Observed Values:

$O_1 = 121$

$O_2 = 288$

$O_3 = 91$

Expected count:

$E_1 = 500 \times \frac{20}{100} = 100.$

$E_2 = 500 \times \frac{30}{100} = 150$

$E_3 = 500 \times \frac{50}{100} = 250.$

All expected values are ≥ 5. Hence we need to perform Chi Square test.

Chi square test formula:

$$\chi^2 = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i}$$

For $i=1$.

$$\chi^2 = \frac{(121-100)^2}{121} = \frac{(21)^2}{121} = \frac{441}{100} = 4.41$$

For $i=2$

$$\chi^2 = \frac{(288-150)^2}{150} = \frac{(138)^2}{150} = \frac{19044}{150} = 126.96$$

For $i=3$

$$\chi^2 = \frac{(91-250)^2}{250} = \frac{(-159)^2}{250} = \frac{25281}{250} = 101.124$$

Overall,

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

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

From chi-square table find χ^2 value for $\alpha = 0.05$ & d.of $= 2$.

$\chi^2 = 5.991$

table χ^2 > calculated χ^2

Hence, we reject the H_0.

Nov-6:

n=500	Less than 18	18-35	>35
Observed	121	288	91
Expected	100	150	250

H_0: The data meets the distribution of 2000 census

H_A: The data doesn't meet the distribution of 2000 census.

$\alpha = 0.05$

d.of $= 3-1 = 2$.

In chi-square table for d.o.f 2 with $\alpha = 0.05$ is 5.991

If $\chi^2 > 5.991$ then reject the H_0 else accept the H_0.

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_o - f_e)^2}{f_e} = \sum_{i=1}^{n} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

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

$$= \frac{(21)^2}{100} + \frac{(138)^2}{150} + \frac{(-159)^2}{250}$$

$$= 4.41 + 126.96 + 101.124$$

$$\boxed{\chi^2 = 232.494}$$

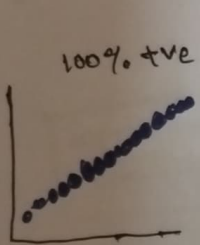
$\therefore \chi^2 > 5.991$ Hence, Reject the H_0.

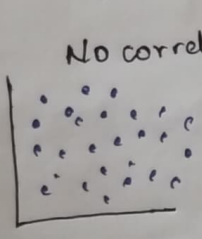
★

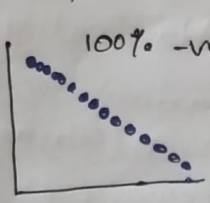
Correlation: To check the relation b/w two numerical features we use correlation.

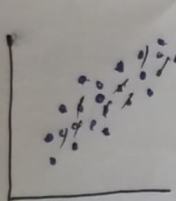
There are 3 types of correlation:

* Positive Correlation
* Negative Correlation
* No correlation.









We have 2 calculations for correlation.

1. Pearson Correlation
2. Spherman Correlation

Always correlation value ranges from -1 to +1

* The value near to +1 is +ve correlation
* The value near to -1 is -ve correlation
* The value near to 0 is no correlation.

Central Limit Theorem: If we are taking n samples and each sample containing x records then average of n samples is calculated, then plot the avg values, that will be following normal distribution. This is commonly known as Central Limit theorem.