

Page No.

Date.

1 What is inferential statistics

→ It is the branch of statistics that allows us to make prediction or draw conclusions about a larger group called a population by analyzing a smaller sample of data.

2

What is hypothesis testing and its components?

→ It is a statistical method used to make decision about population parameters based on sample data.

→ Components:

→ Null hypothesis (H_0)

→ Alternate hypothesis (H_1)

Ex confidence interval 4 critical value

→ A confidence interval gives a range of values where we believe true population parameter lies, with a given level of confidence

→ Critical value:

→ A critical value is a specific value from a probability distribution like standard normal z-distribution or t-distribution) that marks the boundary for a given confidence level.

ii Define P-value

- P-value is a key concept in hypothesis testing. It tells us how likely it is to observe the given data assuming null hypothesis is true.
- If $P\text{-value} < 0.05$ - Reject null hypothesis
- If $P\text{-value} > 0.05$ - Accept null hypothesis

Diff Type-1 & Type-2 error

Type-1	Type-2
→ Rejecting null hypothesis that is actually true	→ Failing to reject a null hypothesis that is actually false
→ Eg: Convicting an innocent person in court	→ Eg: Letting a guilty person go free

- | | |
|---|--|
| → Denoted by alpha (α)
the significance level | → Denoted by beta (β) |
| → False Positive | → False negative |
| → A false claim that something exists or an effect is present | → A missed opportunity to detect a real effect |

6 Description:-

- **Z-test** = A Z-test is a statistical method used to determine whether there is a significant difference between sample data and a population or between two samples.

Formula:-
$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}, \quad \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- **T-test** = A T-test is used when we want to determine whether the mean of a sample is significantly different from a known or hypothesized population mean

Formula:-
$$T = \frac{\bar{x} - \mu}{S/\sqrt{n}}, \quad \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

→ chi-square test :- The chi-square test is a statistical test used to determine if there's a significant association between two categorical variables, or whether dataset fits a certain distribution.

Types:-

- 1) Chi-square goodness of fit
- 2) Chi-square test of independence

7 What is covariance

- It measures how two variables changes together
- Positive cov :- If both increases together
- Negative cov :- If one increases & other decreases
- No pattern :- Covariance is zero or near-zero

Formula :- $\text{Cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$

8 What is correlation:

- It measures the strength and direction of the relationship between two variables.
- Range is -1 to +1
- +1 = Perfect positive relationship
- 0 = No relationship
- -1 = Perfect negative relationship

Formula: $r = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$