Estimation and Test of Hypothesis

Hypothesis: A statement about the nature of a population.

Example: Students who eat breakfast will perform better on a stat exam than students who do not eat breakfast.

Test of hypothesis: The statistical procedure which is used to verify any statement or assumption about population parameter on the basis of sample observations is known as test of significance.

Null hypothesis: The hypothesis which we are going to test for possible rejection under the assumption that it is true.

Null hypothesis is denoted by H_0;

$H_0 : \mu_1 = \mu_2$

Alternative hypothesis: Each of all possible hypothesis other than null hypothesis is called alternative hypothesis and is usually denoted by H_1;

$H_0 : \mu_1 \neq \mu_2$

Type-I error: The error of rejecting H_0 (accepting H_1) which is true is called Type-I error.

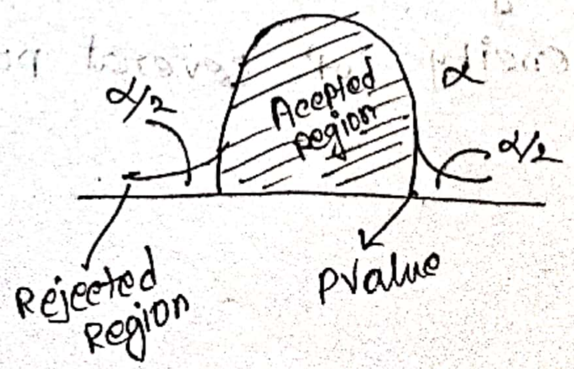
Type-II error: The error of accepting null hypothesis H_1 when it is false is called Type II error.

Type II error is denoted by β.

Level of significance: The probability of Type-I error is called level of significance.

- It's denoted by α
- 'α' could b 1% / 5% / 10%.

P-Value: P value is the least possible values of 'α' for which we can reject the null hypothesis.



Comment on P-value:

1. If $P < 1\%$; It is highly statistical significant.
2. If $P < 5\%$; It is significant
3. If $P > 10\%$; It is not significant.

Commonly used Test statistic:

1. The normal test (z-test)
2. The 't' test
3. Chi-square (χ^2) test
4. F-test

Application of chi-square (χ^2) test:

1. To test the significance of a specified population variance.
2. To test the goodness of fit of a distribution.
3. To test the independence of attributes.
4. To test the equality of several correlation coefficients.
5. To test the homogenity of several tests.
6. To test the homogeneity of several population variance.

Problem: Two hundred engineers were interviewed and classified according to their results and job satisfaction. The distribution of graduates by results and job satisfaction are given in the following contingency table.

Results	Job satisfaction	
	Yes	No
Excellent	20	70
Good	45	65

Compute the value of chi-square for the above data.

Solution: Computation table:

Result	Job satisfaction		Total
	Yes	No	
Excellent	20 (O_{11})	70 (O_{12})	90 (R_1)
Good	45 (O_{21})	65 (O_{22})	110 (R_2)
Total	65 (C_1)	135 (C_2)	200 (N)

We know,

$$\chi^2 = \Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Here, $E_{11} = \frac{R_1 \times C_1}{N} = \frac{90 \times 65}{200} = 29.25$

$E_{12} = \frac{R_1 \times C_2}{N} = \frac{90 \times 135}{200} = 60.75$

$E_{21} = \frac{R_2 \times C_1}{N} = \frac{110 \times 65}{200} = 35.75$

$$E_{22} = \frac{R_2 \times C_2}{N} = \frac{110 \times 135}{200} = 74.25$$

$$\therefore \chi^2 = \frac{(O_{11} - E_{11})^2}{E_{11}} + \frac{(O_{12} - E_{12})^2}{E_{12}} + \frac{(O_{21} - E_{21})^2}{E_{21}} + \frac{(O_{22} - E_{22})^2}{E_{22}}$$

$$= \frac{(20-29.25)^2}{29.25} + \frac{(70-60.75)^2}{60.75} + \frac{(45-35.75)^2}{35.75} + \frac{(65-74.25)^2}{74.25}$$

$$= 2.93 + 1.41 + 2.39 + 1.15$$

$$= 7.88.$$

(Ans)