



*(Knowledge for Development)*

**KIBABII UNIVERSITY  
(KIBU)**

**UNIVERSITY EXAMINATIONS  
2022/2023 ACADEMIC YEAR**

**SPECIAL/SUPPLEMENTARY EXAMINATIONS  
YEAR TWO SEMESTER TWO EXAMINATIONS**

**FOR THE DEGREE OF  
BACHELOR OF SCIENCE  
(INFORMATION TECHNOLOGY)**

**COURSE CODE : BIT 225**

**COURSE TITLE : DATA ANALYSIS TECHNIQUES**

**DATE: 09/08/2023**

**TIME: 2.00 P.M. – 4.0 P.M.**

---

**INSTRUCTIONS TO CANDIDATES  
ANSWER QUESTIONS ONE AND ANY OTHER TWO.**

**QUESTION ONE (COMPULSORY)[30MARKS]**

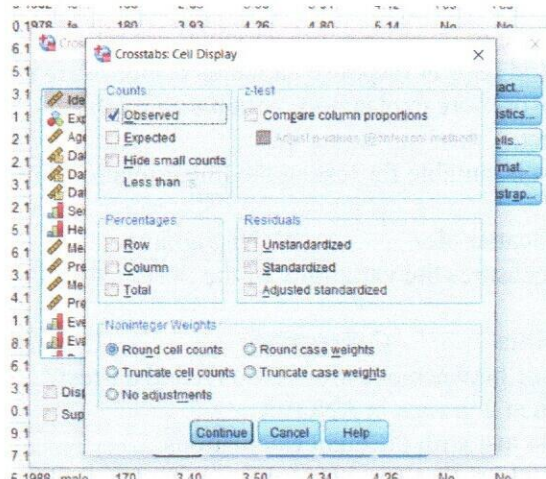
1. In normal distribution:  
A. Mean = Median = Mode  
B. Mean < Median < Mode  
C. Mean > Median > Mode  
D. Mean ≠ Median ≠ Mode
2. In a normal curve, the ordinate is highest at:  
A. Mean  
B. Variance  
C. Standard deviation  
D. Q1
3. The shape of the normal curve depends upon the value of:  
A. Standard deviation  
B. Q1  
C. Mean deviation  
D. Quartile deviation
4. In a normal probability distribution of a continuous random variable, the value of standard deviation is:  
A. Zero  
B. Less than zero  
C. Greater than zero  
D. None of the above
5. The normal curve is symmetrical and for symmetrical distribution, the values of all odd order moments about the mean will always be:  
A. 1  
B. 0.5  
C. 0.25  
D. 0
6. A national achievement test is administered annually to 3rd graders. The test has a mean score of 100 and a standard deviation of 15. If Jane's z-score is 1.20, what was her score on the test?  
A. 82  
B. 88  
C. 100  
D. 118
7. The mean temperature in Glens Falls for the month of February is 23 degrees with a standard deviation of 4.2 degrees. What is the z-score for a temperature of 17 degrees?  
A. 1  
B. 1.43  
C. -1.43  
D. 11.5
8. Which of the following is not part of the calculation for a z-score  
A. Sample score  
B. Population mean  
C. Population standard deviation  
D. Degree of normality
9. What best describes a z-score:  
A. It is the average of all raw scores in a normal distribution  
B. It is the measure of dispersion in a distribution of scores  
C. t is the position of a score relative to the mean  
D. It is the frequency of a score in standardized units
10. Erica's z-score on a test is 1.1; her actual score is 20 and the mean of the test is 15. What is the standard deviation of the test?  
A. -1.21  
B. 1.21  
C. 4.55  
D. -4.55
11. Which of the following is a benefit of converting raw scores to standard scores?  
A. Allows you to be more likely to find a significant result  
B. Allows you to compare scores from different scales  
C. Allows you to check if your data have been entered correctly  
D. None of the above
12. A t-test is a significance test that assesses  
A. The means of two independent groups  
B. The medians of two dependent groups  
C. The modes of two independent variables  
D. The standard deviation of three independent variables
13. Statistical significance or the probability of finding statistical significance is also known as  
A. Degrees of freedom (df)  
B. p-value  
C. Standard deviation (sd)  
D. A constant source of frustration!
14. The three types of t-tests are  
A. One-sample t-tests  
B. Null Hypothesis t-tests  
C. Independent sample t-tests  
D. Paired samples t-tests; E. Variable t-tests
15. Chi-square is used to analyse:  
A. Scores  
B. Ranks  
C. Frequencies  
D. Any of these
16. Using a goodness-of-fit test, we can assess whether a set of obtained frequencies differ from a set of \_\_\_\_\_ frequencies  
A. Mean  
B. Actual  
C. Predicted  
D. Expected

17. A chi-square test is used to test whether a 0 to 9 spinner is "fair" (that is, the outcomes are all equally likely). The spinner is spun 100 times, and the results are recorded. The degrees of freedom for the test will be  
 A. 8      B. 9      C. 10      D. 99
18. If there is a very strong correlation between two variables then the correlation coefficient must be  
 A. any value larger than 1  
 B. much smaller than 0, if the correlation is negative  
 C. much larger than 0, regardless of whether the correlation is negative or positive  
 D. None of these alternatives is correct.
19. The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:  
 $y = -0.0127 + 0.0180x$   
 The above equation implies that:  
 A. each beer consumed increases blood alcohol by 1.27%  
 B. on average it takes 1.8 beers to increase blood alcohol content by 1%  
 C. each beer consumed increases blood alcohol by an average of amount of 1.8%  
 D. each beer consumed increases blood alcohol by exactly 0.018
20. Regression modeling is a statistical framework for developing a mathematical equation that describes how  
 A. one explanatory and one or more response variables are related  
 B. several explanatory and several response variables response are related  
 C. one response and one or more explanatory variables are related  
 D. All of these are correct.
21. Which of the given plots is suitable for testing the linear relationship between a dependent and independent variable?  
 A. Bar chart      B. Scatter plot      C. Histograms      D. All of the above.
22. The.....sum of squares measures the variability of the observed values around their respective treatment means.  
 A. Error      B. Total      C. Treatment      D. Interaction
23. What is the key difference between one-way ANOVA and a t-test?  
 A. You can have more than 2 groups in ANOVA.  
 B. They are the same test just with different calculations.  
 C. T-tests split variance into within and between.  
 D. ANOVA is about the mean and the t-test is about the variance
24. Which is the correct order of steps for calculating and interpreting a one-way ANOVA?  
 A. Determine SS, Calculate df, Calculate MS, Calculate F, Compare F obtained to F critical.  
 B. Calculate F, determine SSS, Calculate df, Compare F obtained to F critical.  
 C. Find F and compare.  
 D. Calculate T, calculate df, compare T obtained to T critical
25. 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.
26. Which menu would you select to run statistical procedures?  
 A. Graph menu      B. Data menu      C. Analyze menu      D. Transform menu
27. Which menu item is used to create bar graphs and scatter plots?  
 A. Graph menu      B. Data menu      C. Transform menu      D. Analyze menu
28. 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

29. Which of the following methods would you use to enter data on gender into SPSS so you could conduct statistical analyses?
- Type male or female into the appropriate column of the data view
  - Type M into the appropriate column for male participants and F for female participants
  - Enter data for male participants into SPSS first, then data for female participants second
  - Enter the data from male and female participants into separate data sets
  - Numerically code male and female with different numbers
30. Which of the following aspects of a variable CANNOT be defined in the Variable View window?
- Value labels
  - Measurement
  - Maximum and minimum values
  - Number of decimal places
  - Missing values

### QUESTION TWO [20 MARKS]

- T-test is used to tell if there is a difference between the two groups in the mean. Briefly explain why can't we just use the "difference" in score (means)? [4 marks]
- Describe how to access the dialog box displayed below. [2 marks]



- Differentiate between frequency table and cross-tabulation. [4 marks]
- 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]
  - The variance. [3 marks]
  - The standard deviation [1 mark]

### QUESTION THREE [20 MARKS]

- Differentiate between ordinal and ratio measurement levels as used in data analysis. [4 marks]
- Outline four properties of a normal distribution. [4 marks]

- c. Explain four important characteristics of a Chi-Square Test. [6 marks]

An unbiased die is thrown 132 times with the following results:

No. of Times	1	2	3	4	5	6
Frequency	16	20	25	14	29	28

work out the value of  $\chi^2$  (Chi-Square). [6 marks]

$$\chi^2 = \sum \frac{(f_e - f_o)^2}{f_e}$$

#### QUESTION FOUR [20 MARKS]

- a. Distinguish between non-parametric and parametric metrics as used in data analysis. [4 marks]
- b. Use the following data set to answer the next three questions

Value	1	2	3	4	5	6	7	8
Frequency	2	0	0	3	4	10	12	9

- i. Calculate the mean [3 marks]
- ii. Calculate media [1 mark]
- iii. Calculate the interquartile range [2 marks]
- iv. Describe the distribution of the data [2 marks]
- c. Explain the requirements that your data must meet to allow you perform One Way ANOVA in data analysis. [8 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
- i. statistics and parameters. [4 marks]
- ii. Descriptive and inferential statistics [4 marks]
- c. Briefly discuss the scenarios when it is wise to use bar graphs and pie charts and not histograms in the presentation of data analysis results. [4 marks]
- d. Compare and contrast continuous scale and categorical scale as used in SPSS. [4 marks]