(*Knowledge for Development*)

KIBABII UNIVERSITY

(KIBU)

UNIVERSITY EXAMINATIONS
2018/2019 ACADEMIC YEAR
END OF SEMESTER EXAMINATIONS
YEAR THREE SEMESTER TWO EXAMINATIONS

FOR THE DEGREE OF
BACHELORS OF SCIENCE
(INFORMATION TECHNOLOGY)

COURSE CODE : BIT 225

COURSE TITLE : DATA ANALYSIS

DATE: 29/05/2019 **TIME: 9.00A.M. – 11.00A.M.**

INSTRUCTIONS TO CANDIDATES

ANSWER QUESTIONS ONE AND ANY OTHER TWO

QUESTION ONE (COMPULSARY) [30 MARKS]

1. What is the advantage of using SPSS over calculating statistics manually?
   a) It equips you with a useful transferable skill.
   b) It reduces the chance of making errors in your calculations.
   c) Many researchers use SPSS as it is a recognised software package.
   d) All of the above.
2. In SPSS, what is the "Data Viewer"?
   a) A table summarizing the frequencies of data for one variable.
   b) A spreadsheet into which data can be entered.
   c) A dialog box that allows you to choose a statistical test.
   d) A screen in which variables can be defined and labeled.
3. How is a variable name different from a variable label?
   a) It is shorter and less detailed.
   b) It is longer and more detailed.
   c) It is abstract and unspecific.
   d) It refers to codes rather than variables.
4. What does the operation "Recode Into Different Variables" do to the data?
   a) Replaces missing data with some random scores.
   b) Reverses the position of the independent and dependent variable on a graph.
   c) Redistributes a range of values into a new set of categories and creates a new variable.
   d) Represents the data in the form of a pie chart.
5. How would you use the drop-down menus in SPSS to generate a frequency table?
   a) Open the Output Viewer and click: Save As; Pie Chart
   b) Click on: Analyze; Descriptive Statistics; Frequencies
   c) Click on: Graphs; Frequencies; Pearson
   d) Open the Variable Viewer and recode the value labels
6. Why might you tell SPSS to represent the "slices" of a pie chart in different patterns?
   a) Because the program tends to crash if you ask it to use colour.
   b) Because the patterns form symbolic visual images of different social groups.
   c) In order to make full use of the facilities that SPSS can offer.
   d) If you do not have a colour printer, it makes the differences between slices clearer.
7. When cross-tabulating two variables, it is conventional to:
   a) represent the independent variable in rows and the dependent variable in columns.
   b) assign both the dependent and independent variables to columns.
   c) represent the dependent variable in rows and the independent variable in columns.
   d) assign both the dependent and independent variables to rows.
8. In which sub-dialog box can the Chi Square test be found?
   a) Frequencies: Percentages
   b) Crosstabs: Statistics
   c) Bivariate: Pearson
   d) Gender: Female
9. To generate a Spearman's rho test, which set of instructions should you give SPSS?
   a) Analyze; Crosstabs; Descriptive Statistics; Spearman; OK
   b) Graphs; Frequencies; [select variables]; Spearman; OK
   c) Analyze; Compare Means; Anova table; First layer; Spearman; OK
   d) Analyze; Correlate; Bivariate; [select variables]; Spearman; OK
10. How would you print a bar chart that you have just produced in SPSS?
    a) In Output Viewer, click File, Print, select the bar chart and click OK
    b) In Variable Viewer, open bar chart, click File, Print, OK
    c) In Chart Editor, click Descriptive Statistics, Print, OK
    d) In Data Editor, open Graphs dialog box, click Save, OK
11. A dependent variable refers to:
    A. the variable being manipulated or varied in some way by the researcher.

B. the variable which shows us the effect of the manipulation.
C. a variable with a single value which remains constant in a particular context.
D. the experimental condition.

12. Complete the following sentence:
Sometimes the difference a researcher has observed in a dependent variable as a result of manipulating the independent variable may not be due to the manipulation but due to:
A. categorical variables. B. confounding variables
C. dichotomous variables D. spurious variables.

13. Variables are:
A. the main focus of research in science B. something that can vary in terms of precision
C. something that we can measure D. all of the above

14. What sort of variable is manipulated by the researcher?
A. Dependent B. Co-dependent
C. Independent D. All variables are manipulated by the researcher.

15. A _____ refers to the people in your study whereas a _____ refers to a distinct group of people.
A. participant; population B. sample; cohort C. sample; population D. population; sample

16. Why do we use inferential statistics?
A. Inferential statistics are used to help us to compare the sample to the whole population.
B. Inferential statistics are used to help us to generalise from the sample to the whole population.
C. Inferential statistics are used to help us to show the difference between the sample and the whole population.
D. All of the above apply to the use of inferential statistics.

17. What is the degree to which sample statistics differ from the equivalent population parameter known as?
A. Selection bias B. Selection error C. Sample bias D. Sampling error.

18. Which of the following graphs would you use to explore the distribution of scores on an inventory measuring physical fitness?
A. histogram B. A box and whisker plot C. A stem and leaf plot D. All of the above.

19. Which of the following plots is most useful for detecting outlying scores?
A. Box and whisker plot B. Scattergram C. Stem and leaf plot D. Histogram.

20. You are doing a study to explore the relationship between attitudes towards exercise and physical fitness levels. Which plot might you use to explore this relationship?
A. Box and whisker plot B. Scattergram C. Histogram D. Stem and leaf plot.

21. Under what circumstances should we be cautious about using the mean as a measure of central tendency?
A. When data is positively skewed B. When the data is skewed
C. When data is negatively skewed D. All of the above

22. Complete the following sentence. It is difficult to interpret whether a distribution deviates from normality when using a:
A. Histogram B. box and whisker plot C. stem and leaf plot D. none of the above.

23. A sample mean is a ____ estimate and we do not know how close it is to the population mean.
B. Confidence B. sample C. distribution D. point

24. Inferential statistics deal with:
A. making conclusions and generalizations about population/s from our sample data.
B. the tabulation and organisation of data in order to demonstrate their main characteristics.
C. giving the best estimate of the population mean.
D. both the second and third statement.

25. To calculate confidence interval we need make use of:
A. z-scores B. histograms C. probability distributions D. none of the above.

26. The null hypothesis is:
A. the assumption that a significant result is unlikely
B. the assumption that there is a relationship or difference between the variables you are testing C. the assumption there is no relationship or difference between the variables you are testing D. the pattern between the variables you are testing.

27. A measure of dispersion is always _____
    A. Zero B. Positive C. Negative D. Infinity
28. Variance is always calculated from
    A. Mode B Media C. Mean D. Geometric mean
29. Suppose for 40 observations the variance is 50. If all observations are increased by 20, the variance of these increased observation will be
    A. 50 B. 70 C. 50/20 D. 30
30. If the standard deviation of the values 2, 4, 6, 8 is 2.58, then the standard deviation of the values 4, 6, 8, 10 is
    A. 0 B. 2.58 C. 5, D. 4.66 E. 2.33

QUESTION TWO [20 MARKS]

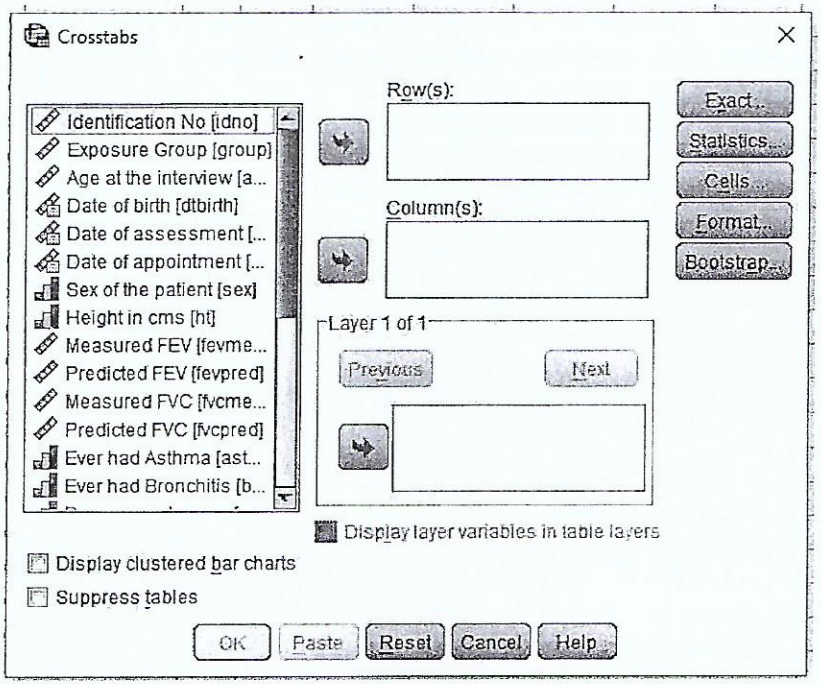
a. Briefly explain the Z-score and its importance in data analysis. [4 marks]
b. What is the importance of confidence interval research data analysis? [2 marks]
c. Using a relevant example distinguish between independent and dependent variables. [4 marks]
d. Variables can be split into categorical and continuous, and within these types there are different levels of measurement. Discuss these levels giving examples in each case. [10 marks]

QUESTION THREE [20 MARKS]

a. Briefly explain the difference between parametric and non-parametric statistical test. [4 marks]
b. Distinguish between standard deviation and standard error as used in statistics [4 marks]
c. Why is it important to calculate a confidence interval in data analysis? [4 marks]
d. Twenty-one heavy smokers were put on a treadmill at the fastest setting. The time in seconds was measured until they fell off from exhaustion: 18, 16, 18, 24, 23, 22, 22, 23, 26, 29, 32, 34, 34, 36, 36, 43, 42, 49, 46, 46, 57 Compute the mode, median, mean, upper and lower quartiles, range and interquartile range. [8 marks]

QUESTION FOUR [20 MARKS]

a. Outline the steps that are required to display the dialog box shown below in SPSS. [2 marks]



b. Compare and contrast Data and Variable views in SPSS [6 marks]

c. In SPSS you can create new variables and you can also modify the values of an existing variable. Briefly describe how to achieve these in SPSS. **[12 marks]**

QUESTION FOUR [20 MARKS]

a. Discuss Data analysis. **[5 marks]**
b. Distinguish between descriptive and inferential statistics **[4 marks]**
c. Discuss test statistics as used in data analysis **[3 marks]**
d. Justify why you would prefer to use SPSS rather than Microsoft Excel for research data analysis. **[8 marks]**