Question1:

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

Ans :

Z-statistics (Z-score) measures that indicate how many standard deviation a datapoint away from the mean

Zscore = (datapoint – mean) / population standard deviation

Relationship between standard deviation and normal distribution is the

1standard deviation is the 68% of data include in ND

2 standard deviation is the 95 % of data include in ND

3 standard deviation is the 99.7% of data include in ND

Using the Z-Statistic in Hypothesis Testing

In hypothesis testing, the z-statistic helps determine whether to reject the null hypothesis. Here's how it's used:

Formulate Hypotheses: Set up the null hypothesis (H0H_0H0) and the alternative hypothesis (H1H_1H1).

H0H_0H0: There is no effect or difference (eg population mean equals a specific value).

H1H_1H1: There is an effect or difference (eg population mean is different).

Question2 :

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)?

Ans:

The p-value measures the probability of observing results as extreme as the current data if the null hypothesis is true. In hypothesis testing, a smaller p-value indicates stronger evidence against the null hypothesis. If the p-value is very small (e.g., 0.01), it implies strong evidence to reject the null hypothesis in favor of the alternative hypothesis, suggesting that the observed effect or difference is statistically significant.

Question3:

Compare and contrast the binomial and Bernoulli distributions.

Ans:

Bernoulli theorem states that in an trail there should only two outcomes

For example tossing a coin

Mainly it can true or false or pass or fail

It can be coded has 1 or 0

Mean = p

Variance = p(1-p)

Pmf of bernouli = p^k (1-p)^1-k

Binomial theorm is the n number of bernouli theorem

For example tossing coin for 10 times

Pmf of binomial = nck *p^k (1-p)^n-k

Mean = np

Variance = npq

Question 4:

Under what conditions is the binomial distribution used, and how does it relate to the Bernoulli distribution?

Ans:

The binomial distribution is used when there are multiple independent trials, each with two possible outcomes, and a constant probability of success. It counts the number of successes in these trials. The Bernoulli distribution is the foundation of the binomial distribution, representing a single trial with a binary outcome. When multiple independent Bernoulli trials are summed, the result follows a binomial distribution.

Question5:

What are the key properties of the Poisson distribution, and when is it appropriate to use this distribution?

Ans:

The poisson distribution is a discrete probability distribution that describes the number of events occurred within in a fixed interval of time or space given a known average rate of occurance

For example :

Number of people entering temple/hospital/play garden

The accident occurred in every hour in country

Question6:

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

Ans:

Probability distribution

Probability distribution irrespective of outcome in nature , draw the outcome in the form of the distribution(data arrangement)

Probability density function

If the outcome is in continuous then probability density function is used

For example:

Calculate the probability for the student height below 140

Probability density functions

- Normal distribution
- Standard normal distribution
- Log normal distribution
- Chi – square distribution
- F – distribution
- Continuous uniform distribution

Probability Mass function

If the outcomes after the experiment is discrete then used Probability Mass Function

Example: tossing coin , throwing dice

- Bernoulli
- Binomial
- Poison distribution
- Discrete uniform distribution
- Geometric distribution

Question7:

Explain the Central Limit Theorem (CLT) with example.

Ans

The CLT states that if you have a population with a mean and standard deviation and we can take sufficient large number of random samples with replacement then the sample mean distribution will be approximately Normal Distribution

Example :

You have a population with mean = 100 & standard deviation 20, if you have sample size = 50 .what is the probability that sample mean will be less the 105 ?

Zscore = (sample mean – population mean)/stdev/square root of size

Step 1: Calculate the Standard Error (SE) of the Mean

SE = 2.83

Step 2: Calculate the Z-Score

To find the probability that the sample mean is less than 105, we need to calculate the **z-score**, which measures how many standard errors the sample mean is from the population mean.

Step 3: Find the Probability Using the Z-Score

Using standard normal distribution tables or a calculator, we can find the probability corresponding to a z-score of 1.77. The probability that a z-score is less than 1.77 is approximately **0.9616**.

Question8:

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

ANS:

Z-scores are appropriate for larger samples or when the population standard deviation is known, while **t-scores** are suitable for smaller samples or when the population standard deviation is unknown.

T-scores account for more variability in small sample sizes by using the t-distribution, which has thicker tails to reflect the greater uncertainty.

Use a Z-Score When:

- The sample size is large (usually n≥30n \geq 30n≥30).
- The population standard deviation is known.
- The population distribution is normal, or the sample size is large enough for the Central Limit Theorem to apply.

Use a T-Score When:

- The sample size is small (usually n<30n < 30n<30).
- The population standard deviation is unknown. In this case, the sample standard deviation (s) is used as an estimate.
- The population is normally distributed, or the data approximately follows a normal distribution.

Question 9 (answer on Jupyter notebook)

Question9: 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.

Question 10 (answer on Jupyter notebook)

Question10: 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.