ZIMBABWE EZEKIEL GUTI UNIVERSITY

FACULTY OF SCIENCE, TECHNOLOGY, AGRICULTURE AND FOOD SYSTEMES DEVELOPMENT

DEPARTMENT OF DATA SCIENCE AND COMPUTER TECHNOLOGY

EXAMINATION PAPER

COURSE CODE	:	BIS122
COURSE TITLE	:	APPLIED STATISTICS
SPECIAL REQUIREMENTS	:	**None**
DURATION	:	3 Hours
LEVEL	:	1.1
DATE	:	2025

28 JUL 2025

INSTRUCTIONS TO CANDIDATES:

1. Answer all questions.
2. Show all your working
3. Calculators and slide rules are allowed
4. Each question carries 20 marks

Question 1

a. With the aid of a Venn diagram explain a non-mutually exclusive event.
[3]

b. A medical centre encourages people to take a flue vaccine each year. The vaccination reduces the likelihood of getting flue from 40% to 10%.

If 45% of the elderly people visiting the medical centre, have been vaccinated, Find the probability that an elderly person chosen at random.
[3]

i. Gets flue
ii. Has been vaccinated, given that that they get flue
[4]

c. Find the probability of drawing a 10 or a spade from a deck of 52 cards.
[3]

d. A box contains 4 red balls, 3 blue balls and 5 green balls. Tobby draws one ball at random from the box.
i. What is the probability that the ball is either green or red?
[2]

e. Define a conditional probability. [2]

f. James took two tests. The probability of him passing both tests is 0.3. The probability of him passing the first test is 0.4. What is the probability of him passing the second test given that he has passed the first test?
[3]

Question 2

a. Given the following data set 8, 10, 12, 13, 13, 15, 15, 16, 16, 18, 13, 23, 24, 22

Find the following

i. Mean [2]
ii. Range [1]
iii. Lower quartile
[3]
iv. Upper quartile
[3]
v. Inter quartile range [1]

b. Below are the ages of women who attended the agricultural show held in Gwanda in 2011:

45, 13, 31, 41, , 29, 10, 12,40, 13, 14, 29, 28, 26, 25,30, 24, 23, 20, 17, 44, 10, 23, 43, 50, 46, 47, 13, 28, 29, 30

i. Using the scale of 10-50 and the class width of 5 make a table to present the grouped data.
[3]
ii. Calculate the mean. [2]
iii. Calculate the variance. [3]
iv. Find the standard deviation.
[2]

Question 3

A discrete random variable X has the following probability distribution

$$P(X=x) = \begin{cases} a(3-x) & x=0,\ 1,\ 2 \\ b & x=3 \end{cases}$$

i. Make a discrete probability table for the above using variables a and b. **[2]**

Given that $E(X)= 1.6$

ii. Find the values of a and b **[4]**
iii. Find $P(X<2)$ **[2]**
iv. Find $P(0.5 <X<3)$ **[2]**
v. Find $E(3X-2)$ **[3]**
vi. Show that the $Var(X) =1.64$ **[4]**
vii. Calculate $Var(3X-2)$ **[3]**

Question 4

a. Define the following terms giving examples
   i. Population **[2]**
   ii. Data **[2]**
   iii. Census **[2]**
   iv. Data set **[2]**
   v. Statistics **[2]**

b. Marks scored by 20 students in a mathematical test are as follows:

Boys	35	46	49	58	60	67	78	85	91	95
Girls	58	64	67	68	69	73	75	80	84	85

   a. Draw a back-to-back stem and leaf plot to show this data. **[5]**
   b. Explain the performance of boys and girls for the data. **[5]**

Question 5

a. Give three advantages of using sampling when doing research rather than using the whole population.

[3]

b. Ben wants to find out about the method of transport that the students use to get to school. The school has a total enrolment of 500 students. She plans to ask the first 50 students arriving at school one morning how they have travelled.

i. What type of sampling method is this?

[1]

ii. Give one advantages of using this method. **[2]**

iii. Give one limitations of using this method. **[2]**

c. Thando thinks it would be better to use the systematic sampling technique in the activity in 5b above. She is going to use the database of all the 500 students at the school to select her sample.

i. Explain how Thando would use the database to select the sample.

[5]

ii. Give one advantage of using this method. **[2]**

iii. Give one disadvantage of using this method.

[2]

d. Identify the sampling procedures that would be appropriate in the following situation explaining your reasoning and how it can be carried out.

A consumer protection body wishes to estimate the proportion of trains that are running late in carrying passengers to work.

[3]

FORMULA SHEET

Probability distribution

$0 \leq P(A) \leq 1$

$\sum P(x)=1$

Mean = $\mu = x' = \sum x * P(x)$

$E(X) = \sum x_i * P_i$

Variance = $\sigma^2 = Var(X) = \sum x^2 * P(x) - \mu$

$= \sum [(x^2 - \mu)^2 P(x)$

$Var(X) = E(X^2) - (E(X))^2$

$Var(aX+b) = a^2 Var(X)$

$E(X^2) = \sum x^2 * P(x)$

$E(aX-b) = aE(X) - b$

Standard deviation = σ = Sqrt (σ^2)

Grouped Data

Mean = $\sum f(x) / \sum f$

Mean deviation = $\sum f(|x - \mu|) / \sum f$

Variance = $\sum f(|x - \mu|)^2 / \sum f$

Standard deviation = Sqrt(variance)

Linear Regression and co-relation

$y = ax + b$

$a = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$

$b = \frac{\sum y \sum x^2 - \sum x \sum xy}{n(\sum x^2) - (\sum x)^2}$

$S_x = Sqrt\ [(\sum x^2 - (\sum x)^2/n)/n-1]$

$S_y = Sqrt\ [(\sum y^2 - (\sum y)^2/n)/n-1]$

$r = S_x * a / S_y$

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[\, n\sum x^2 - (\sum x)^2 \,][\, n\sum y^2 - (\sum y)^2 \,]}}$$

I am