WAHIDA ABUBAKAR MALAM

MAAUN/24/DTS/2001

DATA SCIENCE

STATISTICAL COMPUTING AND MODELLING

QUESTIONS

1. EXPLAIN THE DIFFERENCE BETWEEN A POPULATION AND A SAMPLE WITH REAL WORLD EXAMPLESS.

2. DESCRIBE THREE TYPES OF STATISTICAL MODELS AND GIVE ONE EXAMPLE OF THEIR APPLICATION.

Question 1

A population is the entire group that you want to study, while a sample is a smaller subset of that population selected for analysis. Researchers use samples because studying an entire population is often impractical.

Real-World Examples

Election Polling

Population: All eligible voters in a country.

Sample: A few thousand voters selected to predict the election outcome.

Medical Research

Population: All people with a specific disease.

Sample: A group of 500 patients chosen to test a new treatment.

Customer Satisfaction Survey

Population: All customers of a company.

Sample: A random selection of 1,000 customers who completed a survey.

In each case, the goal is to draw conclusions about the population based on data from the sample while minimizing errors and biases.

Statistical models help analyze data and make predictions. Here are three common types of statistical models, along with their applications:

QUESTION 2

1. Linear Regression Model

Description: This model examines the relationship between a dependent variable and one or more independent variables using a straight-line equation.

Example Application: Predicting house prices based on factors like square footage, number of bedrooms, and location.

2. Logistic Regression Model

Description: Used for classification problems where the outcome is categorical (e.g., yes/no, pass/fail). It estimates probabilities using the logistic function.

Example Application: Predicting whether a customer will buy a product based on demographics and past behavior.

3. Time Series Model

Description: Analyzes data points collected over time to identify trends, seasonality, and future patterns.

Example Application: Forecasting stock prices or sales trends in a retail business.

Each model serves different analytical purposes, depending on the nature of the data and the research question.

QUESTION 3

import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

filename = "data.csv"

df = pd.read\_csv(filename)

data = [10,20,30,40,50,60,70,80,90,100]

mean\_value = np.mean(data)

variance\_value = np.var(data)

std\_deviation = np.std(data)

print(f"Mean: {mean\_value}\n")

print(f"Variance: {variance\_value}\n")

print(f"Standard Deviation: {std\_deviation}\n")

plt.hist(data, bins=5, color='blue', alpha=0.7)

plt.xlabel("Data Values")

plt.ylabel("Frequency")

plt. title("Histogram of Sampe Data")

plt.show()