# UNIVERSITY OF SWAZILAND

## FACULTY OF HEALTH SCIENCES

### FINAL EXAMINATION NOVEMBER 2021

**TITLE OF PAPER: DATA ANALYSIS AND INTERPRETATION**

**COURSE CODE: GNS 612**

**TIME ALLOWED: THREE (3) HOURS**

**PAGES EIGHT (5) INCLUDING COVER PAGE**

**MARKS: 100**

**INSTRUCTIONS:**

1. ENSURE THAT YOU ARE WRITING THE EXAM FOR THE COURSE IN WHICH YOU ARE REGISTERED.

2. THERE ARE FOUR (4) QUESTIONS IN THIS PAPER.

3. ANSWER ALL <u>FOUR (4)</u> QUESTIONS.

4. FOR QUESTIONS 1 – 3, SHOW ALL YOUR WORK.

5. WRITE LEGIBLE.

THIS PAPER IS **NOT** TO BE OPENED UNTIL THE INVIGILATOR HAS GRANTED

PERMISSION.

## QUESTION 1

(a) How does the process of Binomial distribution differ from Poisson distribution?

(7)

(b) We know that IQ is normally distributed with a mean of 100 and a variance of 64. What is the probability that a person chosen at random from the population has an IQ greater than 120? (5)

(c) The score obtained on a test for anxiety neurosis is known to be normally distributed with a mean of 50 and variance 36. Find the probability that a person randomly selected from the population will score on this test:

(i) More than 41,

(ii) More than 72. (6)

(d) Among year 3 Environmental Health Science students at Mbabane Campus, the University of Eswatini, academic year 2019/2020, and the parameters in Table 1 were measured (**Table 1**).

**Table 1**. Summary of some biometric data with reference to the year 3 Environmental Health Science students at Mbabane Campus, the University of Eswatini academic year 2019/2020.

| Parameter | Recording |
|---|---|
| Mean body weight | 67 kg |
| Standard deviation | 2.5 |
| Students on full Government of Eswatini sponsorship (scholarship) | 35 |
| Students on no Government of Eswatini sponsorship (scholarship) | 15 |

(a) If we select 5 students at random from **Table 1**, what is the probability that exactly 3will have no Government of Eswatini sponsorship? (7)

**TOTAL = 25 MARKS**

## QUESTION 2

(a) Given the following data;

| Degrees of freedom | 15 |
|---|---|
| Standard deviation | 2 |
| Mean | 25 |

Construct 95% confidence interval (5)

(b) A study by Nkambule et al. (2010) examined a sample of 8 subjects with open-angle glaucoma and unilateral hemifield defects. The ages (years) of subjects were:

62, 68, 48, 62, 51, 61, 41 & 34.

Can we conclude that the age of the population from which the sample may be presumed to have been drawn is less than 60 years?                              (20)

**TOTAL = 25 MARKS**

## QUESTION 3

A. The table below shows a hypothetical relationship between condom use and the risk of contracting the human papilloma virus (HPV). Therefore, over a 10 year period among 50 people followed 20 of engaged in protected sex and 30 engaged in unprotected sex. The findings were summarized as follows:

|  | HPV | | TOTAL |
|---|---|---|---|
|  | Yes | No |  |
| No Condom | 15 | 5 | 20 |
| Use condom | 3 | 27 | 30 |
| **TOTAL** | **18** | **32** | **50** |

(i)     What is the Risk of contracting HPV among those who use condoms?     (3)

(ii)     Compute the Risk Ratio for your data?                                                     (4)

(iii)     Interpret your finding in (ii)                                                                     (2)

(ii)     What is the 95% Confidence Interval for your Risk Ratio?   Interpret this

Confidence Interval.                                                                                (5)

B. You examined the relationship between the age of the participants and finding life

exciting. The SPSS output below reflects your findings.

**Correlations**

|  |  | Is Life Exciting or Dull | Age of Respondent |
|---|---|---|---|
| Is Life Exciting or Dull | Pearson Correlation | 1 | .098** |
|  | Sig. (2-tailed) |  | .002 |
|  | N | 980 | 979 |
| Age of Respondent | Pearson Correlation | .098** | 1 |
|  | Sig. (2-tailed) | .002 |  |
|  | N | 979 | 1514 |

**. Correlation is significant at the 0.01 level (2-tailed).

(i)     State the key findings from the output.                                      (3)

(ii)    Interpret the findings in (i) for your reader.                               (3)

C.  In another study you investigated the association between the independent variable, taking active part in world affairs, and the dependent variable, general happiness. You analysed the data using Linear Regression.

**ANOVA**[a]

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 8.479 | 1 | 8.479 | 23.008 | .000[b] |
| | Residual | 366.327 | 994 | .369 | | |
| | Total | 374.806 | 995 | | | |

a. Dependent Variable: General Happiness

(i)     In order for you to use Linear Regression, how should be your variables?                                                      (2)

(ii)    Interpret the findings from the SPSS output.              (3)

**TOTAL = 25 MARKS**

## QUESTION 4

**INSTRUCTIONS:** For the following multiple choice questions write the number and the corresponding letter (in upper case, A, B, C, or D) of the correct response. For example, 16. D

1.  There are basically two types of statistics, descriptive and inferential. Which of the following statements is true about descriptive statistics?

    A.  Descriptive statistics enable you to make decisions about your data, e.g. is one group mean significantly different from the population mean.

    B.  Descriptive statistics describe the data

C. Descriptive statistics enable you to draw inferences about your data, e.g. does one variable predict another variable

D. Descriptive statistics help the investigator determine if one variable explains the other

2. You summarized grades in the Data Analysis and Interpretation course for MNSc FNP cohort A (Class A) and cohort B (Class B) in the following graph.

**Questions 2 – 7 refer to the graph shown below:**
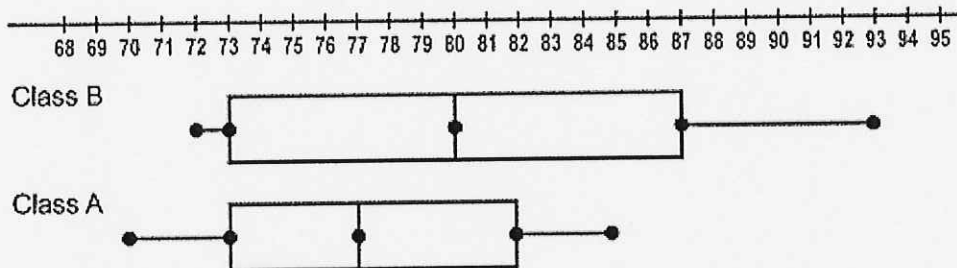


What is this graph known as?

A. A scatterplot

B. A box-whisker diagram

C. A histogram

D. An error bar chart

3. Based on the chart, what was the median grade for Cohort B?

A. 73%

B. 82%

C. 80%

D. 77%

4. Based on the chart, what was the median grade for Cohort A?

A. 73%

B. 82%

C. 80%

D. 77%

5. Based on the chart, what was the IQR for Cohort B?

A. 15

B. 14

C. 9

D. 12

6. Among the two cohorts, which has skewed grades?

   A. Cohort B

   B. Cohort A

   C. Cohort A and B

   D. None

7. The cohort that has skewed grades, has grades that are skewed to the……..

   A. Right

   B. Left

   C. Middle

   D. Platykurtic

8. A shop owner decided to calculate how much revenue he gained selling cabbages each month. What type of variable would the amount of revenue gained from cabbages be?

   A. Categorical

   B. Discrete

   C. Continuous

   D. Nominal

9. Chi-square is used to analyse………..

   A. Scores

   B. Frequencies

   C. Ranks

   D. Any of these

10. On which of the following does the critical-value of chi-square rely?

    A. Degrees of freedom

    B. The sum of the frequencies

    C. The row totals

    D. The number of variables

11. A researcher was interested in stress levels of 15 lecturers during a lecture. She took the same group of 8 lecturers measured their anxiety (out of 15) during a normal lecture and in a lecture she had paid the students to be disruptive and misbehave. What test is best used to compare the mean level anxiety during the two lectures?

    A. Independent t-test

    B. Paired t-test

C. One way independent ANOVA

D. Mann-Whitney test

12. A researcher measured people's physiological reaction to horror films. He split the data into two groups: males and females. The resulting data were normally distributed, and men and women had equal variances. What test should be used to analyse the data?

A. Dependent t-test

B. Independent t-test

C. Mann-Whitney test

D. Wilcoxon signed rank test

13. Which of the following are assumptions underlying the use of parametric tests (based on the normal distribution)?

A. The data should be at least at interval level

B. The samples tested should have approximately equal variances

C. Some feature of the data should be normally distributed

D. All of the options are true

14. Which of the following is least affected by outliers?

A. Median

B. Mean

C. Range

D. Standard deviation

15. If we were to pull all possible samples from the population, calculate the mean of every sample, and construct a graph of the shape of the distribution based on all the means, what would we have?

A. The population distribution of the mean

B. The sampling distribution of the mean

C. The bootstrap distribution of the mean

D. The standard error of the mean

16. A significant Odds Ratio (OR) of 2.5 for BMI as a continuous predictor of heart disease in a binary logistic model would indicate which of the following?

A. The odds of heart disease increase 2.5% for every 1-point increase in BMI

B. Those with heart disease have 2.5 times higher odds of having an increasing BMI compared to those without heart disease

C. The odds of heart disease are 2.5 times higher for every 1-point increase in BMI

    D. There are 2.5 times as many people with heart disease as without among those with higher BMI

17. When does a Confidence Interval (CI) indicate a significant Risk Ratio?

    A. When it includes a 0

    B. When it does not include a 0

    C. When it includes a 1

    D. When it does not include a 1

18. Focus groups are better than questionnaires if the aim is to analyse attitudes.

    A. True

    B. False

19. Focus groups are better than questionnaires if the researcher wants to reach as many people as possible.

    A. True

    B. False

20. Once data is collected the researcher should dive right in onto the analysis process.

    A. True

    B. False

21. Qualitative data obtained from open ended questions should be analysed separately using qualitative analysis methods.

    A. True

    B. False

22. When computing Risk Difference, what is your reference point?

    A. 1

    B. -1

    C. 0

    D. 0.5

23. A Type I error occurs when …………………..

    A. We conclude that there is no effect in the population when in actual fact there is.

    B. We conclude that the test statistic is not significant when in fact it is.

    C. The data we have typed into SPSS is different from the data that we collected

    D. We conclude there is an effect in the population when in fact there is not

24. The Correlation (r) between two variables, A and B is .12 with a significance of p <.01. What can we conclude?

A. There is a substantial relationship between A and B

B. The variable A causes the variable B

C. There is a small relationship between A and B

D. The variable B causes the variable A

25. What does a significant test statistic tell you?

A. There is an important effect

B. The test statistic is larger than we would expect if there were no effect on the population

C. The null hypothesis is true

D. All of these

**TOTAL = 25 MARKS**