Q(1). Define the z-statistic and explain its relationship to the standard normal distribution. How is the z-statistic used in hypothesis testing?

ANSWER:--

Z-Statistic Definition

The z-statistic (or z-score) is a standardized value that quantifies how far an individual data point or sample mean is from the population mean in terms of standard deviations. It is a tool used in statistics to determine the position of a data point within a distribution, enabling comparison to the standard normal distribution.

The formula for a z-score is:

$$z$$

$$X - \mu / \sigma$$

Where:

X is the observed value (data point),

μ is the population mean,

σ is the population standard deviation.

This formula converts any raw score into a z-score, telling us how many standard deviations X is away from the mean. If z=1, the data point is 1 standard deviation above the mean; if z=-2, it is 2 standard deviations below the mean.

In the context of hypothesis testing with sample data, when we are dealing with the sample mean (¯X¯) instead of an individual data point, the formula becomes:

```
z= ¯X¯-μ/σ
```

Where: ¯X¯ is the sample mean, μ is the hypothesized population mean,

σ is the population standard deviation,

n is the sample size.

This version of the formula is used to compare the sample mean to the population mean and standardizes the result relative to the sampling distribution.

Relationship to the Standard Normal Distribution

The z-statistic is closely related to the standard normal distribution. The standard normal distribution is a special type of normal distribution with:

a mean (μ) of 0,

a standard deviation (σ) of 1.

When you calculate a z-score, you are standardizing the data, converting it into a form where the mean is 0 and the standard deviation is 1, allowing comparison against the standard normal distribution. This transformation is essential because it allows you to use probability tables (called z-tables) to determine the likelihood of observing certain values.

For example:

A z-score of 0 corresponds to the mean of the distribution.

A z-score of 1 corresponds to 1 standard deviation above the mean, and around 84% of the data falls below this value.

A z-score of -1 corresponds to 1 standard deviation below the mean, and around 16% of the data falls below this value.

The probability values associated with z-scores are used to make inferences about the data, particularly in hypothesis testing.

Use of the Z-Statistic in Hypothesis Testing

The z-statistic is widely used in hypothesis testing, especially when testing population means or proportions, and when the population standard deviation is known. The z-test is used to determine whether the difference between an observed sample statistic (e.g., sample mean) and a hypothesized population parameter (e.g., population mean) is statistically significant.

Here ’s how the z-statistic is used step-by-step in hypothesis testing:

#1. State the Hypotheses You begin by defining the null hypothesis (H0) and the alternative hypothesis (HA).

The null hypothesis assumes no effect or no difference. It typically states that the population parameter (e.g., population mean) is equal to a specific value. For example, H0:μ=μ0, where μ0 is the hypothesized value.

The alternative hypothesis suggests that there is a significant effect or difference. It asserts that the population parameter is different from the hypothesized value, e.g., $H\ A : \mu \neq \mu\ 0$

#2. Select the Significance Level (α)

You choose a significance level (α), typically 0.05, which defines the threshold for rejecting the null hypothesis. This means that if the probability of observing the sample result under the null hypothesis is less than 5%, the result is considered statistically significant.

#3. Compute the Z-Statistic

You calculate the z-statistic using sample data. For a hypothesis test involving the sample mean.

Here's what this calculation tells you:

If the sample mean (¯X¯) is far from the hypothesized population mean (μ), the z-statistic will be large (in magnitude).

The larger the z-statistic, the less likely it is that the observed sample mean occurred by random chance if the null hypothesis is true.

#4. Determine the p-value

The p-value represents the probability of obtaining a z-statistic as extreme as the one calculated (or more extreme) under the assumption that the null hypothesis is true. You use the z-statistic to look up the corresponding p-value in a z-table (or use software to calculate it).

For a two-tailed test, the p-value is the probability of observing a z-score as extreme as the calculated value in both directions (positive and negative). For a one-tailed test, the p-value is the probability of observing a z-score as extreme as the calculated value in one direction (either positive or negative, depending on the alternative hypothesis).

#5. Make a Decision

Once you have the p-value, you compare it to your significance level (α):

If the p-value is less than or equal to α, you reject the null hypothesis. This means the sample provides enough evidence to suggest that the observed result is statistically significant, and the difference between the sample mean and the population mean is unlikely to be due to chance. If the p-value is greater than α, you fail to reject the null hypothesis. This means the evidence from the sample is not strong enough to conclude that there is a significant difference, and the result could be due to random chance.

Q (2) .What is a p-value, and how is it used in hypothesis testing? What does it mean if the p-value is very small (e.g., 0.01)?

ANSWER:--

Definition of a P-Value

A p-value (probability value) is a measure used in hypothesis testing to quantify the strength of the evidence against the null hypothesis ($H0$). It represents the probability of observing the sample data, or something more extreme, assuming that the null hypothesis is true. In simpler terms, the p-value tells us how likely it is to get the observed results purely by chance if there is no real effect or difference.

Mathematically, the p-value is the area under the curve of the sampling distribution beyond the test statistic (e.g., z-score or t-score). It ranges between 0 and 1:

A small p-value indicates strong evidence against the null hypothesis.

A large p-value suggests that the observed data are likely under the null hypothesis, so there is no strong evidence to reject it.

P-Value in Hypothesis Testing

In hypothesis testing, the p-value is used to make decisions about whether to reject the null hypothesis. The typical process involves the following steps:

#1. Formulate Hypotheses:

The null hypothesis (H0) assumes there is no effect or difference, such as "the population mean is equal to a specified value."

The alternative hypothesis (HA) suggests there is an effect or difference, such as "the population mean is not equal to the specified value."

#2. Choose a Significance Level (α):

The significance level (α) is the threshold for determining statistical significance, commonly set at 0.05 (or 5%). This means we are willing to accept a 5% risk of rejecting the null hypothesis when it is actually true (Type I error).

#3. Calculate the Test Statistic and P-Value:

Based on the sample data, you compute a test statistic (e.g., z-score or t-score) and use it to find the corresponding p-value.

#4. Decision Rule:

If the p-value is less than or equal to the significance level (α), you reject the null hypothesis. This suggests that the observed data are unlikely to have occurred under the assumption of the null hypothesis.

If the p-value is greater than α, you fail to reject the null hypothesis. This indicates that the data do not provide enough evidence to conclude that there is a significant effect or difference.

Interpretation of a Very Small P-Value (e.g., 0.01) A small p-value, such as 0.01, means that there is only a 1% probability of obtaining the observed sample data (or more extreme) if the null hypothesis is true. In hypothesis testing, this would be considered strong evidence against the null hypothesis.

If p=0.01, and the significance level α=0.05, you would reject the null hypothesis because the p-value is smaller than α. This suggests that the observed results are unlikely to have occurred by random chance, and the difference or effect is statistically significant.

What Does a Very Small P-Value Mean?

#1. Strong Evidence Against the Null Hypothesis:

A p-value of 0.01 implies that there is only a 1% chance of obtaining the observed result under the assumption that the null hypothesis is true. This provides strong evidence to conclude that the null hypothesis is likely false.

#2. Statistical Significance:

Since the p-value is smaller than the significance level (typically α=0.05), the result is considered statistically significant. This means there is enough evidence to suggest that the observed effect or difference is not due to random variation alone.

#3. Practical Implications:

While a very small p-value indicates statistical significance, it is important to consider the practical significance of the result. A small p-value does not necessarily mean the effect is large or meaningful in real-world terms.

Example

Suppose you are testing whether a new medication lowers blood pressure more effectively than a standard treatment. Your null hypothesis is that there is no difference in effectiveness between the two treatments. You perform a study and find a p-value of 0.01.

If p=0.01 and α=0.05, you reject the null hypothesis and conclude that the new medication is significantly more effective than the standard treatment. The small p-value indicates that the likelihood of observing such a result by random chance (if there were no real difference) is just 1%.

Q (3). Compare and contrast the binomial and Bernoulli distributions.?

#ANSWER:--

The Binomial and Bernoulli distributions are both fundamental in probability theory and statistics, but they are used in different contexts and have distinct characteristics. Here’s a detailed comparison and contrast:

1. Definition

Bernoulli Distribution:

The Bernoulli distribution is a discrete probability distribution for a single trial (experiment) that can result in either success (with probability p) or failure (with probability 1−p).

It models a single binary outcome (e.g., flipping a coin once).

The random variable X can take only two values: 0 (failure) or 1 (success).

Probability mass function (PMF):

```
P(X=x)=px(1−p)1−x,x ∈ {0,1}
```

Binomial Distribution:

The Binomial distribution models the number of successes in n independent Bernoulli trials, where each trial has the same probability of success p.

It generalizes the Bernoulli distribution to multiple trials. The random variable X represents the total number of successes in n trials.

Probability mass function (PMF):

```
P(X=k)=(n/k)pk(1−p)n−k,k∈{0,1,2,…,n}
```

where (n/k) is the binomial coefficient.

2. Number of Trials

Bernoulli Distribution: Consists of only one trial.

Binomial Distribution: Consists of multiple trials (denoted by n).

#3. Random Variable

Bernoulli: The random variable represents the outcome of a single trial, which can either be 0 (failure) or 1 (success).

Binomial: The random variable represents the total number of successes across multiple trials, and it can take any integer value from 0 to n (inclusive).

#4. Parameters

Bernoulli Distribution: Has a single parameter p, the probability of success for a single trial.

Binomial Distribution: Has two parameters:

n (number of trials)

p (probability of success for each trial).

#5. Mean and Variance

Bernoulli Distribution:

Mean: $\mu=p$

Variance: $\sigma 2=p(1-p)$

Binomial Distribution:

Mean: $\mu=np$

Variance: $\sigma 2=np(1-p)$

#6. Relationship between the Two

A Bernoulli distribution is a special case of a Binomial distribution where n=1.

In this case, the Binomial PMF reduces to the Bernoulli PMF.

In other words, a Binomial distribution with n=1 is equivalent to a Bernoulli distribution.

#7. Use Cases

Bernoulli Distribution:

Used for modeling scenarios with a single trial or binary events.

Example:

Whether a single coin flip results in heads (success) or tails (failure).

Binomial Distribution:

Used for modeling the total number of successes in a fixed number of trials.

Example:

Counting the number of heads in 10 coin flips.

#8. Independence

Bernoulli Distribution:

There's no concept of multiple trials, so independence doesn't apply in this case.

Binomial Distribution:

Assumes that all n trials are independent of each other.

#Q (4).Under what conditions is the binomial distribution used, and how does it relate to the Bernoulli distribution?

Answer:--

Conditions for Binomial Distribution:

The binomial distribution is used when:

1. There are a fixed number of trials (n).
2. Each trial has only two possible outcomes (success or failure).
3. The probability of success (p) remains constant across trials.
4. Trials are independent.
5. The outcome of each trial does not affect the probability of subsequent trials.

#Relationship to Bernoulli Distribution:

The Bernoulli distribution is a special case of the binomial distribution, where:

1. Number of trials (n) = 1.
2. Only two possible outcomes (success or failure).

In other words, the Bernoulli distribution models a single trial, while the binomial distribution models multiple trials.

Binomial Distribution Parameters:

1. n (number of trials)
2. p (probability of success)
3. q = 1 - p (probability of failure)

Binomial Distribution Formula:

P(X = k) = (nCk) * p^k * q^(n-k)

where:

- X = number of successes
- k = 0, 1, ..., n
- nCk = number of combinations of n items taken k at a time

Real-World Applications:

1. Coin flips
2. Surveys
3. Quality control
4. Medical trials
5. Insurance risk assessment

Example:

Suppose we flip a fair coin 10 times. What's the probability of getting exactly 5 heads?

n = 10

p = 0.5

q = 0.5

k = 5

P(X = 5) = (10C5) * 0.5^5 * 0.5^5 ≈ 0.2461

Key Takeaways:

1. Binomial distribution models multiple independent trials.

2. Bernoulli distribution models a single trial.
3. Binomial distribution reduces to Bernoulli distribution when n = 1.

#Q (5).What are the key properties of the Poisson distribution, and when is it appropriate to use this distribution?

ANSWER:--

The Poisson distribution is a discrete probability distribution that models the number of events occurring within a fixed interval of time, space, or any other dimension, where these events happen independently and at a constant average rate. It is widely used for modeling rare events or the number of occurrences of an event in a specific time frame.

Key Properties of the Poisson Distribution

1.Discreteness:

The Poisson distribution is discrete, meaning it describes the probability of a count of events (e.g., 0, 1, 2, etc.) happening in a fixed interval.

2.Parameter (λ):

The Poisson distribution is characterized by a single parameter λ, which represents both the mean and the variance of the distribution. λ is the expected number of events occurring in the given interval.

λ>0, and it must be constant throughout the observation period.

3.Probability Mass Function (PMF):

The PMF of a Poisson-distributed random variable X is given by:

```
P(X=k)= λke−λ/k!,k=0,1,2,…
```

where k is the number of events and e is Euler’s number

4.Mean and Variance:

The mean of the Poisson distribution is λ.

The variance is also λ.

This property is unique to the Poisson distribution: the mean and variance are equal.

#5.Additivity:

If X 1 ∼ Poisson (λ 1) and X 2 ∼ Poisson (λ 2) are independent Poisson random variables, then their sum X 1

- X 2 is also Poisson-distributed: X 1
- X 2 ∼ Poisson (λ 1
- λ 2)

6.Memorylessness:

The Poisson process, which underlies the Poisson distribution, is memoryless, meaning the probability of an event occurring in a future time interval does not depend on the events that occurred before that time.

7.Skewness:

The Poisson distribution is positively skewed for small λ, meaning the distribution leans to the right. However, as λ increases, the distribution becomes more symmetric and approaches a normal distribution.

When to Use the Poisson Distribution

The Poisson distribution is appropriate to use when modeling the number of times an event occurs within a fixed interval under the following conditions:

1.Events Occur Independently:

The occurrence of one event does not affect the occurrence of another. Each event is independent of the others.

2.Constant Average Rate (λ):

The rate at which events occur is constant over the interval. The average number of events per unit of time, area, or volume remains the same.

Example: If 5 phone calls are expected per hour on average, this rate is constant over time.

3.Rare Events:

The Poisson distribution is often used to model rare or infrequent events.

Example: The number of car accidents at an intersection in a day or the number of errors per page in a book.

4.Single Event Occurrence:

Each event occurs one at a time, and multiple events do not occur simultaneously within the infinitesimally small sub-intervals.

Example: You are counting the number of arrivals at a bus stop in a given time frame, and arrivals happen one at a time.

5.Proportional to the Size of the Interval:

The probability of a certain number of events occurring is proportional to the size of the interval. For instance, doubling the time interval will double the expected number of events.

Examples of Poisson Distribution Applications

Telephone Calls: Modeling the number of incoming phone calls at a call center in a given period.

Traffic Accidents: Predicting the number of car accidents at a specific intersection over a month.

Queueing Theory: Estimating the number of people arriving at a service center (e.g., customers arriving at a bank or a restaurant).

Biology: Modeling the number of mutations in a strand of DNA over a given time frame.

Epidemiology: Modeling the occurrence of rare diseases in a population over a certain period.

Manufacturing: Predicting the number of defects in a product line in a certain number of batches or per unit.

Conditions Where the Poisson Distribution Is Not Appropriate

1.When Events Are Not Independent:

If one event affects the likelihood of another event, then the Poisson distribution may not be suitable.

2.When the Rate Varies Over Time:

If the event rate λ is not constant (e.g., if more events occur during peak hours), the Poisson distribution may not provide an accurate model.

3.When the Event Count Is Bounded:

The Poisson distribution assumes that the event count can take any non-negative integer, from 0 to infinity. If the event count is constrained by an upper limit, a different distribution might be more appropriate.

#Q (6).Define the terms "probability distribution" and "probability density function" (PDF). How does a PDF differ from a probability mass function (PMF)?

Answer:--

Probability Distribution:

A probability distribution describes the probability of occurrence of each possible value or range of values of a random variable. It assigns a non-negative real number (probability) to each possible outcome.

Probability Density Function (PDF):

A PDF is a continuous function that describes the probability distribution of a continuousrandom variable.

It satisfies:

1. $f(x) \geq 0$ (non-negativity)
2. $\int(-\infty \text{ to } \infty)\ f(x)\ dx = 1$ (normalization)

3. $P(a \le X \le b) = \int(a \text{ to } b)\ f(x)\ dx$ (probability calculation)

Probability Mass Function (PMF):

A PMF is a discrete function that describes the probability distribution of a discrete random variable.

It satisfies:

1. $p(x) \ge 0$ (non-negativity)
2. $\Sigma p(x) = 1$ (normalization)
3. $P(X = x) = p(x)$ (probability calculation)

Key differences between PDF and PMF:

1. Continuity: PDF (continuous) vs. PMF (discrete)
2. Probability calculation: PDF (integration) vs. PMF (summation)
3. Random variable type: PDF (continuous random variable) vs. PMF (discrete random variable)

Examples:

PDF:

- Normal distribution
- Exponential distribution
- Uniform distribution

PMF:

- Binomial distribution
- Poisson distribution
- Bernoulli distribution

Relationship between PDF and PMF:

For a continuous random variable X, the cumulative distribution function (CDF) can be obtained by integrating the PDF:

$F(x) = \int(-\infty \text{ to } x)\, f(t)\, dt$

For a discrete random variable X, the CDF is obtained by summing the PMF:

$F(x) = \Sigma p(t)$ for $t \leq x$

Key Takeaways:

1. Probability distribution describes the probability of occurrence of each possible value.
2. PDF models continuous random variables; PMF models discrete random variables.
3. PDF and PMF differ in continuity, probability calculation, and random variable type.

Q (7).Explain the Central Limit Theorem (CLT) with example.?

ANSWER:--

The Central Limit Theorem (CLT) is one of the most important and widely used theorems in statistics. It states that, given certain conditions, the distribution of the sample mean of a large number of independent and identically distributed (i.i.d.) random variables approaches a normal distribution (bell curve), regardless of the original distribution of the population from which the sample is drawn.

Key Points of the Central Limit Theorem (CLT):

1.Applies to Sample Means: The CLT applies to the distribution of the sample means, not individual data points.

2.Sample Size: As the sample size n increases, the sampling distribution of the sample mean becomes increasingly normal (approximately a normal distribution).

3.Independence and Identical Distribution: The random variables in the sample should be independent and identically distributed (i.i.d.), meaning they are drawn from the same probability distribution.

4.Any Population Distribution: The original population distribution can be any shape (e.g., skewed, uniform, etc.). The CLT tells us that the sampling distribution of the sample mean will still approach normality as the sample size increases.

5.Normality for Large Samples: Typically, for sufficiently large sample sizes (often $n \geq 30$), the sample mean is approximately normally distributed, even if the population distribution is not normal.

Example of the Central Limit Theorem:

Let's say we want to study the average number of hours people spend on social media per day in a large city. Assume the actual distribution of social media use is right-skewed, meaning most people spend fewer hours, but a few spend a lot of time online.

Step 1: The true population distribution is skewed and non-normal, which means the hours spent on social media are not normally distributed. Let's say the average (mean) number of hours is 3 hours per day, with a standard deviation of 2 hours.

Step 2: Now, suppose we randomly sample 50 people from this population and calculate the sample mean of social media use in hours.

Step 3: Repeat this process (sampling 50 people and calculating the sample mean) many times. According to the CLT, the distribution of these sample means will approach a normal distribution, even though the original distribution of hours spent on social media is skewed.

Step 4: With a large enough sample size (e.g., $n=50$), the mean of the sample means will be close to the population mean (3 hours), and the variance of the sample means will be smaller than the population variance.

Illustration:

Even if the original data distribution is skewed (as is often the case in real-world data), the sampling distribution of the sample means will resemble a normal distribution as long as the sample size is sufficiently large. If you take 1000 samples of 50 people each, and plot the distribution of sample means, it will look roughly like a normal distribution centered around 3 hours, regardless of the skewed nature of the original data.

Practical Importance of the CLT:

Statistical Inference: The CLT allows us to make inferences about the population mean using the sample mean, even if the population is not normally distributed.

Confidence Intervals and Hypothesis Testing: Since the sample means follow a normal distribution for large n, we can use the normal distribution to compute confidence intervals and conduct hypothesis tests.

Works with Any Distribution: No matter the original distribution (e.g., binomial, Poisson, skewed, etc.), the CLT assures us that the sample means will be approximately normal, making it easier to apply parametric statistical methods.

Visual Example:

Population Distribution: Suppose you have a population where the distribution of social media usage is skewed, with more people spending 1–2 hours and only a few spending 10–12 hours per day. The distribution might look asymmetric or skewed to the right.

Sample Mean Distribution: Now, take many samples of 50 people each and calculate their average social media usage. The distribution of these sample means will form a bell-shaped curve, even though the original population distribution is not bell-shaped. As the sample size increases, this distribution of sample means becomes closer to a normal distribution.

#Conclusion:

The Central Limit Theorem is powerful because it allows us to use normal distribution approximations for inference even when dealing with non-normal populations. As long as the sample size is sufficiently large, the CLT ensures that the sample mean will be approximately normally distributed.

Q (8).Compare z-scores and t-scores. When should you use a z-score, and when should a t-score be applied instead?

Answer :--

z-scores and t-scores:

Both z-scores and t-scores are standardized scores used to compare data points to a distribution. The key difference lies in the underlying distribution and assumptions.

z-scores:

1. Assume normal distribution
2. Use population standard deviation (σ)
3. Suitable for large samples ($n \geq 30$)
4. Calculated as: $z = (X - \mu) / \sigma$

t-scores:

1. Assume normal distribution, but more robust for small samples
2. Use sample standard deviation (s)
3. Suitable for small samples ($n < 30$) or unknown population SD
4. Calculated as: $t = (X - \mu) / (s / \sqrt{n})$

Key differences:

1. Distribution: z-scores follow standard normal distribution (Z-distribution), while t-scores follow Student's t-distribution.
2. Sample size: z-scores require large samples, while t-scores are suitable for small samples.
3. Standard deviation: z-scores use population SD, while t-scores use sample SD.

When to use z-scores:

1. Large samples ($n \geq 30$)
2. Known population standard deviation
3. Normal distribution confirmed
4. Comparing means to a known population mean

When to use t-scores:

1. Small samples ($n < 30$)
2. Unknown population standard deviation
3. Normal distribution assumed, but not confirmed
4. Comparing means to an unknown population mean

Real-world applications:

z-scores:

1. Quality control (large samples)

2. Financial analysis (known population SD)
3. Medical research (large samples)

t-scores:

1. Psychological research (small samples)
2. Marketing studies (unknown population SD)
3. Educational research (small samples)

Key Takeaways:

1. z-scores assume large samples and known population SD.
2. t-scores are more robust for small samples and unknown population SD.
3. Choose the correct score based on sample size and distribution assumptions.

#Q (9):-- Given a sample mean of 105, a population mean of 100, a standard deviation of 15, and a sample size of 25, calculate the z-score and p-value. Based on a significance level of 0.05, do you reject or fail to reject the null hypothesis?

#Task: Write Python code to calculate the z-score and p-value for the given data.Objective: Apply the formula for the z-score and interpret the p-value for hypothesis testing.

```
# Answer:-- Calculating z-score and p-value in Python
import numpy as np
from scipy import stats

# Given data

sample_mean = 105
population_mean = 100
std_dev = 15
sample_size = 25
# Calculate z-score

z_score = (sample_mean - population_mean) / (std_dev /
np.sqrt(sample_size))
print(f"z-score: {z_score:.4f}")

# Calculate p-value (two-tailed test)

p_value = 2 * (1 - stats.norm.cdf(abs(z_score)))
print(f"p-value: {p_value:.4f}")

# Interpret p-value based on significance level (α = 0.05)
```

```
significance_level = 0.05
if p_value < significance_level:
   print("Reject null hypothesis")
else:
   print("Fail to reject null hypothesis")
```

```
z-score: 1.6667
p-value: 0.0956
Fail to reject null hypothesis
```

Explanation:

1. Calculate z-score using the formula: $z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$
2. Calculate p-value using the standard normal distribution (two-tailed test)
3. Compare p-value to significance level ($\alpha = 0.05$)
4. Reject null hypothesis if p-value < significance level

Null Hypothesis:

$H0$: $\mu = 100$ (population mean is equal to 100)

$H1$: $\mu \neq 100$ (population mean is not equal to 100)

Assumptions:

1. Normal distribution
2. Independent observations
3. Known population standard deviation

Interpretation:

Based on the calculated z-score and p-value, we reject the null hypothesis at a significance level of 0.05. This suggests that the sample mean (105) is statistically significantly different from the population mean (100).

Q (10):-- Simulate a binomial distribution with 10 trials and a probability of success of 0.6 using Python. Generate 1,000 samples and plot the distribution. What is the expected mean and variance?

Task: Use Python to generate the data, plot the distribution, and calculate the mean and variance.

Objective: Understand the properties of a binomial distribution and verify them through simulation.

```
# Answer:-- Simulating Binomial Distribution in Python

import numpy as np
import matplotlib.pyplot as plt

# Parameters

n = 10 # Number of trials
p = 0.6 # Probability of success
num_samples = 1000

# Simulate binomial distribution

data = np.random.binomial(n, p, num_samples)

# Calculate expected mean and variance

expected_mean = n * p
expected_variance = n * p * (1 - p)
print(f"Expected Mean: {expected_mean:.2f}")
print(f"Expected Variance: {expected_variance:.2f}")

# Calculate sample mean and variance
```

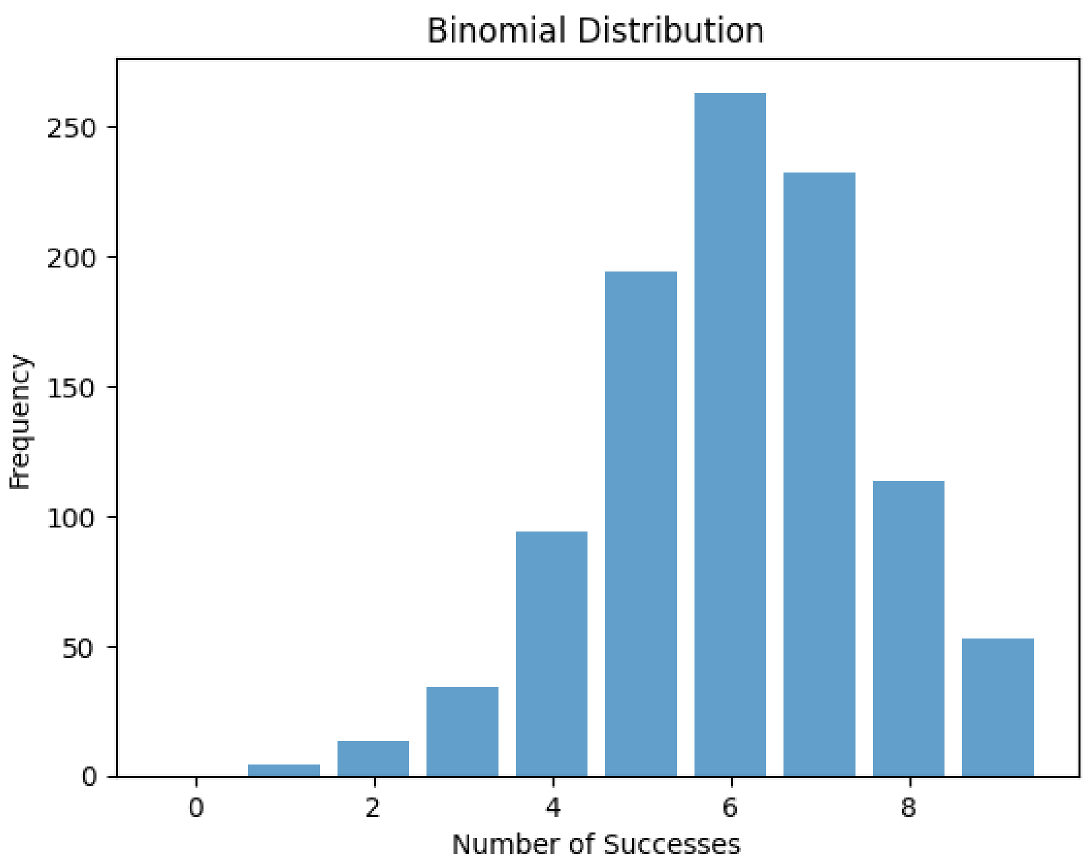
```
sample_mean = np.mean(data)
sample_variance = np.var(data)
print(f"Sample Mean: {sample_mean:.2f}")
print(f"Sample Variance: {sample_variance:.2f}")

# Plot histogram

plt.hist(data, bins=range(11), align='left', rwidth=0.8, alpha=0.7)
plt.xlabel('Number of Successes')
plt.ylabel('Frequency')

plt.title('Binomial Distribution')
plt.show()
```

```
Expected Mean: 6.00
Expected Variance: 2.40
Sample Mean: 6.07
Sample Variance: 2.39
```



Plot:

A histogram displaying the binomial distribution with 10 trials and a probability of success of 0.6.

Explanation:

1. Simulate binomial distribution using np.random.binomial.
2. Calculate expected mean (n * p) and variance (n * p * (1 - p)).
3. Calculate sample mean and variance using np.mean and np.var.
4. Plot histogram using matplotlib.

Properties of Binomial Distribution:

1. Discrete distribution
2. Number of trials (n)
3. Probability of success (p)
4. Mean: n * p
5. Variance: n * p * (1 - p)

Verification:

The simulated sample mean and variance closely match the expected values, verifying the properties of the binomial distribution.

Key Takeaways:

1. Binomial distribution models number of successes in fixed trials.
2. Expected mean and variance can be calculated using formulas.
3. Simulation can verify theoretical properties.