



CS

(Knowledge for Development)

KIBABII UNIVERSITY
(KIBU)

UNIVERSITY EXAMINATIONS
2019/2020 ACADEMIC YEAR

END OF SEMESTER EXAMINATIONS
YEAR TWO SEMESTER TWO EXAMINATIONS

FOR THE DEGREE IN
(INFORMATION TECHNOLOGY)

COURSE CODE : BIT 225

COURSE TITLE : DATA ANALYSIS TECHNIQUES

DATE: 04/02/2021

TIME: 2.00 P.M. – 4.00 P.M.

INSTRUCTIONS TO CANDIDATES

ANSWER QUESTIONS ONE AND ANY OTHER TWO.

QUESTION ONE (COMPULSARY) [30 MARKS]

1. Ordinal level data are characterized by:
 - A. Equal intervals between each adjacent score.
 - B. A fixed zero.
 - C. Data that can be meaningfully arranged by order of magnitude.
 - D. None of the above.
2. For what is the 'variable view' in IBM SPSS's data editor used?
 - A. Entering data.
 - B. Writing syntax.
 - C. Viewing output from data analysis.
 - D. Defining characteristics of variables.
3. What are the two main windows in SPSS?
 - A. Data view and variable view
 - B. Data editor and output viewer
 - C. Data view and output viewer
 - D. Variable view and output viewer
4. Which menu would you select to run statistical procedures?
 - A. Graph menu
 - B. Data menu
 - C. Analyze menu
 - D. Transform menu
5. Which menu item contains procedures to manipulate variables?
 - A. Graph menu
 - B. Data menu
 - C. Transform menu
 - D. Analyze menu
6. Which menu item is used to create bar graphs and scatter plots?
 - A. Graph menu
 - B. Data menu
 - C. Transform menu
 - D. Analyze menu
7. Which menu item contains the split file and select cases command?
 - A. Graph menu
 - B. Data menu
 - C. Transform menu
 - D. Analyze menu
8. SPSS will recognize which of the following mistakes?
 - A. Incorrect type of variable selected
 - B. Incorrect data entered
 - C. Incorrect statistical test selected
 - D. Incorrect form of instructions
 - E. Incorrect data set selected
9. What is the purpose of the toolbar in SPSS?
 - A. To list procedures not included in the procedures across the top of the Data Editor
 - B. To explain procedures more fully than in the procedures across the top of the Data Editor
 - C. To remind the researcher to perform certain procedures, e.g. save regularly
 - D. To display all the statistical analyses available
 - E. To enable procedures to be carried out without going to the options and selecting the appropriate procedure
10. Which of the following methods would you use to enter data on gender into SPSS so you could conduct statistical analyses?
 - A. Type male or female into the appropriate column of the data view
 - B. Type M into the appropriate column for male participants and F for female participants
 - C. Enter data for male participants into SPSS first, then data for female participants second
 - D. Enter the data from male and female participants into separate data sets
 - E. Numerically code male and female with different numbers
11. Which of the following is the best way to deal with missing data?
 - A. Leave the cell where data is missing blank
 - B. Only enter data for cases where you have a complete data set
 - C. Use numbers not used elsewhere to represent different kinds of missing data
 - D. Remove variables where data is missing
 - E. Make up a response

12. Which of the following aspects of a variable CANNOT be defined in the Variable View window?
- A. Value labels
 B. Measurement
 C. Maximum and minimum values
 D. Number of decimal places
 E. Missing values
13. Which of the following is the name given to the record of procedures conducted by SPSS which appears first in the Output Viewer?
- A. SPSS commands
 B. Syntax commands
 C. Command file
 D. Analysis commands
 E. Output commands
14. What is the difference between interval/ratio and ordinal variables?
- A. The distance between categories is equal across the range of interval/ratio data
 B. Ordinal data can be rank ordered, but interval/ratio data cannot
 C. Interval/ratio variables contain only two categories
 D. Ordinal variables have a fixed zero point, whereas interval/ratio variables do not
15. What is the name of the test that is used to assess the relationship between two ordinal variables?
- A. Spearman's rho
 B. Phi
 C. Cramer's V
 D. Chi square
16. A test of statistical significance indicates how confident the researcher is about:
- A. The inter-coder reliability of their structured interview schedule
 B. Passing their driving test
 C. Understanding the difference between bivariate and multivariate analysis
 D. Generalizing their findings from the sample to the population
17. Quantitative data refers to:
- A. numerical data that could usefully be quantified to help you answer your research question(s) and to meet your objectives.
 B. statistical analysis.
 C. any data you present in your report.
 D. graphs and tables.
18. Which of these is not one of the four main reasons for missing data?
- A. The analyst ignored its presence on the data form.
 B. The respondent may have missed a question by mistake.
 C. The respondent did not know the answer or did not have an opinion.
 D. The data was not required from the respondent, perhaps because of a skip generated by a filter question in a survey.
19. Computers are essential for quantitative data analysis because:
- A. increasingly data analysis software contains algorithms that check the data for obvious errors as it is entered.
 B. they are fun to use.
 C. they enable easy calculation for those of us not too good with figures.
 D. they are so powerful.
20. Which one of these is not a way of measuring central tendency?
- A. A. Regression analysis.
 B. Measuring the value, often known as the average, that includes all data values in its calculation (mean).
 C. Measuring the value that occurs most frequently (mode).
 D. Measuring the middle value or mid-point after the data have been ranked (median).

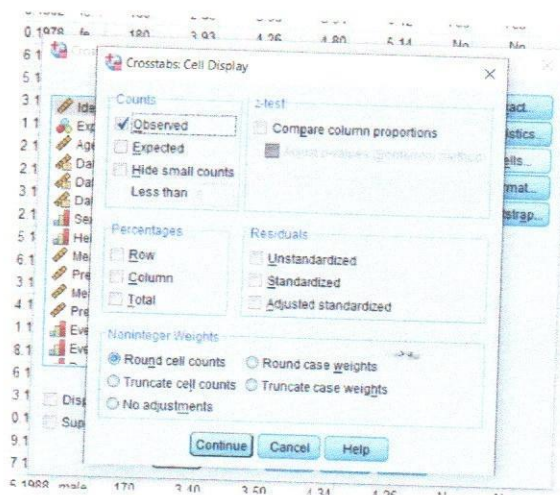
21. Standard deviation is:
- inappropriate in management and business research.
 - a way of illustrating crime statistics.
 - a way of describing those phenomena that are not the norm.
 - a way of measuring the extent of spread of quantifiable data.
22. Parametric and non-parametric are:
- two main groups of statistical significance tests.
 - terms used in medical practice.
 - general tests of statistical relevance.
 - alternatives to standard deviation tests.
23. Testing the probability of a relationship between variables occurring by chance alone if there really was no difference in the population from which that sample was drawn is known as:
- significance testing.
 - chi-squared tests.
 - correlation coefficients.
 - multiple regression analysis.
24. ANOVA is:
- a government body which collects social statistics.
 - a two-way analysis of variance.
 - a one-way analysis of variance.
 - the name of a statistical software package.
25. A correlation coefficient enables you to:
- measure the difference between two variables.
 - assess whether two variables measure the same phenomenon.
 - quantify the strength of the linear relationship between two ranked or quantifiable variables.
 - establish whether the data is telling you what you think it should tell you.
26. The coefficient of determination (sometimes known as the regression coefficient) enables you to:
- assess whether two variables measure the same phenomenon.
 - establish whether the data is telling you what you think it should tell you.
 - assess the strength of relationship between a quantifiable dependent variable and one or more quantifiable independent variables.
 - measure the difference between two variables.
27. The level of significance can be viewed as the amount of risk that an analyst will accept when making a decision
- True
 - False
28. By taking a level of significance of 5% it is the same as saying
- We are 5% confident the results have not occurred by chance
 - We are 95% confident that the results have not occurred by chance
 - We are 95% confident that the results have occurred by chance
29. The probability of rejecting the null hypothesis when it is true is called:
- Level of confidence
 - Level of significance
 - Power of the test
 - Difficult to tell
30. Which of the following factors do not affect the width of the confidence interval?
- Sample mean
 - Population variance
 - Sample size
 - Confidence level

QUESTION TWO [20 MARKS]

- a. When you view data in SPSS, each row in the Data View represents a case, and each column represents a variable. Using relevant examples, distinguish between cases and variables. [6 marks]
- b. Discuss the importance of Split option in SPSS and describe how to access it. [4 marks]
- c. Nominal or ordinal categorical variables can be coded numerically (e.g., recording a subject's gender as 1 if male or 2 if female) would be classified as numeric variables with zero decimal places. Explain the dangers of doing these and how to address them in SPSS. [4 marks]
- d. There are three ways of defining information about variables in SPSS. Briefly explain the three ways. [6 marks]

QUESTION TWO [20 MARKS]

- a. Describe how to access the dialog box displayed below. [2 marks]



- b. Differentiate between frequency table and *cross-tabulation*. [4 marks]
- c. Briefly explain circumstances when it is more useful to look at percentiles than it is to look at means. [4 marks]
- d. A dataset of 9, x, 10, 9, 20, 12, x, 16, x and 10 has a mean of 11. Calculate
- The value of x [3 marks]
 - The mode [1 mark]
 - The median [2 marks]

- iv. The variance [3 marks]
v. The standard deviation [1 mark]

QUESTION FOUR [20 MARKS]

- a. Sometimes you may need to compute a new variable based on existing information (from other variables) in your data. Explain any three situations where you may want to compute a new variable. [6 marks]
- b. Suppose you have test cumulative scores as percentages, and want to convert those percentages to classes. A typical grading scheme in Kenya is:
- Below 40: Fail ($\text{test} < 40$)
 - 40 to 49: Pass ($40 \leq \text{test} < 50$)
 - 50 to 59: 2nd Class Lower ($50 \leq \text{test} < 60$)
 - 60 to 69: 2nd Class upper ($60 \leq \text{test} < 70$)
 - 70 or higher: 1st Class ($\text{test} \geq 70$)

Briefly describe how you would convert the percentages into classes in SPSS.

[10 marks]

QUESTION FIVE [20 MARKS]

- a. Discuss the importance of research data analysis [4 marks]
- b. Distinguish between the terms below as used in research data analysis
- statistics and parameter [4 marks]
 - Descriptive and inferential statistics [4 marks]
- c. Briefly discuss the scenarios when it is wise to use bar graph and pie charts and not histogram in presentation of data analysis results. [4 marks]
- d. Compare and contrast continuous scale and categorical scale as used in SPSS. [4 marks]