Statistics Advanced - 2| Assignment

Question 1: What is hypothesis testing in statistics?
Answer:

Hypothesis testing in statistics is a method used to make decisions or draw conclusions about a population based on sample data. It involves testing an assumption (the null hypothesis, H_0) against an alternative hypothesis (H_1) using probability and sample evidence, and then deciding whether to reject H_0 or not.

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

The null hypothesis (H_0) is the assumption that there is *no effect, no difference, or no relationship* in the population — it represents the status quo.

The alternative hypothesis (H_1 or Ha) is the statement that *there is an effect, a difference, or a relationship* — what the researcher wants to prove.

Difference:

- **H_0 assumes nothing new is happening.**
- **H_1 challenges H_0 and suggests something new or significant is present.**

Example:

- **H_0: The new drug has no effect on blood pressure.**
- **H_1: The new drug does affect blood pressure.**

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

Answer:

The significance level (α) in hypothesis testing is the threshold probability used to decide whether to reject the null hypothesis.

- **It represents the risk of making a Type I error (rejecting H_0 when it is actually true).**
- **Common values are 0.05 (5%) or 0.01 (1%).**

Role in outcome:

- **If the p-value ≤ α, we reject H_0 (evidence supports H_1).**
- **If the p-value > α, we fail to reject H_0 (not enough evidence against H_0).**

Example: At α = 0.05, we are willing to take a 5% chance of wrongly rejecting the null hypothesis.

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

Answer:

In hypothesis testing, errors occur when we make wrong decisions about the null hypothesis:

1. Type I Error (False Positive):

- **Rejecting the null hypothesis (H_0) when it is actually true.**

- **Probability of this error = α (significance level).**
  Example: Concluding a new medicine works when in reality it has no effect.

2. Type II Error (False Negative):

- **Failing to reject the null hypothesis when the alternative hypothesis (H_1) is actually true.**

- **Probability of this error = β.**
  Example: Concluding a new medicine has no effect when in reality it does work.

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

Answer:

A Z-test and a T-test are both statistical tests used to compare sample data with population parameters or between groups, but they differ mainly in sample size and knowledge of population variance:

Z-test:

- **Used when the population variance (σ^2) is known and/or the sample size is large ($n > 30$).**

- **Assumes data follows a normal distribution.**
  Example: Testing whether the average height of 1,000 students differs from a known population mean.

T-test:

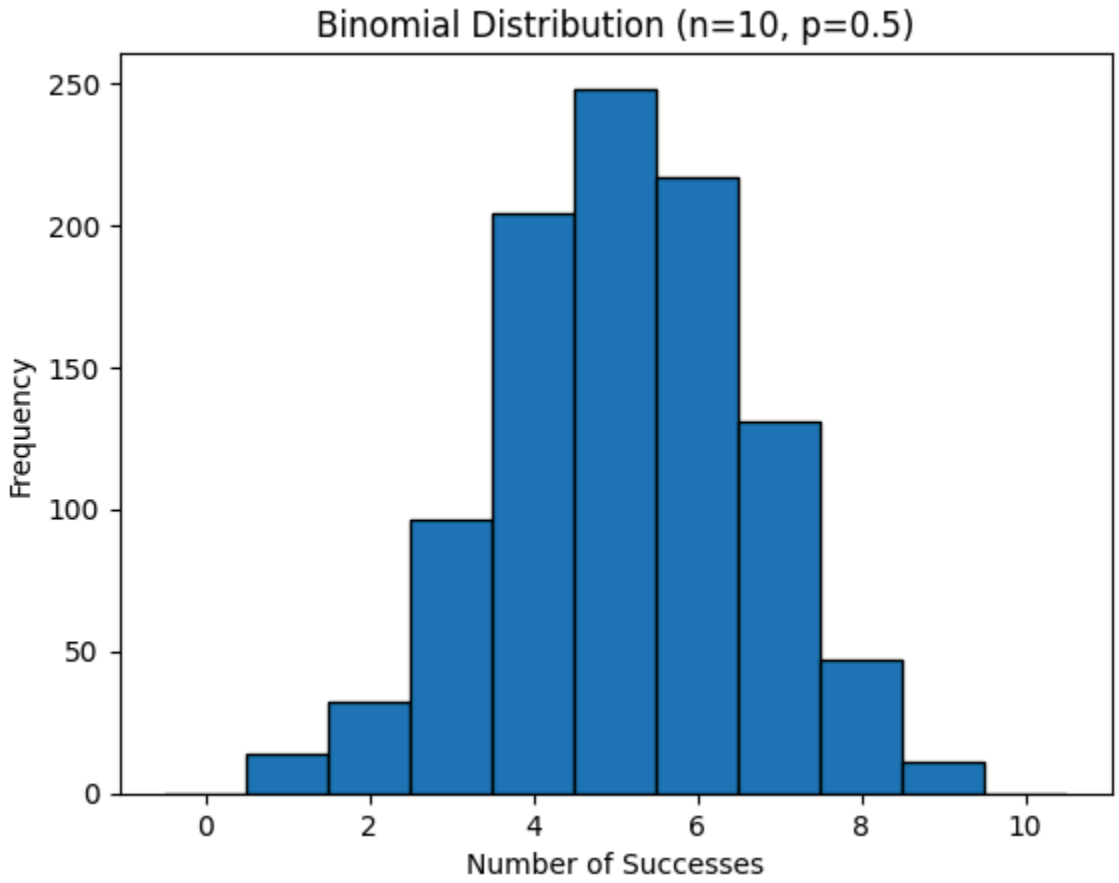
- **Used when the population variance is unknown and the sample size is small ($n \leq 30$).**

- **Relies on the t-distribution, which accounts for more variability in small samples.**
  Example: Testing whether the mean exam score of a class of 20 students differs from the expected averal

Question 6: Write a Python program to generate a binomial distribution with n=10 and p=0.5, then plot its histogram.

Answer: code

Output



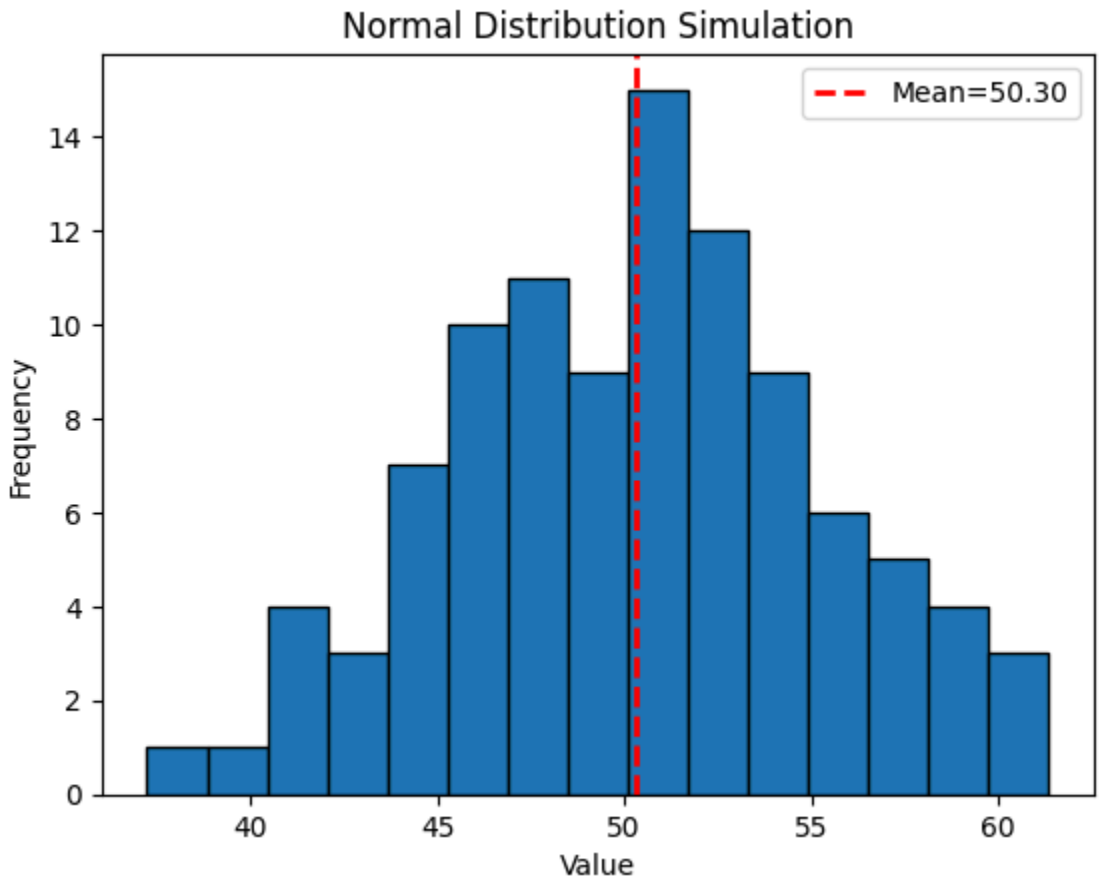
Question 7: Implement hypothesis testing using Z-statistics for a sample dataset in Python. Show the Python code and interpret the results.
sample_data = [49.1, 50.2, 51.0, 48.7, 50.5, 49.8, 50.3, 50.7, 50.2, 49.6, 50.1, 49.9, 50.8, 50.4, 48.9, 50.6, 50.0, 49.7, 50.2, 49.5, 50.1, 50.3, 50.4, 50.5, 50.0, 50.7, 49.3, 49.8, 50.2, 50.9, 50.3, 50.4, 50.0, 49.7, 50.5, 49.9]
Answer: code

Output:

```
Mean=50.09, Z=0.99, p=0.320
```

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: code

Output:



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: code

Output