"Investing in Africa's Future"

COLLEGE OF ENGINEERING AND APPLIED SCIENCES

NCIS 208: APPLIED STATISTICS

END OF FIRST SEMESTER EXAMINATIONS

NOVEMBER 2025

LECTURER: MR. E. CHIKAKA

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 – Short Answer Questions [40 Marks]
(Answer ALL questions. Each question carries 5 marks.)

Question 1
Briefly explain the role of statistics in IT decision-making, giving one example from **cybersecurity**. [5]

Question 2
Differentiate between **categorical data** and **numeric data**, providing one Computer Information Systems (CIS) example for each. [5]

Question 3
Define **percentile** and **quartile**, and explain their usefulness in **user experience (UX) analysis**. [5]

Question 4
List and describe any **two data quality issues** commonly encountered in CIS datasets, and suggest how they may be addressed. [5]

Question 5
A monitoring system reports that the probability of a server crash on a given day is **0.1**. What is the probability that the server will run without crashing for **5 consecutive days**? [5]

Question 6
Give one real-world example each of a situation in CIS where the following probability distributions can be applied:
(a) Binomial
(b) Poisson
(c) Normal [5]

Question 7
What is the difference between a **one-sample t-test** and a **two-sample t-test**? Provide one CIS-related application for each. [5]

Question 8
Define **multicollinearity** in multiple regression and explain why it can be problematic when modelling CIS data. [5]

SECTION B – Structured Questions [60 Marks]

(Answer ANY THREE questions. Each question carries 20 marks.)

Question 9 – **Descriptive Statistics & Visualization**

A system administrator collects **response times (in milliseconds)** from a web server during peak hours:
120, 140, 110, 150, 160, 130, 170, 180, 125, 135.
(a) Compute the **mean, median, variance, and standard deviation** of the dataset. [*8*]
(b) Construct a box plot and identify any outliers. [*6*]
(c) Explain how such descriptive statistics can be used for **performance monitoring and decision-making**. [*6*]

Question 10 – **Probability Distributions**

(a) A firewall detects an average of **4 intrusion attempts per hour**. Assuming a **Poisson distribution**, calculate the probability of:
i. Exactly 3 intrusion attempts in an hour. [*3*]
ii. At most 2 intrusion attempts in an hour. [*4*]
(b) A user claims that **60% of emails received are spam**. If a sample of 10 emails is taken, find the probability that **exactly 7 are spam** using the **Binomial distribution**. [*5*]
(c) Discuss why the **Normal distribution** is commonly applied in **network traffic analysis**. [*8*]

Question 11 – **Inferential Statistics & Hypothesis Testing**

A company wants to test whether a new **data compression algorithm** reduces average file size compared to the current method. The current method produces an average file size of **5.0 MB**. A sample of **20 files** processed with the new algorithm yields a mean of **4.6 MB** with a standard deviation of **0.5 MB**.
(a) Formulate the **null and alternative hypotheses**. [*4*]
(b) Conduct a **one-sample t-test** at the 5% significance level. Show all steps. [*10*]
(c) Interpret the results in the context of IT decision-making. [*6*]

Question 12 – **Regression & Correlation**

A UX team collects data on **page load time (X, in seconds)** and **user satisfaction rating (Y, on a scale of 1–10)** from 8 web pages.

Page Load Time (X)	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
User Satisfaction (Y)	9	8	7	6	6	5	4	3

(a) Compute the **Pearson correlation coefficient** between X and Y, and interpret its value. [*8*]
(b) Fit a **simple linear regression model** of Y on X. [*6*]
(c) Use the model to **predict satisfaction rating** when load time = 10 seconds. [*3*]
(d) Discuss the limitations of using regression for decision-making in UX. [*3*]

Question 13 – **Time Series & Forecasting**

The number of daily logins to a cloud system over **8 consecutive days** is recorded as follows:

Day	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Logins	200	220	210	250	240	230	260	270

(a) Plot the time series and comment on any **trend or pattern** observed. [*4*]
(b) Compute the **3-day moving average** for Days 3–8. [*6*]
(c) Apply a **simple exponential smoothing** method with $\alpha = 0.3$, assuming the initial forecast for Day 1 = 200. Forecast logins up to Day 5. [*6*]
(d) Explain how **time series forecasting** can be applied in **capacity planning for cloud systems**. [*4*]

END OF EXAMINATION