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)

# 1 CHI Square TEST

Chi Square Test claims about population proportions.

Why Chi Square Test

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

Question, In the 2000 Indian Census, the age of the individual in a small town were found to be the following.

Less than 18 = 20%

18-35 = 30%

> 35 = 50 %

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

< 18 = 121

18 – 35 = 288

> 35 = 91

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

Answer

|  |  |  |
| --- | --- | --- |
| < 18 | 18-35 | * 35 |
| 20% | 30% | 50% |

This is expected. From the population data 2000

|  |  |  |
| --- | --- | --- |
| < 18 | 18-35 | >35 |
| 121 | 288 | 91 |

Observed data. n= 500 (500 sample). Based on the population data. The expected data will be.

|  |  |  |  |
| --- | --- | --- | --- |
| 500 x 0.2 | 500 x0.3 | 500 x .5 | Observation |
| 100 | 150 | 250 | Expected |

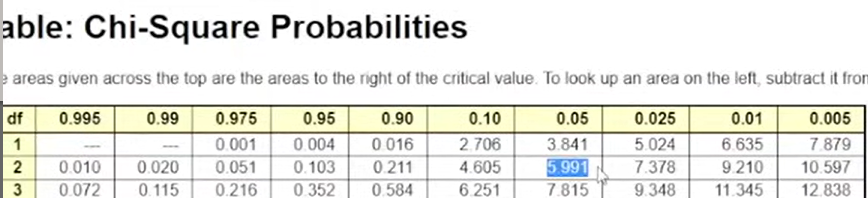
1 Ho = The data meets the distribution 2010 census.

Hi = The data does not meet distribution 2000 census.

2 ∞ = 0.05 (95% Confidence Interval).

3 Degree of freedom = n-1 = 3-1= 2 { df=2, ∞=0.05}

4 Decision Boundary {Check chi square table to fond decision boundary}.



Two tail tests

If X2 is greater than 5.99 reject Ho

2 Tail test

5 Calculate test statistics.

Fo – Observed

Fe - Expected

X2 (fo-fe)2/fe = (121-100)2/100 + (288-150)2/150 + (91-250)2/250

= 232.94

X2= 232.94 > 5.99 { Reject the Null Hypothesis }

Python test

Suppose the IQ in a certain population is normally distributed with a mean of µ = 100 and standard deviation of ճ =15.

A researcher wants to know if a new drug affects IQ levels, so he recruits 20 patients to try it and records their IQ levels.

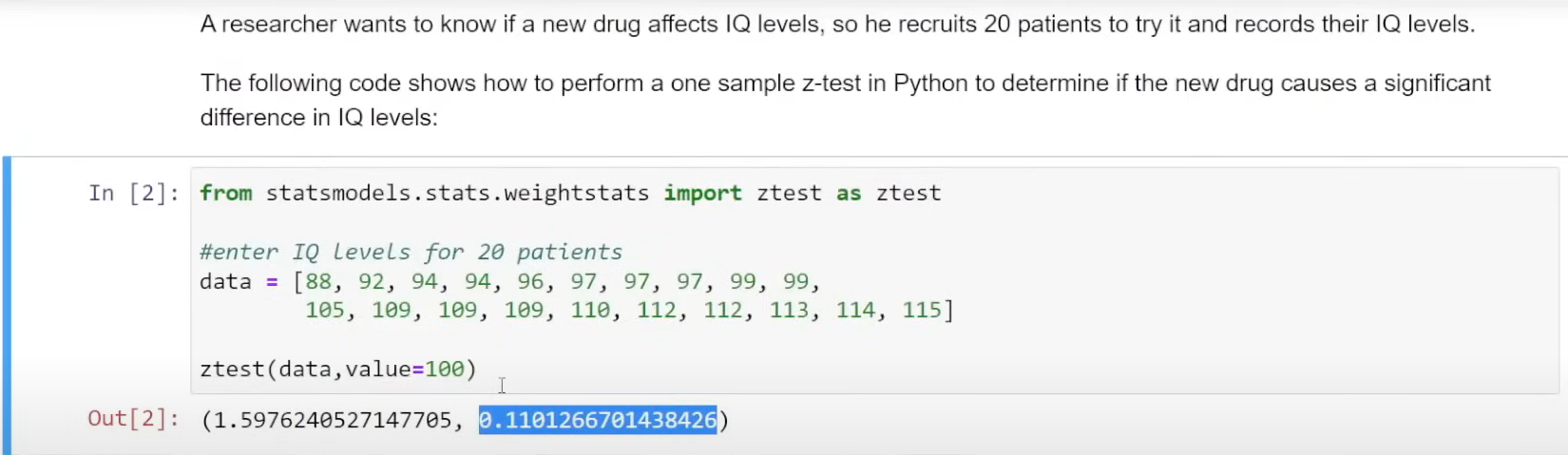
The following code shows how to perform a one sample z-test in python to determine if the new drug causes a significant difference in IQ levels.

from statsmodels.stats.weightstats import ztest as ztest

enter IQ Levels for 20 patients.

data = [88, 92, 94, 94, 96, 97, 97, 97, 99, 99, 105, 109, 109, 110, 112, 112, 113, 114, 115]

ztest(data, value=100)



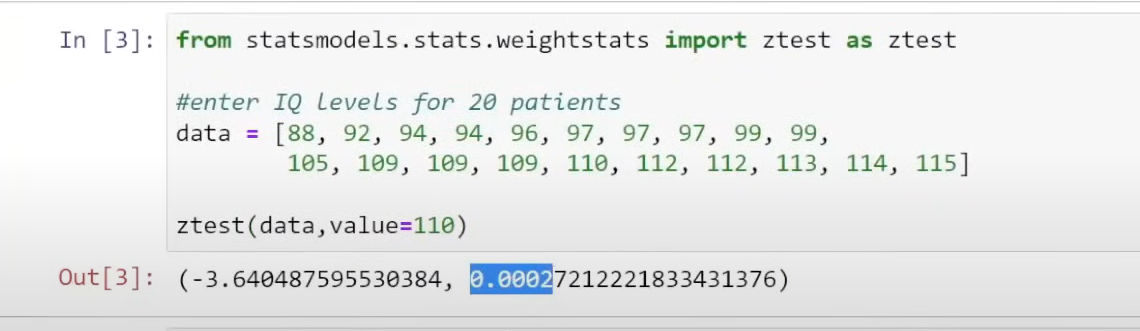
Z test value – 1.59……; P value =0.110….

∞ = 0.05

0.11 > 0.05 = Reject the null hypothesis.

Because it falls on

In another case



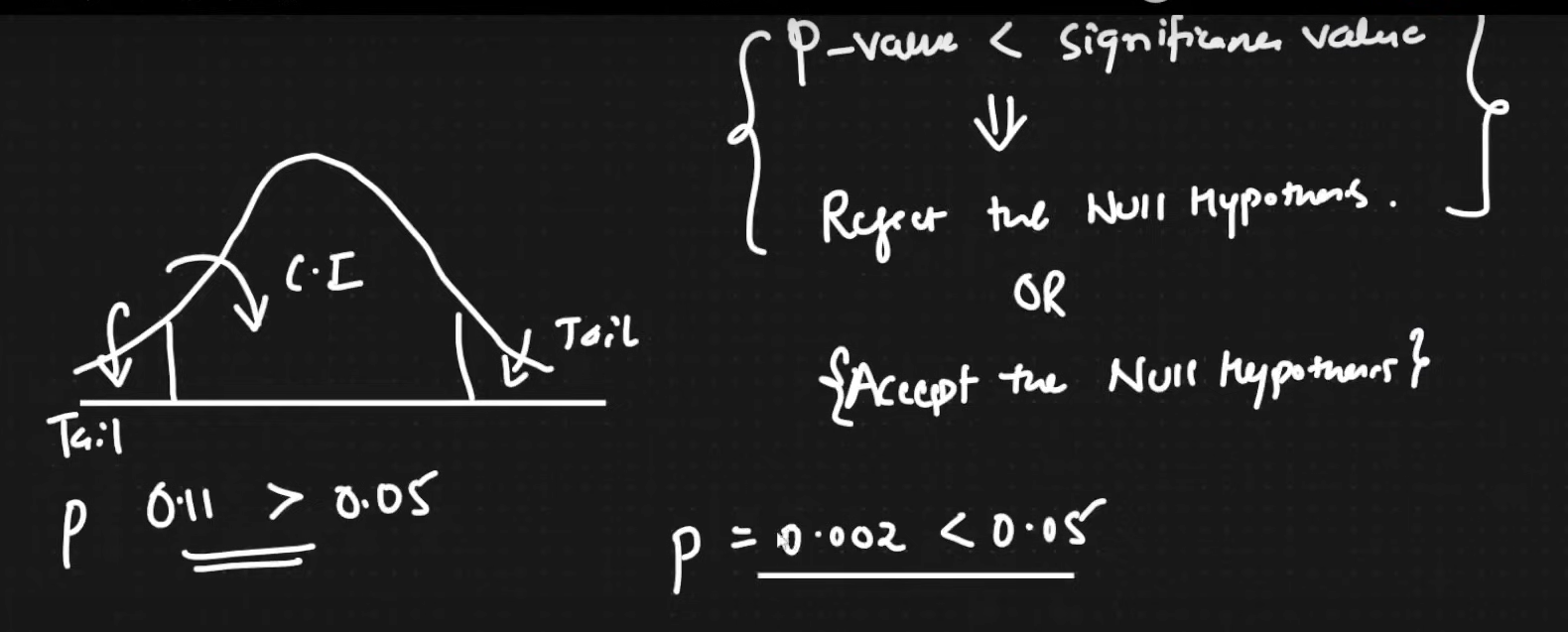
Ztest value – - 3.640……; P value =0.00027…

∞ = 0.05

0.0002 < 0.05 = {Accepted the null hypothesis}.

# 2 Covariance

P-value < significance value = Reject the Null hypothesis OR Accept the Null hypothesis.



Before medication

P value - 0.11 > 0.05 = Reject the null hypothesis.

Because it falls on

After medication P value – 0.002 < 0.05 Accepted because inside.

Weight and height Scenario.

X Weight – 50 60 70 75

Y Height – 160 170 180 181

Relationship

When,

X increasing Y also increasing

X decreasing Y also decreasing

No. of hours Study X 2 3 4

Play Y 6 4 3

Relationship

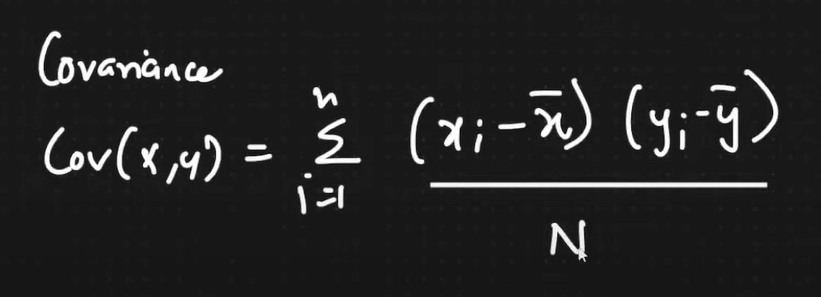
When.

X is increasing Y is decreasing

X is Decreasing Y is increasing

How can we quantify relationship between X & Y?

Covariance



Result of this formula is Positive Value or Negative value or 0

Positive Value indicate two things. Positive Correlation

1 X increasing Y also increasing.

X decreasing Y also decreasing.

Negative value indicates. Negative Correlation

1 X is increasing Y is decreasing.

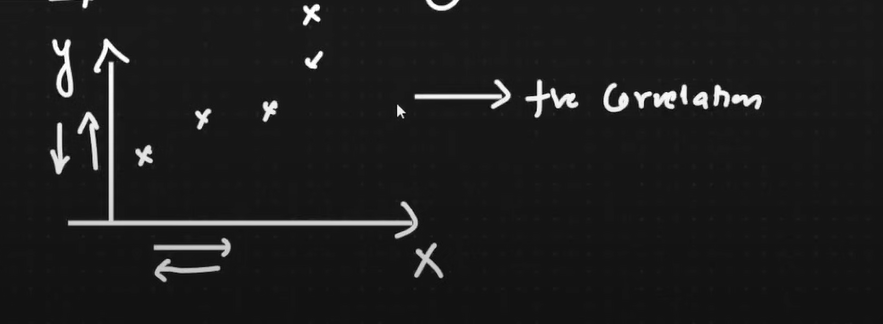
X is Decreasing Y is increasing.

When value is 0

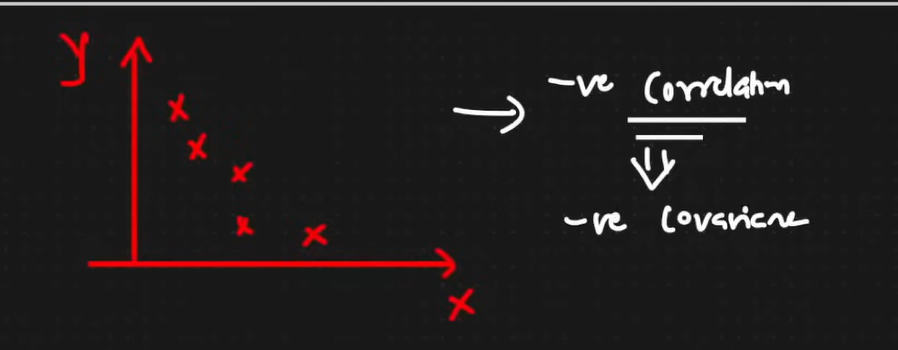
X increasing Y no Change

X is decreasing Y no Change

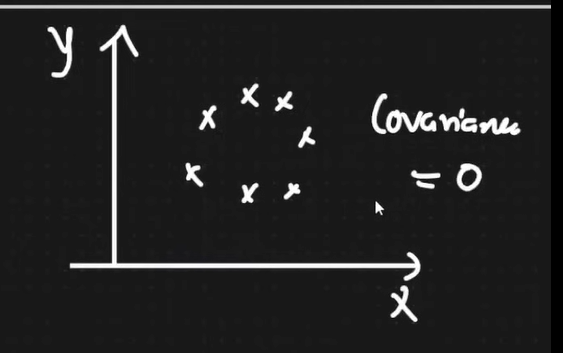
Example of Positive Corelation



Example of Negative Corelation ( Negative co variance).



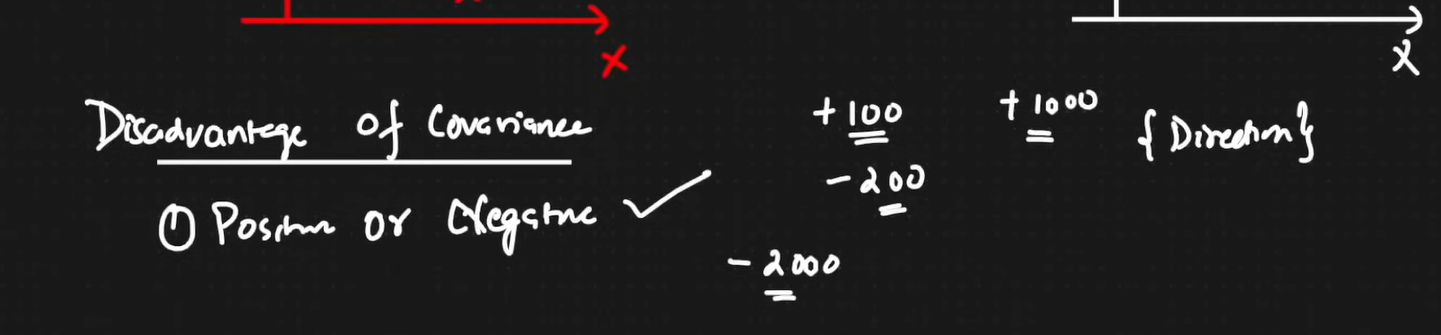
O Covariance



Disadvantage of Covariance

1 Positive of Negative

“With respect to magnitude there is not such limited; But able to see the direction positive or negative.”



That is the reason we really need to restrict these values between some range so for that specific region we use.

2 Pearson Correlation Coefficient which restrict all your value to -1 to +1.

The more towards +1 more positively corelated.

The more towards -1 more negatively corelated.