UNIVERSITY OF SWAZILAND

FACULTY OF HEALTH SCIENCES

FINAL EXAMINATION NOVEMBER. 2021

TITLE OF PAPER: BIOSTATISTICS FOR EVIDENCE-BASED PRACTICE

COURSE CODE: GNS 638

TIME ALLOWED: THREE (3) HOURS

PAGES: 10 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 FOUR (4) QUESTIONS.

4. SHOW ALL YOUR WORK.

5. WRITE LEGIBLE.

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

PERMISSION.

QUESTION 1

You have two (2) hypothetical clients, Njalo and Njabu who are on antiretroviral therapy (ART), your facility has been monitoring their CD4 cell (mm^3) count for the past five (5) years.

Njalo: 22, 39, 50, 68, 70, 90, 120, 150

Njabu: 69, 72, 77, 100, 115, 120, 144, 160.

Using the mean, median, range, and interquartile range, compare the CD4 count of Njalo with that of Njabu.

(i) Mean (3)

(ii) Median (2)

(iii) Range (2)

(iv) Interquartile range (IQR) (4)

A. Between Njalo and Njabu who has a higher mean, and what does having a higher mean signify? (2)

B. Between Njalo and Njabu who has a higher interquartile range, and what does this mean? (2)

C. The summary from the socio-demographic data of your study is as following:

Socio-demographic variables		Frequency(n)	Percentage (%)	Mean	Standard deviation
Gender	Male	31	43.7		
	Female	40	56.3		
Age	18-21	44	62	21.2	2.66
	22-25	23	32.3		
	26-32	4	5.6		
Marital status	Single	67	94.4		
	Married	2	2.8		
	Divorced	2	2.8		
Number of friends at university	0-5	48	67.6	4.70	3.30
	6-12	23	32.4		

(i)   What was the key finding on marital status?   (2)

(ii)   Show the findings on marital status in an appropriate graph / chart   (4)

(iii)   For the variable age, which could be one of the best graphs/ charts to present this
        variable?   (2)

(iv)   Justify your response to (iii)   (2)

TOTAL = 25 MARKS

QUESTION 2

A. You studied the association between the independent variables (IVs) gender and drinking problem with the dependent variable (DV), unemployment (DV). You then analysed your data using Logistic Regression. The findings are presented in the SPSS outputs below.

Classification Table[a]

Observed			Predicted		
			Unemployed and Looking for Work a Month +		Percentage Correct
			Yes	No	
Step 1	Unemployed and Looking for Work a Month +	Yes	0	59	.0
		No	0	914	100.0
	Overall Percentage				93.9

a.   The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	sex(1)	-.833	.276	9.109	1	.003	.435	.253	.747
	Drinking problem	1.158	.659	3.093	1	.079	3.184	.876	11.577
	Constant	.879	1.310	.450	1	.502	2.408		

a. Variable(s) entered on step 1: sex, drinking problem. Sex: 1 = Male, 2 = Female; Drinking Problem: 1 = Yes, 2 = No

(i)   What is the meaning of the findings in the Classification Table.   (4)

(ii)   Explain the findings obtained in Variables in the Equation table.   (8)

B. You strongly suspect that counselling confounds the association between your IV, demotion and DV, unemployment. So you ran two logistic regressions.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1ᵃ	Demotion	-2.619	.424	38.202	1	.000	.073	.032	.167
	Constant	2.929	.148	390.891	1	.000	18.708		

a. Variable(s) entered on step 1: Demotion.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1ᵃ	Demotion	-2.589	.427	36.851	1	.000	.075	.033	.173
	Counselling	.757	.453	2.794	1	.095	2.132	.878	5.178
	Constant	1.475	.871	2.866	1	.090	4.370		

a. Variable(s) entered on step 1: Demotion, Counselling.

(i) From the two (2) SPSS outputs above, what is your conclusion, does counselling confounds the association between your IV, demotion and DV, unemployment. (2)

(ii) Show all the computation that justifies your response to (i). (5)

C. In your research you are investigating the association between social adjustment (IV) and depression (DV) among high school students residing in hostels. Both variables Social Adjustment and Depression are measured on a continuous scale. Which data analysis method will be appropriate to apply?

(i) Logistic Regression

(ii) Linear Regression

(iii) ANOVA

(iv) Risk Ratio (2)

D. In another study you investigated the same variables the association between social adjustment (IV) and depression (DV) among high school students residing in hostels. However, Social Adjustment was measured as categorical and Depression was a continuous scale. Which data analysis method will be appropriate to apply?

(i)   Logistic Regression

(ii)   Linear Regression

(iii)   ANOVA

(iv)   Risk Ratio   (2)

E.   What makes the researcher bother him /herself about the presence of a confounding variable in a study? The confounding variable can....................

    (i)   Reverse the relationship between the IV and DV

    (ii)   Exaggerate the relationship between the IV and DV

    (iii)   Not influence the association between IV and DV

    (iv)   Underestimate the association between IV and DV
        (Select all responses that apply)   (3)

TOTAL = 25 MARKS

QUESTION 3

A.   The table below shows a hypothesized relationship between using contraceptives and the odds of developing deep vein thrombosis (DVT).

	DVT		TOTAL
	Yes	No	
Oral contraceptives	7	43	50
No oral contraceptives	4	46	50
TOTAL	11	89	100

    (i)   What are the Odds of DVT among those who took oral contraceptives?   (2)

    (ii)   What is the Odds Ratio for the above data?   (4)

    (iii)   Interpret the Odds Ratio you computed in (ii)   (3)

    (iv)   What is the 95% Confidence Interval for your Odds Ratio?   (3)

B. You are interested in the accuracy of the rapid test that is used to determine COVID-19 antibodies, in a setting where the disease prevalence is 10%. You have the following information about the rapid test.

	COVID-19 Results		
	Positive	Negative	
COVID-19 exposed	16	4	20
COVID-19 exposed'	5	35	40
Total	21	39	60

(i)   Differentiate sensitivityfrom specificity of a diagnostic test?   (3)

(ii)   What is the sensitivity of the given COVID-19 antibody test?   (3)

(iii)   Compute the specificity of the above COVID-19 antibody test?   (3)

(iv)   Based on your findings in (ii) and (iii) is the COVID-19 antibody test a good test? Explain.   (2)

TOTAL = 25 MARKS

QUESTION 4

1. Logistic Regression is used when you want to:
   A. Predict a dichotomous variable from a continuous or dichotomous variable
   B. Predict a continuous variable from a dichotomous variable
   C. Predict a categorical from several other categorical variables
   D. Predict a continuous variable from a continuous or dichotomous variable

2. The Odds Ratio is.........
   A. The ratio of the probability of an event not happening to the probability of an event happening
   B. The probability of an event occurring
   C. The ratio of the odds after a unit change in the predictor to the original odds
   D. The ratio of the probability of an event happening to the probability of an event not happening

3. In binary Logistic Regression
   A. The dependant variable is continuous
   B. The dependent variable is divided into equal subscales

C. The dependent variable consists of two categories

D. There is no dependent variable

4. What does Multiple Linear Regression examine?

A. The relationship between more than one dependent variable and only one independent variable

B. The relationship between one or more than one dependent variable and only one independent variable

C. The relationship between one dependent variable and more than one independent variables

D. The relationship between one dependent variable and one independent variable

5. Which one of the following statistical tests allows causal inferences?

A. Analysis of Variance

B. Regression

C. T-test

D. None of these

6. What is skewness?

A. A pointy distribution

B. A flat distribution

C. Lack of symmetry of a distribution

D. A platykurtic distribution

7. Which of the following methods is recommended to use when running an analysis of of missing data?

A. Pairwise deletion

B. Multiple imputation

C. Single imputation

D. Dummy variable adjustment

8. What does multiple imputation do?

A. Removes any observations that has missing in one or more of the variables in the model

B. Uses information from all values from other variables to predict values on missing variable(s)

C. Calculations are based on all available data pairwise for all pairs of variables

D. Insert a new value for all missing observations in a variable (e.g. 0 or the mean) as well as including a dummy variable coded 1if the original data is missing

9. What exactly is an interaction / moderation effect?

A. When a third variable and an independent variable affect the dependent variable simultaneously

B. When a third variable reduces the effect of an independent variable on the dependent variable

C. When a third variable affects the relationship between the independent variable and the dependent variable

D. When a third variable affects an independent variable but not the dependent variable

10. Imagine you conducted a study to look at the association between whether an expectant mother eats breakfast (or not) and the gender of her baby. The results were Cramers'v = 0.22. How would you interpret this value?

A. There was a medium to large association between gender of the baby and whether or not the mother ate breakfast everyday

B. There was a small to medium association between gender of the baby and whether the mother ate breakfast everyday

C. 22% of the variation in frequency counts of the baby gender (boy or girl) can be explained by whether or not the mother ate breakfast everyday

D. 22% of the variation in frequency counts of the baby gender (boy or girl) can be predicted by whether or not the mother ate breakfast everyday

11. What is the primary purpose of Pearson's and Spearman Correlation Coefficient?

A. Examining the relationship between two non-categorical variables

B. Identify deviation from normality for continuous variables

C. Examining the relationship between two categorical variables

D. Comparing means across groups

12. Which of the following would be considered a strong negative correlation?

    A. .89

    B. .09

    C. -.89

    D. -.09

13. Which of the following is not an assumption for Pearson's Correlation Coefficient?

    A. Normally distributed variables

    B. Monotonic relationship

    C. Linear relationship

    D. Constant variance

14. Which of the following is not an assumption for simple Linear Regression?

    A. Multicolinearity

    B. Linear relationship

    C. Constant variance

    D. Normally distributed variables

15. A confidence interval (CI) indicates significance when……

    A. It includes 1

    B. It includes 0

    C. It does not include 1

    D. It does not include 0

16. For a categorical predictor for Logistic Regression, what is the group that other groups are compared to called?

    A. Null group

    B. Independent group

    C. Standard group

    D. Reference group

17. What is the primary purpose of ANOVA?

    A. Comparing means across three or more groups

    B. Comparing medians across three or more groups

C. Examining the relationship between two categorical variables

D. Identifying normally distributed data

18. Chi-squared can be used to understand the relationship between ......

    A. Any two variables

    B. Two categorical variables

    C. Two continuous variables

    D. One continuous and one categorical variable

19. Chi-squared is computed by first squaring the difference between ......

    A. Observed frequencies and expected frequencies

    B. Observed frequencies and the total sample size

    C. Observed frequencies and observed percentages

    D. Expected values and observed percentages

20. Which of the following will be an appropriate measure of central tendency of a variable that is continuous and normally distributed?

    A. Mean

    B. Variance

    C. Median

    D. Mode

21. Which of the following will be most appropriate for describing the spread of variable that is extremely skewed to the right?

    A. Standard deviation

    B. Range

    C. IQR

    D. Mode

22. In a data frame containing information on the age and height of 100 people, the people are .... and age and height are the ....

    A. Observations, variables

    B. Variables, observations

    C. Data, factors

D. Factors, data

23. Which of the following statements about outliers are NOT true?

    A. Outliers are values very different from the rest of the data
    B. Influential cases will always show as outliers
    C. Outliers have an effect on the mean
    D. Outliers have an effect in regression parameters

24. In a small data sample (N = 20) what can you say about a z-score of 2.37?
A. It is significant at $p < .05$
B. It is significant at $p < .001$
C. It is significant at $p < .01$
D. It is not significant

TOTAL = 25 MARKS