Assignment Code: DA-AG-006

Statistics Advanced - 1| **Assignment**

Instructions: Carefully read each question. Use Google Docs, Microsoft Word, or a similar tool to create a document where you type out each question along with its answer. Save the document as a PDF, and then upload it to the LMS. Please do not zip or archive the files before uploading them. Each question carries 20 marks.

Total Marks: 200

Question 1: What is a random variable in probability theory?

Answer:

A random variable is a function that assigns a numerical value to each outcome of a random experiment.

Example: In tossing a coin, let (X = 1) for Head and (X = 0) for Tail.

Question 2: What are the types of random variables?

Answer:

1. Discrete Random Variable – Takes countable values (finite or countably ir
   Example: Number of students in a class, number of heads in coin tosses.

2. Continuous Random Variable – Takes any value within a range.
   Example: Height, weight, time.

Question 3: Explain the difference between discrete and continuous distributions.

Answer:

Discrete distribution deals with random variables that take countable values (finite or co
Example: Number of students, number of defects.

Continuous distribution deals with random variables that can take any value within a giv
Example: Height, weight, time.

Question 4: What is a binomial distribution, and how is it used in probability?

Answer:

A binomial distribution is a discrete probability distribution that describes the number of successes in a fixed number of independent trials, where each trial has only two outcomes (success or failure) and a constant probability of success.
Use in Probability:
It is used to calculate the probability of getting a certain number of successes, such as the number of heads in coin tosses or the number of defective items in a batch.

Question 5: What is the standard normal distribution, and why is it important?

Answer:

The standard normal distribution is a normal distribution with mean 0 and standard deviation 1.
It is denoted by (Z \sim N(0,1)).

Importance:
It is important because any normal distribution can be converted into a standard normal distribution using the z-score, which makes it easy to calculate probabilities using standard normal tables.

Question 6: What is the Central Limit Theorem (CLT), and why is it critical in statistics?

Answer:

Central Limit Theorem (CLT):
The Central Limit Theorem states that the sampling distribution of the sample mean approaches a normal distrib
size becomes large, regardless of the population's original distribution, provided the samples are independent a
distributed.

Why it is critical:
It is critical because it allows statisticians to use normal distribution methods for inference (confidence intervals
testing) even when the population distribution is unknown.

Question 7: What is the significance of confidence intervals in statistical analysis?

Answer:

Significance of Confidence Intervals in Statistical Analysis:
A confidence interval provides a range of values within which the true population parameter (such as mean or prop
with a specified level of confidence (e.g., 95%).
Significance:
It shows the precision and reliability of an estimate.
It helps in decision-making and inference by indicating the possible range of the true value.

Question 8: What is the concept of expected value in a probability distribution?

Answer:

Expected Value:
The expected value is the average or mean value of a random variable, calculated by weighting each possible
outcome by its probability.
Concept:
It represents the long-run average outcome if the random experiment is repeated many times.

For a discrete variable:
(E(X) = \sum x , P(x))

Question 9: Write a Python program to generate 1000 random numbers from a normal distribution with mean = 50 and standard deviation = 5. Compute its mean and standard deviation using NumPy, and draw a histogram to visualize the distribution.

(*Include your Python code and output in the code box below.*)

Answer:

```
import numpy as np
import matplotlib.pyplot as plt
data = np.random.normal(loc=50, scale=5, size=1000)
mean = np.mean(data)
std_dev = np.std(data)
print("Mean:", mean)
print("Standard Deviation:", std_dev)
plt.hist(data, bins=30)
plt.xlabel("Values")
plt.ylabel("Frequency")
plt.title("Histogram of Normal Distribution")
plt.show()
Output (sample):Mean ≈ 50.06
Standard Deviation ≈ 5.21 The histogram shows a bell-shaped curve, confirming a nor
```

Question 10: You are working as a data analyst for a retail company. The company has collected daily sales data for 2 years and wants you to identify the overall sales trend.

daily_sales = [220, 245, 210, 265, 230, 250, 260, 275, 240, 255,

235, 260, 245, 250, 225, 270, 265, 255, 250, 260]

- Explain how you would apply the Central Limit Theorem to estimate the average sales with a 95% confidence interval.
- Write the Python code to compute the mean sales and its confidence interval.

(*Include your Python code and output in the code box below.*)

Answer:

```
By the Central Limit Theorem, the sample mean of daily sales follows a normal distribution, allowing us to estimate the average sales and calculate a 95%
the sample mean and standard deviation.
import numpy as np
import math
daily_sales = [220, 245, 210, 265, 230, 250, 260, 275, 240, 255]
mean = np.mean(daily_sales)
std = np.std(daily_sales, ddof=1)
n = len(daily_sales)
margin = 1.96 * (std / math.sqrt(n))
print("Mean Sales:", mean)
print("95% CI:", (mean - margin, mean + margin))
```