Semester End Examination Autumn 2017
Royal University of Bhutan
Paro College of Education
Paro: Bhutan

Module: MAT408, Descriptive Statistics **Program**: B.Ed (Secondary) **Level:** IV
Writing Time: Three Hours **Full Marks:** 100

Instructions:
In this paper there are two sections, A and B. You are required to answer ALL questions from section A and only FIVE questions from section B. Do not write for the first 15 minutes; use this time for reading the 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. Do not leave the examination hall before you are certain that all the questions as directed in the paper, have been answered. You are allowed to use calculator for this paper.

SECTION A (20 marks)

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

Question 1

a. The following are the examples of a continuous variable, EXCEPT the

A time of flight of a missile.

B temperature of a person's palm.

C number of sisters in a family.

D response time in a test of reflexes.

b. The formula for finding the arithmetic mean of a set of ungrouped data is

A $\overline{X} = \frac{\sum Xi}{N}$

B $\overline{X} = A + \frac{\sum fidi}{N}$

C $\overline{X} = \frac{\sum fiXi}{N}$

D $\overline{X} = A + \frac{\sum fidi}{N} \times c$

c. The square of the standard deviation of a set of data gives us the value of

A coefficient of variation.

B median.

C mode.

D variance.

d. A normal distribution assumes the value of coefficient of kurtosis equals to

A 3.0

B 2.0

C 0.1

D 1.0

e. The correlation coefficient value which shows a very weak negative relationship between two variables X and Y is:

A r_{xy} = -0.19

B r_{xy} = -0.09

C r_{xy} = -9.00

D r_{xy} = -0.90

f. The regression coefficient for the line of regression Y on X is given by the formula

A $\frac{r\sigma_y}{\sigma_y}$

B $\frac{r\sigma_y}{\sigma_y}$

C $\frac{r\sigma_y}{\sigma_x}$

D $\frac{r\sigma_x}{\sigma_x}$

g. Which of the following is not the example of secondary data?

A Civil registry records

B Researcher's field notes

C Crime records

D Any official documents

h. The crudest measure of dispersion of a given set of distribution is the

A standard deviation.

B range.

C quartile deviation.

D inter-quartile range.

i. The most appropriate graph to represent the budget allotments for various departments of a government is

A line graph.

B pie chart.

C bar-graph.

D histogram.

j. The central moment of a random variable X, which gives us the value of variance is calculated using the formula

A $\frac{\sum f(Xi - \overline{X})}{N}$

B $\frac{\sum f(Xi - \overline{X})^2}{N}$

C $\frac{\sum fi(Xi - \overline{X})^3}{N}$

D $\frac{\sum fi(Xi - \overline{X})^4}{N}$

SECTION B

FIVE Questions – 80 Marks

Instructions: There are SEVEN questions in this section. Answer any **FIVE** questions. All questions carry equal marks. The intended mark for each sub-question is given in the brackets.

Question 2

a. Explain the two major methods of quantitative data collection. (4)

b. Enlist the general procedures for dealing with the statistical information? (4)

c. Describe the practical significance of 'Statistics' in our everyday life. (5)

d. When is it necessary for us to group the given data set in the process of conducting statistical analysis? (3)

Question 3

a. What is a frequency distribution? Explain briefly. Also outline the steps involved in constructing a frequency distribution table for a continuous data. (6)

b. Prepare a frequency distribution taking an appropriate class-interval for the following data, which are the weights of 30 children in a certain class. Also represent them in a histogram. (7)

48	51	43	46	41	49	43	45	46	49
48	46	42	49	43	42	50	45	43	48
43	42	46	45	40	44	55	42	53	44

c. "Representing data using the stem-and-leaf plot is not that practically useful." Do you agree or disagree with this statement? Justify your view. (3)

Question 4

a. Find the range, inter-quartile range and median for a given data set of 12, 15, 23, 14, 25, 15, 24, 18, 15, 16, 26. (6)

b. Calculate the value of mean and mode for the following data. (5)

Class-intervals	0-8	8-16	16-24	24-32	32-40
Frequency	6	9	12	10	5

c. A distribution consists of two components with frequencies 50 and 100, having their means 54.1 and 50.3, and standard deviations 8 and 7 respectively. Prove that the standard deviation of the combined distribution is 7.564. (5)

Question 5

a. Records were kept on two employees, Dorji and Tenzin in an office. The study gives the following data. Which employee was more productive? Why? (4)

Employee	Mean	SD
Dorji	45	5.86
Tenzin	32	3.54

b. Calculate the quartile deviation & its coefficient from the following data. (4)

Weights/kg	12	13	14	18	16	11
Frequency	3	4	10	6	4	3

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 by calculating them, an item 13 was misread as 30. Find the correct mean and standard deviation. (8)

Question 6

a. The following data represents the weight (kg) of 100 students in a certain school. Find the moment coefficient of skewness and kurtosis, and comment on the nature of the distribution. (7)

Xi	60-62	63-65	66-68	69-71	72-74
fi	5	18	42	27	8

b. Explain the term 'kurtosis' using an appropriate illustration and highlight the interpretation of its coefficient values. (4)

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. (5)

Question 7

a. What is correlation coefficient? Explain the interpretation of the values of the correlation coefficient? (5)

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. Given the following information relating to frequency distribution of 10 observations: $\overline{X} = 5$, $\overline{Y} = 4$, $\sum X^2 = 298$, $\sum Y^2 = 176$ and $\Sigma XY = 213$, find the value of r_{xy}. (5)

Question 8

a. Prove that the arithmetic mean of the regression coefficient is greater than the correlation coefficient. (8)

b. In a correlation study the following data were obtained. (8)

	X	Y
Mean	65	67
SD	2.5	3.5
Correlation Coefficient	0.8	0.8

Find the regression equations that are associated with the above values.

Key formulae to be used

$$r_{xy} = \frac{\Sigma\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)}{\sqrt{\Sigma\left(X-\overline{X}\right)^2}.\sqrt{\left(Y-\overline{Y}\right)^2}} \qquad r_{xy} = \frac{N\sum XY - \sum X.\sum Y}{\sqrt{N\sum X^2-(\sum X)^2}.\sqrt{N.\sum Y^2-(\sum Y)^2}}$$

$$\mathrm{r_{xy}} = \frac{\Sigma xy - n.x.y}{\sqrt{\Sigma x^2 - n.x^2}.\sqrt{\Sigma y^2 - n.y^2}} \qquad r_{xy} = \frac{\sum XY - n.\overline{X}.\overline{Y}}{\sqrt{\sum X^2 - n.\overline{X}^2}.\sqrt{\sum Y^2 - n.\overline{Y}^2}}$$

$$\ell = 1-6\left[\sum d^2 + \frac{M_1^3 - m1}{12} + \frac{M_2^3 - M_2}{12} + \frac{M_3^3 - M_3}{12} + \ldots\ldots\right] \qquad R/\ell = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

$$\mathrm{b_{yx}} = \frac{n.\Sigma xy - (\Sigma x)(\Sigma y)}{n.\Sigma x^2 - (\Sigma x)^2} \qquad \mathrm{b_{xy}} = \frac{n.\Sigma xy - (\Sigma x)(\Sigma y)}{n.\Sigma y^2 - (\Sigma y)^2}$$

$$bxy == \frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sum(Y-\overline{Y})^2} \qquad byx = \frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sum(X-\overline{X})^2}$$

$$(Y-\overline{Y}) = \frac{r\sigma y}{\sigma x}.(X-\overline{X}) \qquad (X-\overline{X}) = \frac{r\sigma x}{\sigma y}.(Y-\overline{Y})$$

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

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

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