6th DAY Live SESSION

1. CHI SQUARE ✓
2. Covariance ✓
3. Pearson Correlation Coefficient ✓
4. Spearman Rank Correlation ✓
5. Practical Implementation

   Z-test, t-test, chi square test
6. F Test (ANOVA)

CHI SQUARE TEST

1. Chi Square Test claims about population proportions

   It is a non parametric test that is performed on categorical (nominal or ordinal) data.

Q) In the 2000 Indian Census, the age of the individual in a small town were found to be the following:

less than 18	18-35	>35
20%	30%	50%

In 2010, age of n=500 individuals were sampled. Below are the results

<18	18-35	>35
121	288	91

Using $\alpha = 0.05$, would you conclude the population distribution of ages has changed in the last 10 years?

Ans)

<18	18-35	>35
20%	30%	50%

{Population} 2000

Expected

<18	18-35	>35
121	288	91
100	500 × 0.3	500 × 0.5
	150	250

n=500

Observed

Expected

<18	18-35	>35
121	288	91
100	150	250

Observation

Expected

{Chi Square table}

① H_0 = The data meets the distribution 2000 census

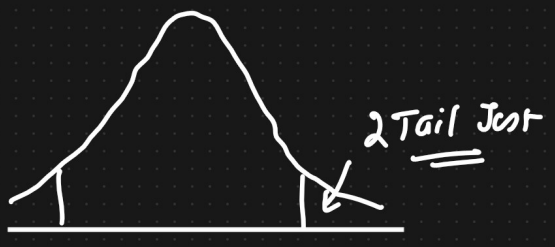
H_1 = The data does not meet " " "

② $\alpha = 0.05$ (95% C.I)

df: 2, $\alpha = 0.05$

③ Degree of freedom = $n - 1 = 3 - 1 = 2$

④ Decision Boundary



If χ^2 is greater than 5.99 reject H_0

5) Calculate Test Statistics

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$= \frac{(121-100)^2}{100} + \frac{(288-150)^2}{150} + \frac{(91-250)^2}{250}$$

$$= 232.94$$

$\chi^2 = 232.494 > 5.99$ {Reject the Null Hypothesis}

$\alpha = 0.05$ ↙ 0.01, 0.10 } Domain.

0.11 > 0.05 ⇓ Accept the Null

0.11 ✗ 0.0 ⇓ Reject the Hypothesis

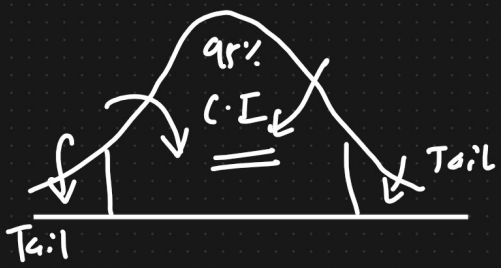
0.002 < 0.05 {Reject the null hypothesis}

I Was Correct Here

{ p-value < Significance value ⇓ Reject the Null Hypothesis. }

OR

{Accept the Null Hypothesis}



p 0.11 > 0.05 ⇓ Accept

p = 0.002 < 0.05 ⇓ Reject the null hypothesis

② Covariance

X	Y
Weight	Height
50	160
60	170
70	180
75	181

{ X↑ Y↑ ; X↓ Y↓ }

No. of hour study	play
2	6
3	4
4	3

X↑ Y↓ ; X↓ Y↑

Quantity relationship between X & Y

Covariance

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

+ve ⇒ positive Correlation

X↑ Y↑ ; X↓ Y↓

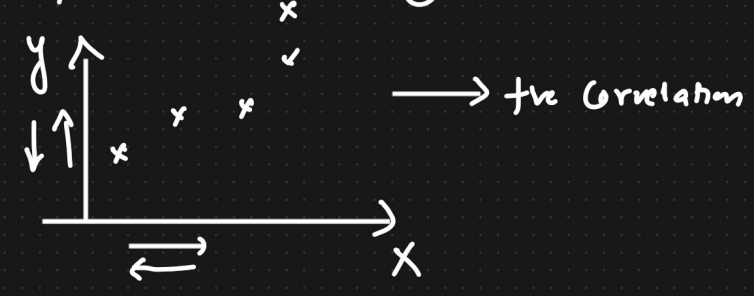
-ve ⇒ negative Correlation

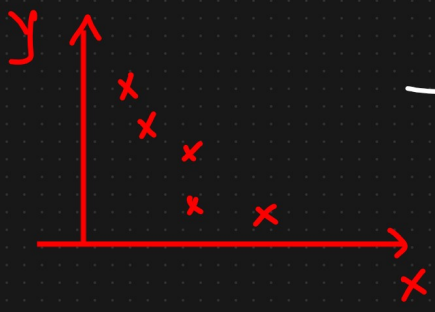
X↓ Y↑ ; X↑ Y↓

= +ve or -ve

X ⇕ & Y

O

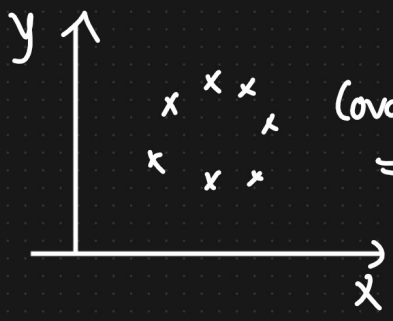




→ -ve Correlation

⇓

-ve Covariance



Covariance = 0

Disadvantage of Covariance

① Positive or Negative ✓

+100

-200

-2000

+1000 {Direction}

② Pearson Correlation Coefficient

(-1 to 1)

The more towards +1 more positively correlation

The more towards -1 more negative correlated

$$\rho_{(x,y)} = \frac{Cov(x,y)}{\sigma_x \sigma_y} \quad \{-1 \text{ to } 1\}$$

$$Spear(x,y) = \frac{Cov(R(x), R(y))}{R_{\sigma_x} \times R_{\sigma_y}}$$

X Height	Y Weight	R(x)	R(y)
170	75	2	2
160	62	3	3
150	60	4	4
145	55	5	5
180	85	1	1

Non linear properties

Coding

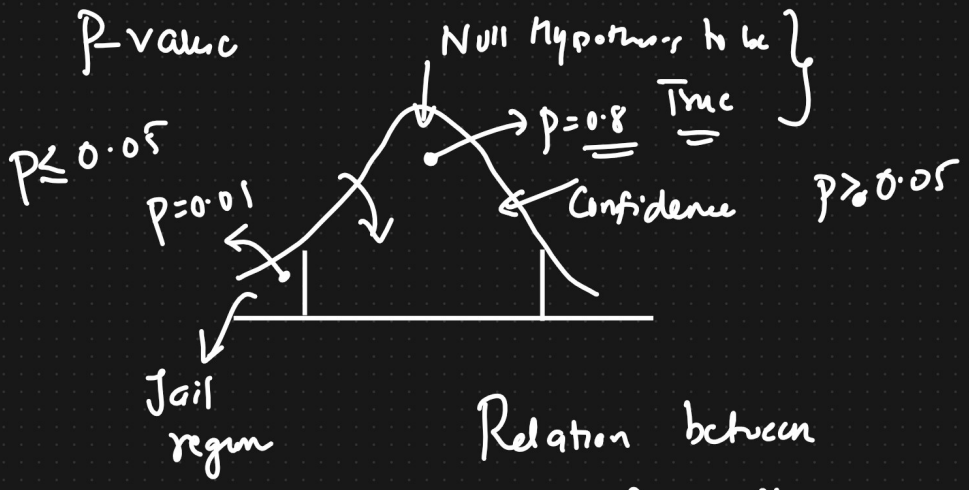
$P \leq 0.05 \rightarrow$ Reject the NULL Hypothesis

probability

5% probability the null Hypothesis is Correct

$\alpha = 0.05$

$P \geq 0.05 \rightarrow$ Accept the NULL Hypothesis



Relation between pvalue & significance

P-value ≤ Significance [C.I]

↳ Reject the Null Hypothesis

P≥α ↳ Accept the Null

① P value and Significance value