Type 1 and Type 2 Error ⇒ Also gets used in Machine learn.

Type 1 and Type 2 Error gets its reference from Confusion matrix.

Type 1 → False Positive → Result is 0 and Model predicts 1

Type 2 → False Negative ← Actual value is 1 and predicted is 0

Confusion matrix

Predicted \ Actual	1	0
1	TP	FP
0	FN	TN

Actual
1 → True
0 → False
→ Kind of Same Positive & Negative

- Lets see how Type 1 and Type 2 Error are related to confusion Matrix.

So according to this there are 4 Scenarios

But first let us understand whats our Reality & Decision

Decision :- Null hypothesis is True or Null Hypothesis is False

Reality :- Null hypothesis is True or Null Hypo is False

↓

4 Scenarios

Note - Actual & Predicted can take/switch their positions.

Outcome - 1 : we reject the Null hypothesis ⟶ when in reality is false [0, 0] ⇒ True Negative

we have done correct classification

Outcome 2: we fail to reject Null hypothesis/ Retain the Null hypothesis ⟶ But in Reality it is True [1, 1] ⇒ True Positive

Here also we have done correct classificatn

Outcome 3 → We reject the Null Hypothesis ⟶ But in reality it was True [0, 1] ⇒ False Positive ⇒ Actual is 0 and Predicted 1

↳ Type 1 Error

Outcome 4 ⇒ we fail to reject the Null hypothesis/ accepted it ⟶ Reality it was False [Null Hypo is accepted when it is False]

↳ Type 2 Error

Reject Kodi Jab Shi thi is — Type I

Accept Kri Jab Galat thi ⟶ Type II

- Confidence Interval and margin of Error.

we have used confidence interval in z-test and t-test.

To understand confidence interval we have to first understand

① Point - Estimate ⇒ a value of any statistics that estimates the value of an unknown population is called Point Estimate.

Its hard to calculate population parameters thats why we sample data.

$\bar{x} \longrightarrow \mu$ {conclusion}

2.95 3

Conclusions drawn from Outputs (mean, sd, median) received from sample data for population can be considered as point estimate.

Confidence Interval

we construct a CI to help estimate what the actual value of unknown population Mean is.

Ex If mean falls between 80 ⟷ 120 then we say we fail to reject Null Hypothesis.

Confidence Interval help us calculate this 80 to 120 value.

Confidence Interval = Point Estimate $\pm$ Margin of Error

$\bar{x} \pm$ Margin of Error

$\bar{x}$: Sample mean

Margin of Error $= Z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$

$$= \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

alpha = significance value = 1 − confidence Interval

$= Z_{\boxed{\text{some value}}}$ → Get this through Z table

Using these we find values of confidence interval.

Ex- On verbal section of CAT exam the SD is 100. A sample of 25 test takers has a mean of 520. Construct 95% CI around the mean

Ans $\bar{x} = 520$ $\sigma = 100$ $n = 25$ $CI = 95\%$

or 0.95 $\alpha = 0.05$

$$CI = \bar{x} \pm Z_{\alpha/2}\, \sigma/\sqrt{n}$$

$\bar{x} -$ "

$\bar{x} +$ "

$Z_{(0.025)} \frac{100}{25}$

$\boxed{Z_{(0.025)} \times 4} + 520$

$520 -$ "

CI = 480.8 & 559.2

Conclusion is Mean of population lies in these value & I am 95% sure abt that. → Imp

Chi-Square Test

The Chi Square Test for Goodness of fit test claims about population proportion {categorical variables} ⟶ ordinal, nominal

It is a non parametric test that is performed on categorical variables [ordinal, nominal]

Eg → Population of male that likes different colors of Bike.

proportion of data — related to population

	Theory → P	Sample
Yellow	1/3	22
Orange	1/3	17
Red	1/3	59

Theory ⇓ Theory categorical Distribution

Sample → observed categorical Distribution

Main Goal → If population proportion given in "Theory" is "True" to our "Sample" distribution

⇒ we check this through Chi-Square Goodness of Fit.

Q.J In 2010 census of city, the weight of Individuals in small city were found to be following

≤ 50kg	50–75	>75
20%	30%	50%

On 2020 ages of n = 500 individuals were Sampled. Below are the

< 50	50–75	>75
140	160	200

Using $\alpha = 0.05$ or $CI = 0.95$, would you conclude the population difference of weights has changed in the last 10 years?

⇒ Theory or Sample is different or Not

↳ Expected

< 50	50–75	>75
20%	30%	50%

Sample
↳ Observation

< 50	50–75	>75
140	160	200

what Could be expected As per theory from Sample

< 50	50–75	>75
0.2×500	0.3×500	0.5×500
= 100	= 150	250

⟹ There is bit change

Now we use Chi Square Test to test our hypothesis.

① Null Hypothesis : H_0 : The data meets expectatⁿ

Alternate " : H_1 : The data does not meet expectation

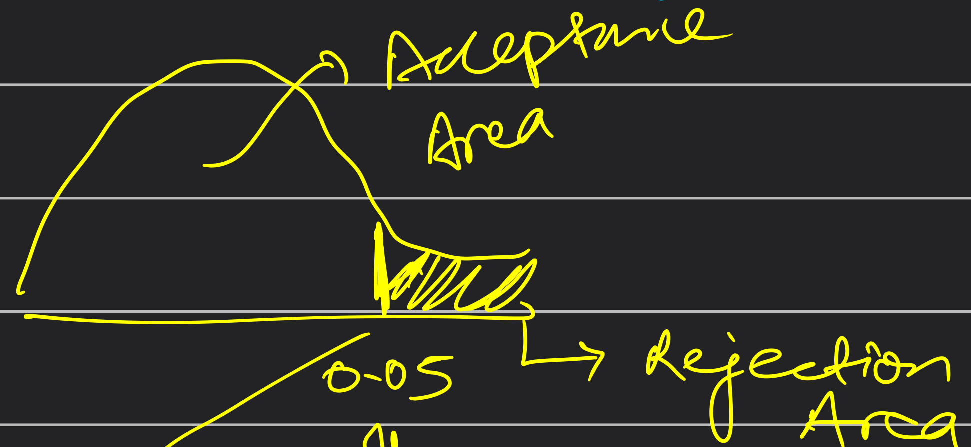
② $\alpha = 0.05$ $CI = 95\%$

③ Degree of Freedom = No of Categories − 1

$\quad$ || $\qquad\qquad\qquad$ = 3 − 1 = 2

$\quad$ df $\qquad\qquad\qquad$ <30, 30-75, >75

④ Decision Boundary ⇒ For Chi Square Test the graph is Right Skewed



$0.05 \approx \alpha$

To calculate this value → Critical Value we need Chi Square Table

Degree of Freedom and α we get value 5.99.

⑤ If χ^2 is greater than 5.99 (right skewed graph), Reject H_0 else we fail to reject H_0.

χ^2 ↙ Chi-Square

⑥ Calculate χ^2 - Test Stats

$$= \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{Expected}}$$

$$= \frac{(140-100)^2}{100} + \frac{(160-150)^2}{150} + \frac{(200-250)^2}{250}$$

$$\chi^2 = 26.66 \geq 5.99$$

(7) Conclusion → We reject Ho & there is population difference and it has increased when compared to census in 2010