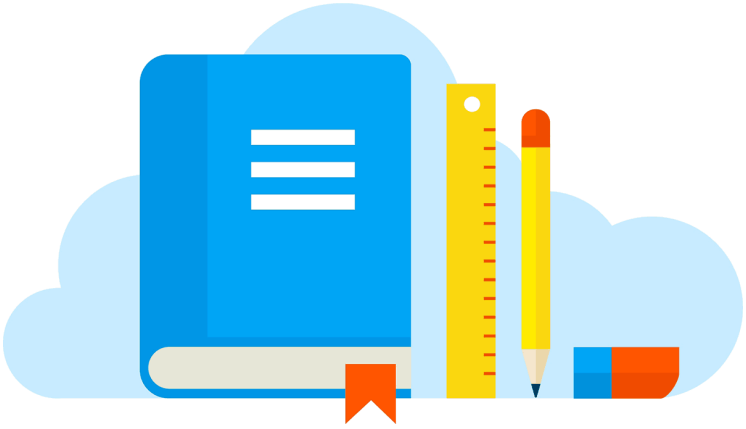

Data Science Notes by Sarowar Ahmed

📊 Chapter: inferential statistics

📊 Topic: Chi-Square Test

📢 What is a Chi-Square Test?

▪ A chi-square test is a statistical method used to determine if there's a significant association between two categorical variables. It's essential in fields like marketing, medicine, and social sciences, helping to make decisions based on observed data versus expected outcomes if there were no relationship.

🍒 ♒ Types of Chi-Square Tests:

▪ Chi-Square Test of Independence: Determines if two categorical variables are independent.
▪ Chi-Square Goodness of Fit Test: Checks how well observed data fit a model or distribution.

➕ Formula:

The formula for the Chi-Square Test of Independence is:
$\chi 2=\Sigma(Oi-Ei)^2/Ei$

Where:
▪ Oi is the observed frequency.
▪ Ei is the expected frequency under the null hypothesis.

🔰 Example:
▪ Imagine a marketer wants to see if there is an association between age group and preference for a new beverage type. They survey 100 people divided into two age groups and two beverage preferences (New Drink vs. Traditional).

Observed Data:
▪ Under 30: 30 prefer New Drink, 20 prefer Traditional
▪ 30 and Over: 10 prefer New Drink, 40 prefer Traditional

Expected Data (if no preference association):

▪ Under 30: 25 for New Drink, 25 for Traditional

▪ 30 and Over: 25 for New Drink, 25 for Traditional

Using the formula:

$\chi 2=(30-25)^2/25+(20-25)^2/25+(10-25)^2/25+(40-25)^2/25=14$

This chi-square statistic is then compared to a critical value in chi-square tables based on degrees of freedom and the significance level.

💡 Why is this useful?

▪ Chi-square tests help determine if differences in survey responses, like our beverage preference, are likely due to actual preferences or just random chance. This informs better decisions in policy-making, business strategies, and beyond.

Got any questions on the Chi-square tests? Feel free to ask me via Linkedin! Let's keep learning together.

My LinkedIn Date: 07/05/2024