

6/8/2025 Part-A Theoretical Function (Short Notes and explanation)

Q1 - What is inferential statistics?

Ans Inferential Statistics is the branch of statistics that allows you to make predictions or conclusions about a large group called population.

→ By analyzing a smaller sample taken from the large group.

Q2 - What is hypothesis testing and its components?

Ans Hypothesis testing is a statistical method used to make decision about population parameter based on a sample.

Components of hypothesis testing:

- 1) Null hypothesis (H_0) = It is initial claim or default.
- 2) Alternative hypothesis (H_1) = It is what you want to prove or test.

Q3 - Explain Confidence interval and Critical value?

Ans Confidence Interval is a range of values used to estimate a population parameter (mean, or

proportion) based on a sample.

- Critical-value : A Critical value is the cut-off point on a probability distribution used to determine the boundary of the confidence interval to make decisions.

Q4 Define -P-value ?

Ans P-value (Probability value) is a key concepts in hypothesis testing.

- It tells us how likely it is to observe the given data assuming the Null hypothesis is True.

$P \leq 0.05$ (Reject the Null hypothesis)

$P > 0.05$ (We fail to reject null hypothesis)

Q5 Differentiate Type 1 error and Type 2 ?

Ans 1) Type 1 error (False positive) :

- Rejecting a True Null hypothesis
- which means in reality H_0 was true but it was rejected by a predicted model.

2) Type 2 error (False negative) :

- Failing to reject a false null H_0
- which means in reality the null hypothesis was rejected but the

model predicted it True.

→ Example (Type 1): A person has COVID-19 positive but after test it was negative

→ Example (Type 2): A person was ^{not} drunk but after test he was found out drunk.

Q6 Brief descriptions of Z-Test, t-test, Chi-square test and ANOVA Test?

Ans 1) A Z-test is a statistical test used to determine whether there is a significant difference between a sample mean and known population mean.

Criteria: σ = Should be known
Sample Size ≥ 30

One sample Z-test: Two sample Z-test:

$$Z = \frac{\bar{X} - \mu}{\sigma \sqrt{n}}$$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

when $\mu_1 = \mu_2$

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- 2) T-test is a statistical test used to compare means of one or two groups when the population STD is unknown.

Criteria: $\sigma = \text{unknown}$
Sample size < 30 .

one sample T-test: Two sample T-test

$$T = \frac{\bar{x} - \mu}{s \sqrt{n}}$$

(Pooled)

$$T = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$\rightarrow \frac{\sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}}{n_1 + n_2 - 2}$$

2) (Welch's)

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- 3) A Chi-Square test is used to examine the relationship between two categorical variables. or to check

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

4) ANOVA (Analysis of Variance): ANOVA is a statistical method used to compare the means of three or more groups to determine if at least one group mean is significantly different.

$$F = \frac{\text{Variance between groups}}{\text{Variance within groups}}$$

Q7 What is Co-Variance?

Ans Co-Variance measures how two variables change together.

→ If both increase together → Positive Co-Variance

→ If one increase → Negative Co-Variance

Q8 What is Co-Relation?

Ans
$$\text{Cov}(X, Y) = \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y})$$

Q8 What is Co-Relation?

Ans Correlation measures the strength and direction of the relationship between two variables.

$$\text{Pearson correlation} = r = \frac{\text{Cov}(X, Y)}{\sigma_X \times \sigma_Y}$$