Statistics Advanced - 1| Assignment

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

Answer: A random variable in probability theory is a function that assigns a numerical value to each possible outcome of a random experiment. It converts outcomes like "heads or tails" or "win or lose" into numbers so we can analyze them mathematically.

Question 2: What are the types of random variables?

Answer: Types of random variables:-
1.Discrete random variable – takes countable values.
Example: number of students present in a class.
2.Continuous random variable – takes values from a continuous range.
Example: height or weight of a person.

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

Answer:

Discrete Distribution	Continuous Distribution
It is used for discrete random variables.	It is used for continuous random variables.
Takes countable, separate values (usually integers).	Takes any value within a range (uncountable).
Represents using Probability Mass Function (PMF).	Represents using Probability Density Function (PDF).
The sum of all probabilities equals 1.	The total area under the curve equals 1.

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

Answer:A binomial distribution is a discrete probability distribution that models the number of successes in a fixed number of repeated, independent trials of the same experiment.
Each trial has only two possible outcomes:
- Success
- Failure

Conditions for a Binomial Distribution :-
A random experiment follows a binomial distribution if :
1. The number of trials nnn is fixed.
2. Each trial is independent.
3. Each trial has two outcomes (success/failure).
4. The probability of success ppp is constant for every trial.

Example : If a coin is tossed 10 times and the probability of getting heads is 0.5, the binomial distribution can be used to find the probability of getting exactly 6 heads.

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

Answer: The standard normal distribution is a special case of the normal distribution that is centered at zero with a standard deviation of one.

<u>Why It Is Important :-</u>

1. Standardization (Z-scores) : Any normal distribution can be converted into the standard normal distribution.

2. Probability Calculation : Probabilities for all normal distributions can be found using one standard normal table (Z-table).

3.**Statistical Analysis :** Widely used in hypothesis testing, confidence intervals, and inferential statistics.

4. Real-World Applications : Used in exam scores, quality control, finance, and social science.

Example : If a student's test score is converted into a Z-score, the standard normal distribution tells us how that score compares to the average.

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

Answer: The Central Limit Theorem states that : If we take a sufficiently large number of random samples from any population (with finite mean and variance), the distribution of the sample means will approach a normal distribution, regardless of the population's original distribution.

<u>Why CLT Is Critical in Statistics :</u>

1.Foundation of Inferential Statistics - Enables hypothesis testing and confidence intervals.

2.Works for Any Distribution - Even if the population is skewed or unknown.

3. Simplifies Probability Calculations - Allows use of the normal distribution in many practical cases.

4.Real-World Decision Making - Used in quality control, surveys, finance, and scientific research.

Example: If we repeatedly take samples of size 50 from a population with unknown distribution and calculate their means, those means will form a normal distribution.

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

Answer: A confidence interval is a range of values, calculated from sample data, that is likely to contain the true population parameter (such as a mean or proportion).

Example: A 95% confidence interval means that if we repeated the sampling many times, about 95% of the intervals would contain the true parameter.

<u>Significance of Confidence Intervals:</u>

1. Measure of Uncertainty - They show the **margin of error** around an estimate.

2. More Informative Than Point Estimates - A single value (like a sample mean) hides uncertainty; a CI reveals it.
3. Supports Decision-Making - Helps assess whether a value is plausible or not (e.g., whether a new process meets quality standards).
4. Used in Hypothesis Testing - If a hypothesized value lies outside the CI, it may be rejected at that confidence level.

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

Answer: The expected value (EV) of a probability distribution represents the long-run average outcome of a random variable if an experiment is repeated many times.
Example: If a game pays ₹100 with probability 0.2 and ₹0 with probability 0.8:
$E(X)=100(0.2)+0(0.8)=20$
So, the expected value is ₹20 per game in the long run.

It is used in decision-making, risk analysis and helps to compare different strategies or experiments.

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.

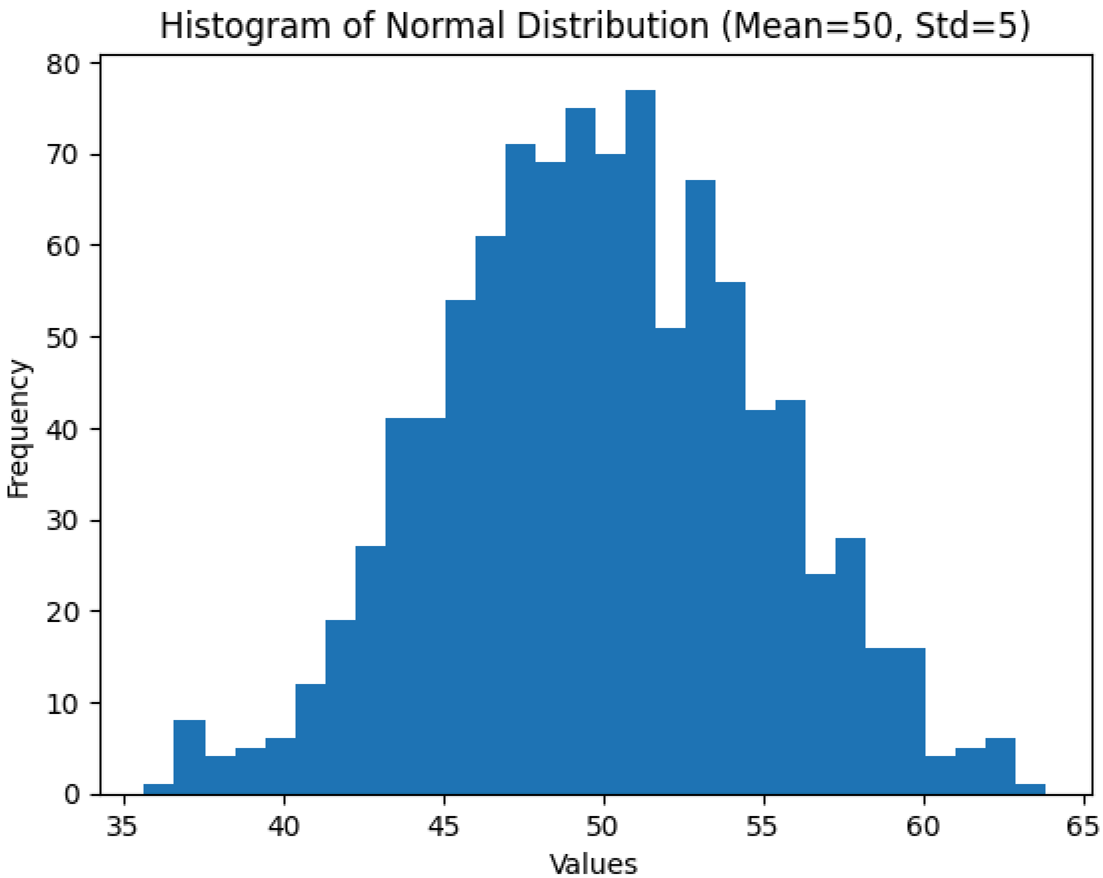
Answer:
```
import numpy as np
import matplotlib.pyplot as plt

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

mean_value = np.mean(data)
std_value = np.std(data)

print("Mean:", mean_value)
print("Standard Deviation:", std_value)

plt.hist(data, bins=30)
plt.xlabel("Values")
plt.ylabel("Frequency")
plt.title("Histogram of Normal Distribution (Mean=50, Std=5)")
plt.show()
```
Output:-
```
Mean: 49.910133017949235
Standard Deviation: 4.939680919876479
```



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

Answer: Using CLT to estimate average sales

1. Treat the given daily sales as a sample.
2. Compute: Sample mean and Sample standard deviation
3. Use the normal distribution to construct a 95% confidence interval:

Now we are 95% confident that the true average daily sales lies within this interval.import numpy as np

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

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

data = np.array(daily_sales)

mean_sales = np.mean(data)

std_sales = np.std(data, ddof=1)

n = len(data)

z = 1.96

margin_error = z * (std_sales / np.sqrt(n))

lower_bound = mean_sales - margin_error

upper_bound = mean_sales + margin_error

print("Mean Daily Sales:", mean_sales)

print("95% Confidence Interval:", (lower_bound, upper_bound))
```

Output:-

Mean Daily Sales: 248.25

95% Confidence Interval: (np.float64(240.68312934041109), np.float64(255.81687065958891))