’ and ‘secondary’ data? Explain with appropriate examples of each. [6]
- b. What are six procedural steps involved in the process of dealing with the quantitative data? List them in order. [3]
- c. How is ‘Statistics’ as a science of information practically useful in our everyday life? Explain four of its uses. [4]
- d. “It becomes necessary and important for us to group the given data set in the process of conducting statistical analysis.” Justify the statement. [3]

Question 3

- a. The number of saplings planted by PCE students in different years on June 2 is given below:

Years	2000	2001	2002	2003	2004	2005	Total
No. of trees planted	400	450	700	750	900	1500	4700

Outline the procedural steps for constructing a bar graph using the above data. [5]

- b. The frequency distribution given below shows ranges of English test scores of 70 undergraduate students of a college. [7]

Class-intervals	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99
# of students	1	2	4	4	7	9	16	10	7	6	4

- i. Express the frequency distribution as its relative frequency, cumulative frequency and cumulative percentage distributions.
 - ii. Represent the above data in the form of a histogram.
- c. “Representing data using the stem-and-leaf plot is practically not feasible.” Do you agree or disagree with this statement? Justify your view. [4]

Question 4

- a. At a health club, 80% of the members are men and 20% of the members are women. If the average age of the men is 30 and the average age of the women is 40, what is the average age of all the members? [5]
- b. Find the median and the inter-quartile range of the given data set: 12, 15, 23, 14, 25, 15, 24, 18, 15, 16, 26. [5]
- c. A distribution consists of components with frequencies 28, 36 and 28 having their means 2.9, 5.6 and 5.6 respectively. Prove that the mean of the combined distribution is 4.78 approximately. [6]

Question 5

- a. Describe the relationship between standard deviation and variance of the given data set. [3]
- b. Records were kept on three employees: A, B and C in a factory. The study gives the following data. [5]

Employees	Mean	SD
A	23	1.45
B	45	5.86
C	32	3.54

- i. Which employee was the most productive?
- ii. Which employee was the most consistent?
- iii. What measure did you choose to answer the question ii? Why?
- c. The arithmetic mean and standard deviation of a series of 20 items were calculated by students as 20 cm and 5 cm respectively. But in the process of calculating them, an item 13 was misread as 30. Find the correct mean and standard deviation. [8]

Question 6

- a. Explain the term 'skewness' using appropriate illustrations. Also explain how we can interpret the Bowley's coefficient of skewness. [5]
- b. Find the standard deviation and kurtosis of the following series by the method of moments, and comment on the nature of the distribution. [5]

Classes	0-10	10-20	20-30	30-40	40-50
Frequency	10	20	40	20	10

- c. In a distribution, Pearson's coefficient of skewness is - 0.6 and its mean is 65, and median is 70. Find the value of its mode and coefficient of variation. [6]

Question 7

- a. For the following data, show that the correlation coefficient r is 0.92. [5]

X_i	6	2	10	4	8
Y_i	9	11	5	8	7

- b. The coefficient of rank correlation of marks obtained by 10 students in Statistics and Science was found to be 0.8. It was later found that the difference in ranks in two subjects obtained by one of the students was wrongly taken as 7 instead of 9. Find the correct rank correlation coefficient. [6]
- c. The coefficient of correlation between two variables X and Y is 0.38. Their covariance is 10.2. If the variance of variable X is 16, find the standard deviation of Y series. [5]

Question 8

- a. Using the following data, find the two lines of regression and from them compute Karl Pearson's coefficient of correlation. [8]

$$\begin{array}{lll} \sum X_i = 250 & \sum Y_i = 300 & \sum X_i Y_i = 7900 \\ \sum X_i^2 = 6500 & \sum Y_i^2 = 10000 & N = 10 \end{array}$$

- b. Given the following data, calculate the SE of the estimate of X for the given value of Y and the SE of the estimate of Y for the given value of X . [8]

X_i	4	8	2	5	6
Y_i	9	12	1	5	8

Key formulae to be used

$$\bar{X}_{12} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

$$\sigma_{12} = \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2 + n_1 (\bar{X}_1 - \bar{X}_{12})^2 + n_2 (\bar{X}_2 - \bar{X}_{12})^2}{n_1 + n_2}}$$

OR

$$\sigma_{12} = \sqrt{\frac{n_1 \sigma_1^2 + n_2 \sigma_2^2 + n_1 (\bar{X}_1 - \bar{X}_{12})^2 + n_2 (\bar{X}_2 - \bar{X}_{12})^2}{n_1 + n_2}}$$

$$r_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \cdot \sqrt{\sum (y - \bar{y})^2}} \quad r_{xy} = \frac{N \sum XY - \sum X \cdot \sum Y}{\sqrt{N \sum X^2 - (\sum X)^2} \cdot \sqrt{N \sum Y^2 - (\sum Y)^2}}$$

$$r_{xy} = \frac{\sum xy - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{\sum x^2 - n \cdot \bar{x}^2} \cdot \sqrt{\sum y^2 - n \cdot \bar{y}^2}}$$

$$r_{xy} = \frac{\sum XY - n \cdot \bar{X} \cdot \bar{Y}}{\sqrt{\sum X^2 - n \cdot \bar{X}^2} \cdot \sqrt{\sum Y^2 - n \cdot \bar{Y}^2}}$$

$$\ell = 1 - 6 \left[\sum d^2 + \frac{M_1^3 - 3M_1}{12} + \frac{M_2^3 - 3M_2}{12} + \frac{M_3^3 - 3M_3}{12} + \dots \right]$$

$$R / \ell = 1 - \frac{6 \sum D^2}{n(n^2 - 1)}$$

$$b_{yx} = \frac{n \cdot \sum xy - (\sum x)(\sum y)}{n \cdot \sum x^2 - (\sum x)^2}$$

$$b_{xy} = \frac{n \cdot \sum xy - (\sum x)(\sum y)}{n \cdot \sum y^2 - (\sum y)^2}$$

$$b_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (y - \bar{y})^2}$$

$$b_{yx} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

$$(y - \bar{y}) = \frac{r \sigma_y}{\sigma_x} \cdot (x - \bar{x})$$

$$(x - \bar{x}) = \frac{r \sigma_x}{\sigma_y} \cdot (y - \bar{y})$$

$$a = \frac{(\sum X^2) \cdot (\sum Y) - (\sum X)(\sum XY)}{n \sum X^2 - (\sum X)^2}$$

$$b = \frac{n \sum XY - (\sum X)(\sum Y)}{n \cdot \sum X^2 - (\sum X)^2}$$