24-01-25 Day-6

Page No.
Date

Today's topics

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 proportion. It is a non parametric test that is performed on categorical (Nominal or Ordinal) data

Q. In the 2000 Indian census, the ages of the individual in a small town were found to be the following:

1. less than 4
2. 18-35
3. > 35

Ans:-

less than 4	18-35	>35
20%	30%	50%

In 2010, ages of n=500 individuals was sampled. Below are the results

<18	18-35	>35
121	288	91

Page No.
Date

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

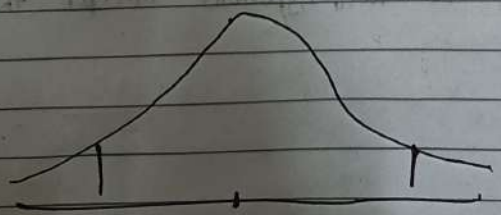
Answer:-

<18	18-35	>35	2000 Population
20 %	30 %	50 %	Expected

<18	18-35	>35	n=500
121	288	91	Observed
500 × 0.2 → 100	500 · 03 → 150	500 · 05 → 250	→ Expected.
100	150	250	

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

1. H_0 = The data meets the distribution 2000 census
   H_1 = The data doesn't meet the distribution 2000 census
2. α = 0.005 (95 % Confidet Interval)
3. Degree of Freedom $n-1 \Rightarrow 3-1 = 2$
4. Decision Boundary



2 tail test

If x^2 is great than 5.99 reject H_0

5. Calculate test statistics

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

{ f_o = observed
f_e = Expected }

$$\chi^2 = \left(\frac{121-100}{100}\right)^2 + \frac{(258-400)^2}{150} + \frac{(91-200)^2}{200}$$

$$\chi^2 = 4.41 + 126.96 + 101.124$$
$$\chi^2 = 232.494$$

$\chi^2 = 232.94 > 5.99$ ⇒ Reject the Null hypothesis

6. Conclusion

p-value < Significant value
⇓
In greater than Reject the Null hypothesis
(0.1)

In less than Accept the Null hypothesis

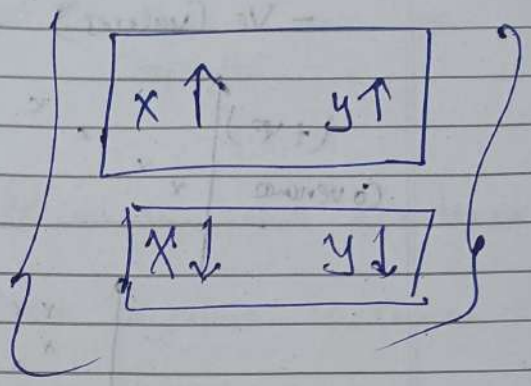
P 0.4 > 0.05
⇓
Reject the Null hypothesis

P = 0.0002 < 0.05
⇓
Accept the Null hypothesis

Page No.
Date

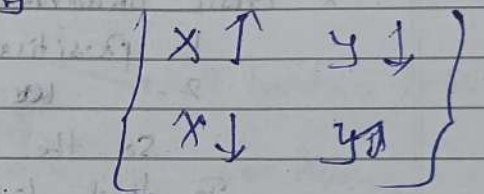
* Covarience :-

X	y
weight	height.
50	180
60	170
70	180
75	181



Es:-

No. of hrs Study	No of hrs play
2	6
3	4
4	4



Quantify relationship between x & y
For that time we use

Covarience

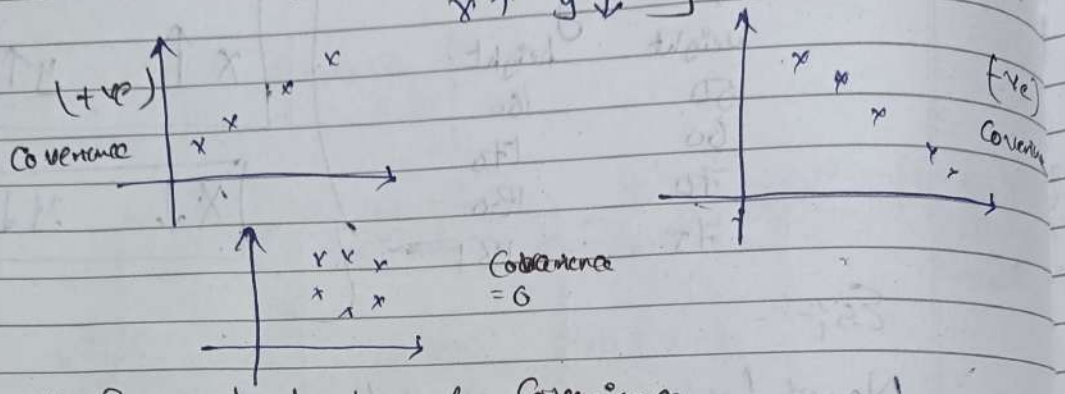
Formula $Cov(x,y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n}$

for Sample:

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

+ Ve (values) => $x\uparrow$ $y\uparrow$ } Positive Correlation
$x\downarrow$ $y\downarrow$

- Ve (values) => $x\downarrow$ $y\uparrow$ } Negative Correlation
$x\uparrow$ $y\downarrow$



* Basic disadvantage of Covariance
1. Positive or Negative
2. the value is unlimited (+ve) or (-ve)
So the we dont difficult to Solve.
In that time we use
* Pearson Correlation Coefficient
(-1 to 1) -> The more towards +1 more positively Correlation

-> The more towards -1 more Negative Correlative

Formula :-

$$\rho(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y} \Rightarrow [-1, +1]$$

Formula : Pearson

$$\rho Pearson(x,y) = \frac{Cov(x,y)}{R\sigma_x \times R\sigma_y} \quad \frac{Cov(R(x), R(y))}{R\sigma_x \times R\sigma_y}$$

X Height	Y weight	R(x)	R(y)
170	75	2	
160	62	3	2
150	60	4	3
150	61	5	4
145	55	4	5
180	85	1	1

Page No.
Date

Q. Why do use Person correlation coef ?

→ It captures "non-linear properties

Probably $P \leq 0.05$ → Reject the Null hypothesis

5% probably the null Hypothesis is correct

$\alpha = 0.05$

$P \geq 0.05$ → Accept the Null hypothesis