



“Investing in Africa’s Future”

COLLEGE OF HEALTH, AGRICULTURE AND NATURAL SCIENCES

DEPARTMENT OF PUBLIC HEALTH AND NURSING [DPHN]

NSPH 544: APPLIED BIOSTATISTICS [MPH]

END OF FIRST SEMESTER EXAMINATIONS

NOVEMBER - DECEMBER 2023

LECTURER: DR. Z.M. ZINGONI

DURATION: 3 HOURS

INSTRUCTIONS

ANSWER **ALL** QUESTIONS FROM SECTION A AND **ANY THREE (3)** FROM SECTION B

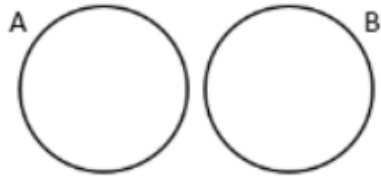
THE MARK ALLOCATION FOR EACH QUESTION IS INDICATED AT THE END OF THE QUESTION

CREDIT WILL BE GIVEN FOR LOGICAL, SYSTEMATIC AND NEAT PRESENTATION

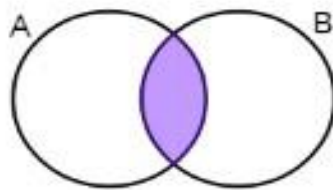
SECTION A: 40 marks-Answer ALL questions

1. Two events, A and B, can be described in different ways illustrated below:

Option A



Option B



- i. Which of the options indicates mutually exclusive events? [1]
- ii. Which option indicates independent events? [1]
- iii. Which option indicates joint events? [1]
- iv. Which of the options indicated conditional events? [1]

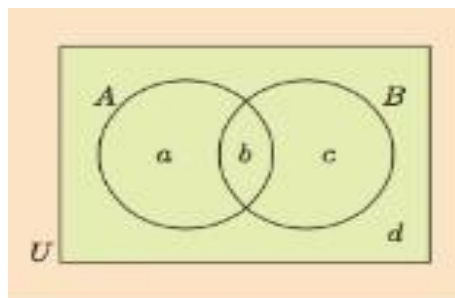
Option B was used to describe the relationship between A and B events as shown below. Assuming that A=HIV infection and B=Tuberculosis infection and the corresponding probabilities being

$$a=0.12$$

$$b=0.18$$

$$c=0.42$$

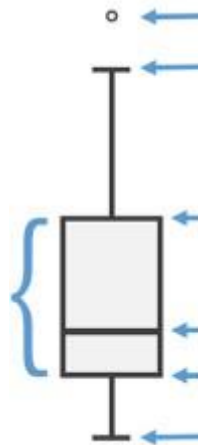
$$d=0.28$$



- v. Calculate the following
 - a. $P(A)$ [2]
 - b. $P(B)$ [2]

- c. $P(A \text{ and } B)$ [2]
- d. $P(A \setminus B)$ [2]
- e. $P(A \text{ or } B)$ [2]

2. A random variable can follow a discrete or continuous distribution. State 1 discrete distribution the random variable X can follow. [1]
3. Describe each of the following plots indicating the best they can be used in research with examples. *You can add illustrations*
 - i. Ogive plot [2]
 - ii. Frequency polygon [2]
4. A box and whisker plot is one of the most common plots used to describe the distribution of the data. Label the pointed parts on the diagram below which normally show on the box plot. [7]



5. For each of the following statements, give the best possible conclusion **with a justification**.
 - i. A variable has a mean of 25, a mode of 19 and a median of 22. State the skewness of this data. [2]
 - ii. A variable has a kurtosis of 4.7. Which type of kurtosis does this variable have? [2]
 - iii. A variable has a skewness of 1.5. Is it left or right-skewed? [2]
 - iv. What is the name of a distribution with a mean of 0 and a variance of 1? [2]
 - v. If you are testing the hypothesis using a p-value approach and the data analysis gives a p-value of 0.78 at 5%. Do you reject or fail to reject the null hypothesis? [2]

vi. If you are using the 95% confidence interval to test the following hypothesis statement

$$H_0: \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 - \mu_2 \neq 0$$

If the 95% confidence interval obtained ranged from (-3.5 to 2.5), do you reject or fail to reject the null hypothesis? [2]

i. If the ANOVA analysis gives significant results which show mean differences between groups. Do you perform a post-estimation test or not? [2]

SECTION B: Answer any 3 of the following

Question 6: 20 marks

The Normal distribution has several properties that can be used to describe it. Can you write short notes on the Normal distribution with respect to the following:

- a) The two main parameters used to describe a normally distributed variable [2]
- b) The histogram plot with a super-imposed curve [3]
- c) Skewness [2]
- d) Kurtosis [2]
- e) The total area under the curve [1]
- f) The total area under the curve within 1SD; 2SD and 3SD (*you can add an illustration*). [3]
- g) The standard normal transformation [2]
- h) The sampling distribution of the mean [2]
- i) The Z-values linked to the 90%, 95% and 99% assuming a two-tailed level of significance. [3]

Question 7: 20 marks

A medical practitioner sets out to determine the linear relationship between the birth weight and age of the mother at the Mpilo Hospital. The data for the 20 women who delivered during the study period is shown below. The medical practitioner approached you to perform an analysis to answer their research question. The student consulted you to perform the analysis to establish if the linear relationship exists. The following data is provided to you for analysis.

Birth weight (Kg)	Age in years at delivery (years)
1.6	25
2.4	36
1.8	29
2.6	31
1.7	30
3.7	42
2.3	32
2.6	30
1.4	20
2.7	29
3.1	30
2.9	37
2.8	26
2.5	35
3.3	26
3.6	22
4.2	23
2.7	28
3.1	33
3.9	34

- Estimating the effect of the mother's age on the birth weight, that is the model slope. [7]
- Estimating the model intercept [5]
- Write down the model fully and interpret the slope and the intercept [5]
- Predict the weight of an unborn child if the mother's age is 27 years [1]

Question 8: 20 marks

In a study to investigate viral load suppression among patients on first-line and those on second-line after being followed up for a year. At the end of the study, the results below were reported:

Viral load	First line drug	Second line drug
Mean	740	155
SD	57	23
n	15	15

- a) Is there a significant mean difference in the viral load suppression among those on first-line drugs and those on the second-line drug? Assume a two-tailed test at the 5% level of significance and equal (pooled) variance [8]
- b) What is the 95% confidence interval of the two mean values tested in part (a) [5]
- c) What is the 90% confidence interval of the means values tested in part (a) [4]
- d) Comment on the results from a-c fully [3]

Question 9-20 marks

- a) A data analysis short course on Factor analysis which is a data reduction data analysis approach was introduced at AU in 2022. To assess the level of understanding of the participants, the facilitator gave pre- and post-module assessments whose marks are shown in the table below for 10 students.

Student	Pre-module score	Post-module score
1	18	22
2	21	25
3	16	17
4	22	24
5	19	16
6	24	29
7	17	20
8	21	23
9	23	19
10	18	20

- i. Determine the mean score difference of the pre- and post-module assessment marks. [2]
 - ii. By performing hypothesis testing, determine if there is a significant change in the mean scores before the module training and after the module training at the 5% level of significance assuming a two-tailed test. Show all steps [5]
 - iii. Determine the 95% confidence interval for the estimate tested in part (iii) [5]
- b) The rate of COVID-19 cases was estimated to be 5 people in a month at mine X in Shurugwi. What is the probability that 4 people will test positive for COVID-19 at mine X? [3]
- c) Suppose a random sample $n=25$ is selected from a population that is normally distributed with a mean equal to 106 and a variance of 144. What is the probability that the sample mean is between 102 and 110? [5]