Tutorial letter 011/0/2021

Descriptive Statistics and Probability

STA1501

Year module

Department of Statistics

Assignment 1 Questions

Define tomorrow.

UNISA | university of south africa

Please study Chapters 1 to 5 and corresponding parts of the study guide before answering the following questions.

Remember that in this module, to get examination admission you must have a semester mark of at least 40%. You semester mark is calculated as the average of the five best marks of your 6 assignments.

This assignment consists of two parts, namely; the true/false question (part A) – 20 marks and the written questions (part B) – 100 marks, thus the total for the assignment is 120 marks.

PART A: 20 Marks

State whether the following statements are true or false.

QUESTION 1

A qualitative variable for which there is some kind of ranking for the different values is called an interval variable.

QUESTION 2

In general, probability samples are samples in which the elements to be selected are drawn on the basis of probability in such a way that each element in the population has the same probability of being selected obtained as part of the sample.

QUESTION 3

All the individual data values can be read out from a stem-and-leaf display.

QUESTION 4

As a result of surveying 1000 first year students at UNISA, it was found that 32% owned a laptop. The value 32% is an example of a parameter.

QUESTION 5

A histogram can be used to summarise ordinal qualitative data.

QUESTION 6

The interquartile range of a data set is always half of the range of the data set.

QUESTION 7

Measures of position are used to describe the position a specific data value possesses in relation to the mean of the data.

QUESTION 8

Bias in self-selected sampling can be reduced by increasing the sample size.

QUESTION 9

Assume that the least squares line fitted trough a scatter diagram of two variables, X and Y, is given by $\widehat{y} = 5.2 - 0.3x$. Then the coefficient of correlation for the two variables is negative.

QUESTION 10

The coefficient of correlation for a data set with two variables is always smaller than the covariance of the same data set.

QUESTION 11

A data set can have several mode values.

QUESTION 12

The median of a data set is always equal in value to one of the original data values.

QUESTION 13

A box plot is plotted using the values of the mean, the first quartile and the second quartile.

QUESTION 14

Sampling error can be reduced by increasing the sample size.

QUESTION 15

Cluster sampling is an example of simple random sampling.

QUESTION 16

If a data set consists of only negative values, then its standard deviation is negative.

QUESTION 17

If a distribution is symmetric, then the distance between first and second quartiles if roughly the same as the distance between the second and third quartiles.

QUESTION 18

In a simple random sample, all members of the population have the same probability of being selected.

QUESTION 19

A positively skewed data set is one where most of the values are positive.

QUESTION 20

If the covariance between two data variables is zero then the scatter diagram shows all the points lined along a horizontal line.

PART B: 100 Marks

Study Chapter 1, 2, 3, 4 and 5 of the text book before attempting these questions.   Give full calculations, and use correct mathematical and statistical notation.

QUESTION 21

A group of UNISA students were asked what device they mostly use for accessing the myUnisa module sites. The responses were coded as 1 for laptop computer, 2 for desktop computer and 3 for a smart phone. The following data was collected: 1, 2, 1, 1, 1, 1, 2, 2, 3, 3, 1, 2, 2, 3, 1, 3, 3, 3, 3, 3, 2, 1, 3, 1, 1, 1, 2, 2, 3, 2.

  (a) Construct a table showing the frequency distribution and the relative frequency distribution for this data.   (8)

  (b) Draw a bar chart of the data. Ensure you label the axes properly and give the title of the graph.   (6)

  (c) What does the data tell us about the use of the different devices by the students?   (2)

[16]

QUESTION 22

The following data set shows the table of frequencies obtained when a group of men and women were asked to select their favourite soft drink out of three different brands, labelled here as A, B and C.

Brand	A	B	C
Male	23	7	11
Female	40	47	19

(a) Draw a cross-classification table of row proportions. (6)

(b) Draw a two-dimensional bar chart indicating the preferences of the different brands for males and females. Remember to label the axes and give a title to the graph. (6)

(c) Interpret your findings. (3)

[15]

QUESTION 23

The scores on a mathematics test for a group 30 learners are as follows:

11,	74,	62,	67,	29,	36,	55,	55,	72,	33,
63,	82,	10,	96,	78,	14,	68,	88,	26,	66,
25,	87,	53,	63,	63,	35,	57,	19,	29,	46.

(a) Construct a stem and leaf plot of the data. (5)

(b) Calculate the mean, median and mode of the data values. (7)

(c) Comment on the skewness of the data. (1)

(d) Calculate the standard deviation of the marks. (5)

(e) Determine the coefficient of variation of the marks. (2)

(f) Comment on how well the Empirical Rule applies to this data set. (5)

[25]

QUESTION 24

The following data set gives the distance a group of employees travel to and from work during one week.

$$303 \quad 53 \quad 49 \quad 15 \quad 157 \quad 227 \quad 146 \quad 29 \quad 189$$
$$88 \quad 97 \quad 44 \quad 46 \quad 232 \quad 229 \quad 117 \quad 87 \quad 75$$

(a) Find the first, second and third quartiles of the data. (9)

(b) Does the data set have outliers? Justify your answer! (2)

(c) Construct a box plot of the data and comment on the skewness of the data. (7)

[18]

QUESTION 25

We wish to investigate the relationship between income and education. A sample of 8 individuals is selected at random, and their monthly income (in thousands of rands) and education (in years) are shown below:

Education	15	10	7	8	10	11	14	12
Income	23	18	15	15	21	28	29	27

(a) Determine the dependent and independent variables. (2)

(b) Draw a scatter diagram for these data. (5)

(c) Describe the relationship between income and education. (2)

(d) Calculate S_x^2, S_y^2 and S_{xy}. (10)

(e) Determine the least squares line in the form of $\hat{y} = b_0 + b_1 x$. (3)

(f) Use the estimated least squares line to predict the income of a person with 17 years of education. (1)

(h) Calculate the coefficient of determination and interpret it. (3)

[26]

Total marks [120]