Statistics Advanced – 2 | ASSIGNMENT

Question 1: What is hypothesis testing in statistics?

Answer:

Hypothesis testing is a statistical method used to make inferences about population parameters based on sample data. It evaluates whether observed results are due to random chance or represent true population effects.

Question 2: What is the null hypothesis, and how does it differ from the alternative hypothesis?

Answer:

The null hypothesis (H_0) assumes no effect or difference, while the alternative hypothesis (H_1) states that there is an effect or difference. Example: H_0: $\mu = 50$, H_1: $\mu \neq 50$.

Question 3: Explain the significance level in hypothesis testing and its role in deciding the outcome of a test.

Answer:

The significance level (α) represents the threshold probability for rejecting H_0. It defines how much risk of a Type I error (false positive) is acceptable, typically 0.05 or 0.01.

Question 4: What are Type I and Type II errors? Give examples of each.

Answer:

Type I error: Rejecting a true H_0 (false positive). Example: Declaring a medicine effective when it isn't. Type II error: Failing to reject a false H_0 (false negative). Example: Missing the effect of a real medicine.

Question 5: What is the difference between a Z-test and a T-test? Explain when to use each.

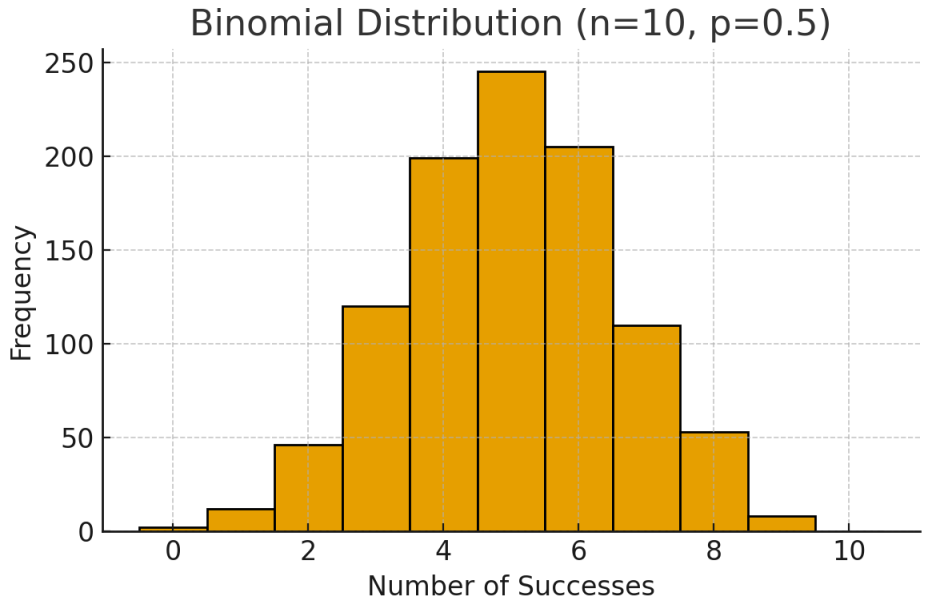
Answer:

A Z-test is used when the population standard deviation is known or $n \geq 30$. A T-test is used when the population standard deviation is unknown and $n < 30$. Both compare sample and population means.

Question 6: Write a Python program to generate a binomial distribution with n=10 and p=0.5, then plot its histogram. (Include your Python code and output in the code box below.)

Answer:

Generated 1000 random values from Binomial(n=10, p=0.5).
Mean = 4.9810, Std Dev = 1.6085.



Question 7: Implement hypothesis testing using Z-statistics for a sample dataset in Python. Show the Python code and interpret the results.

Answer:

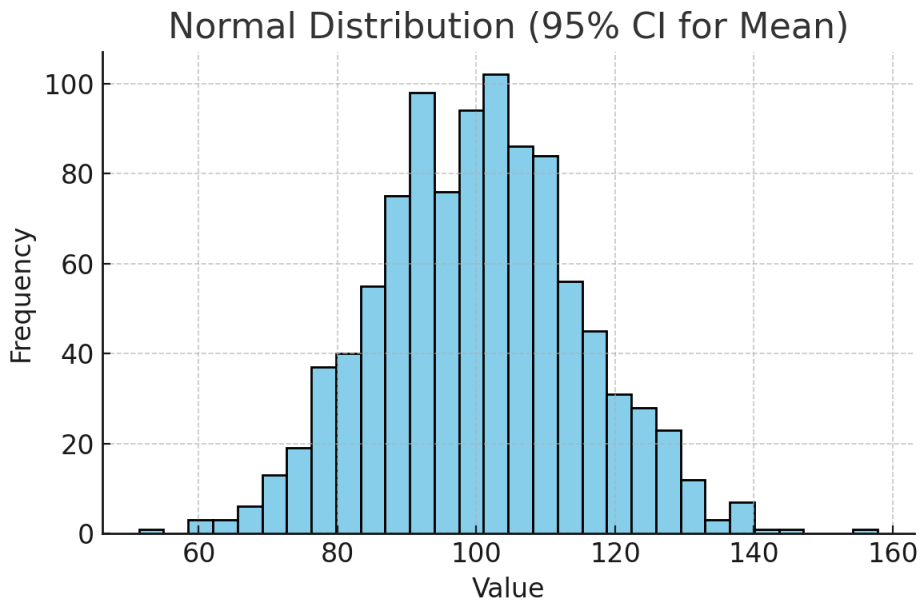
Z = 0.9940, p-value = 0.3202.

Conclusion: Fail to reject the null hypothesis (no significant difference).

Question 8: Write a Python script to simulate data from a normal distribution and calculate the 95% confidence interval for its mean. Plot the data using Matplotlib.

Answer:

Mean = 100.2900, Std Dev = 14.6882, 95% Confidence Interval = (99.37850677904946, 101.2014548956203).



Question 9: Write a Python function to calculate the Z-scores from a dataset and visualize the standardized data using a histogram. Explain what the Z-scores represent in terms of standard deviations from the mean.

Answer:

Z-scores indicate how many standard deviations each value is from the mean (mean = 0, std = 1).