GWANDA STATE UNIVERSITY

Faculty of Computational Sciences

DEPARTMENT OF MATHEMATICS AND STATISTICS

Applied Statistics

CMS 1202

Examination Paper

April 2025

This examination paper consists of 3 printed pages

Time Allowed: **3 hours**

Total Marks: **100**

Examiner's Name: Mr. E. Utete

INSTRUCTIONS

Answer **ALL** questions in Section A and **ANY THREE** questions in Section B

ADDITIONAL REQUIREMENTS

Scientific calculator

Graph papers

Statistical Tables

Section A: Answer ALL Questions

Total Marks: 40

A1 The following data represents the ages (in years) of 8 individuals:

$$22, 25, 27, 30, 32, 35, 40, 25$$

(a) Calculate the mean, median, and mode of the dataset. [4]

(b) Compute the range, variance, and standard deviation. [5]

(c) Construct a box plot for the data and identify any outliers, if present. [5]

A2 (a) A dataset has a mean of 60 and a standard deviation of 12. If each value in the dataset is multiplied by 3 and then decreased by 5, what will be the new mean and standard deviation? [4]

(b) Explain the difference between **population parameters** and **sample statistics**. Provide an example of each. [4]

A3 The number of emails received per hour follows a Poisson distribution with a mean of 4.

(a) Calculate the probability of receiving exactly 3 emails in an hour. [3]

(b) Calculate the probability of receiving at least 2 emails in 30 minutes. [4]

A4 The following table shows the observed frequencies of a categorical variable:

Category	Observed Frequency
A	25
B	40
C	35

Test whether the categories are equally likely using a chi-square goodness-of-fit test at a 5% significance level. [11]

Section B: Answer THREE out of FOUR Questions

Total Marks: 60

Each question in this section is worth **20 marks**.

B5 (a) The following data represents the monthly sales (in thousands) for a company over 6 months:

$$50, 55, 60, 65, 70, 75$$

i. Calculate the 3-month moving average. [4]

ii. Fit a linear trend line to the data and predict the sales for the 7th month. [8]

(b) Explain the components of a time series and how they can be identified in a dataset. [8]

B6 (a) The following table shows the cross-tabulation of two categorical variables, A and B:

	$B1$	$B2$	Total
$A1$	20	30	50
$A2$	25	25	50
Total	45	55	100

i. Calculate the expected frequencies for each cell under the assumption of independence. [4]

ii. Perform a chi-square test for independence at a 5% significance level. [10]

(b) Explain the limitations of the chi-square test. [6]

B7 (a) A dataset is normally distributed with a mean of 100 and a standard deviation of 15.

i. Calculate the probability that a randomly selected value is greater than 120. [3]

ii. Calculate the probability that a randomly selected value is between 90 and 110. [4]

(b) Explain the concept of **skewness** in a dataset. How does positive skewness differ from negative skewness? [7]

(c) Define kurtosis, how does it differ from skewness? [6]

B8 (a) The following data represents the heights (in cm) of 12 individuals:

$$160, 165, 170, 175, 180, 185, 190, 195, 174, 182, 184, 178$$

i. Construct a stem and leaf plot for the data. [5]

ii. Calculate the median of the data. [3]

iii. What are the advantages to using the median as a measure for central tendency compared to the mean ? [6]

(b) Explain the importance of graphical representation of data in statistics. Provide two examples of graphs and their uses. [6]

End of Examination Paper

Total: 100 marks.