Statistics Advanced - 1| Assignment Answers

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

Answer:-A random variable is a variable that represents the outcomes of a random phenomenon. It assigns numerical values to each outcome in a sample space. There are two types of random variables: **discrete** (which takes specific values, like 1, 2, 3) and **continuous** (which can take any value within a range, like 2.5, 3.14, etc.). Random variables help in quantifying uncertainty and are essential in defining probability distributions.

Question 2: What are the types of random variables?

Answer:- The two main types of random variables are:

1. **Discrete Random Variables**: These take a countable number of distinct values. Example: Number of heads in a coin toss.
2. **Continuous Random Variables**: These take an infinite number of possible values within a range. Example: Height of a person, temperature

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

Answer:-1 .**Discrete distributions** describe the probability of outcomes of a discrete random variable.

Example: Binomial distribution.

2.**Continuous distributions** describe the probability of outcomes of a continuous random variable.

Example: Normal distribution.

- The main difference is that **discrete** distributions use **probability mass functions (PMFs)**, while **continuous** distributions use **probability density functions (PDFs)**.

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

Answer:-The binomial distribution models the number of successes in a fixed number of independent Bernoulli trials, where each trial has two outcomes: success or failure. It is used when the probability of success is constant in each trial.

Formula:

P(X = k) = C(n, k) * p^k * (1-p)^(n-k)

Example: Probability of getting 3 heads in 5 coin tosses.

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

Answer:-The standard normal distribution is a normal distribution with a **mean of 0** and **standard deviation of 1**. It is important because:

1. It simplifies calculations using **Z-scores.**
2. It is used in hypothesis testing and confidence intervals.
3. Many statistical methods assume normality, and raw data can often be standardized to fit this distribution.

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

Answer:-The Central Limit Theorem states that the sampling distribution of the sample mean will tend to be normally distributed, even if the population distribution is not, **provided the sample size is sufficiently large** (typically $n \geq 30$).

It is critical because it allows us to use normal probability techniques for inference, even when the population is not normal.

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

Answer:-Confidence intervals provide a **range of values** within which we expect the true population parameter (like mean or proportion) to lie, with a given level of confidence (e.g., 95%).

They help in:

1. Estimating unknown parameters.
2. Indicating the reliability of the estimate.
3. Supporting decision-making under uncertainty.

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

Answer:-The expected value is the **average** or **mean** value of a random variable over many repetitions of an experiment.

For a discrete random variable:

$E(X) = \Sigma [x * P(x)]$

It gives the center or "balance point" of the distribution and helps in understanding the long-term average outcome.

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


# Generate 1000 random numbers from normal distribution

data = np.random.normal(loc=50, scale=5, size=1000)


# Compute mean and standard deviation

mean = np.mean(data)

std_dev = np.std(data)


print(f"Mean: {mean}")

print(f"Standard Deviation: {std_dev}")


# Plot histogram

plt.hist(data, bins=30, edgecolor='black')
```

```
plt.title("Normal Distribution (μ=50, σ=5)")

plt.xlabel("Value")

plt.ylabel("Frequency")

plt.show()
```

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:-

```
import numpy as np

import scipy.stats as stats


# Daily sales data

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

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

```
# Compute mean and standard error
mean = np.mean(daily_sales)
std_dev = np.std(daily_sales, ddof=1)
n = len(daily_sales)
standard_error = std_dev / np.sqrt(n)


# 95% confidence interval using CLT
confidence_level = 0.95
z_score = stats.norm.ppf(1 - (1 - confidence_level) / 2)
margin_of_error = z_score * standard_error
confidence_interval = (mean - margin_of_error, mean + margin_of_error)


print(f"Mean Sales: {mean}")
print(f"95% Confidence Interval: {confidence_interval}")
```

Explanation:

Using the Central Limit Theorem, we assume that the sample mean of sales will follow a normal distribution. We calculate the mean and standard error, then apply the Z-distribution to estimate the 95% confidence interval.