3rd November:

Hypothesis Testing:

H_0/H_N → Null Hypothesis (True statement)

H_1/H_A → Alternative Hypothesis.

H_A will prove that either we have to accept the H_0 (or) to reject the H_0.

H_0: Statements

H_1 : opp. statements.

Examples:

1. H_0: You are going to score more than 90M in upcoming test.

   H_1: No, you are not going to score more than 90M in upcoming test.

2. H_0: If you get a job, your life will get settled.

   H_1: No, If you get a job also your life will not get settled.

3. H_0: Today it is going to be rainny.

   H_1: No, today it's not going to be rainny.

P-Value: (0-1)-range of P.

If $p <= \alpha$ we can reject H_0

If $p > \alpha$ we have to accept H_0

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

n : sample size.

$\hat{p}$: sample proportion

p_0 : Assumed population proportion in H_0.

α → Significance value.

$$\alpha = 1 - CI$$

C.I : Confidence Interval (68%, 95%, 99.7%) recommended CI.

For 68% CI:

$\alpha = 1 - 0.68$

$\alpha = 0.32$

For 95% CI:

$\alpha = 1 - 0.95$

$\alpha = 0.05$

For 99.7% CI: 0.997

$\alpha = 1 - 99.7$

$\alpha = 0.003$.

* Always we will go with 95% CI but if we want means we can calculate for any CI.

For 87% CI:

$\alpha = 1 - 0.87$

$\alpha = 0.13$

1. Assume $p = 0.25$, will you accept /reject the Null Hypothesis.

Here, $p > \alpha$ Hence, we will accept the Null Hypothesis.

2. Assume $p = 0.05$.

Here $\alpha = 0.13$

$p = 0.05$

$p < \alpha$ Hence, we can reject the Null Hypothesis.

Example:

H_0: The avg. height of Indian people is 5.4

H_1: The avg. height of Indian people is not 5.4

One tailed test → only one condition should satisfy.

two " " → two conditions should satisfy

Conclusion:

Hypothesis testing is framework for making inferences about data & models (ML models) in Machine learning, it helps in model evaluation, feature selection, Assumption validation and ensuring the robustness & reliability of conclusions drawn from models.

Type I & Type II Error:

Reality — Decision

Both happens on H_0.

1. R H_0 True and D H_0 True (✓)
2. R H_0 True and D H_0 False → Type I error
3. R H_0 False and D H_0 True → Type II error.
4. R H_0 False and D H_0 False (✓)

Error statements.

Ex:

H_0: Besant BTM is no.1 institute in Bangalore.

H_1: Besant BTM is not no.1 institute in Bangalore.

* Most dangerous error is type II error.

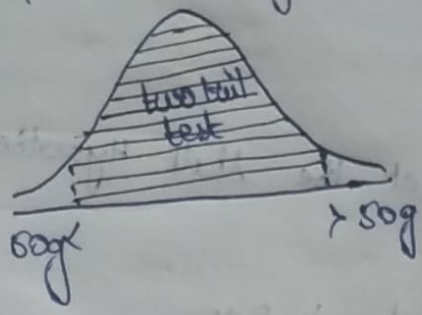
Note:

* If you failed to accept H_0, it is type I error.
* If you failed to reject H_0, it is type II error.

One tail test & two tail test :

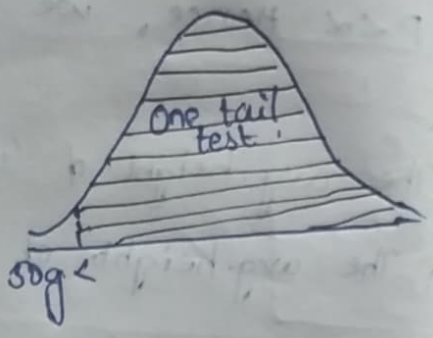
Ho: The chips packet weight is 50g

H1: No, the chips packet weight is not 50g.



Ho: The chips packet weight is >50g

H1: No, The chips packet weight is <50g.



Z-test & T-test :

1. The avg. age of college students is 24 years with the SD 1.5. Sample of 36 students. the mean is 25 years with 95% CI do the age will vary or not?

Ho: The avg age is 24 years.

H1: No, the avg age is not 24 years.

$\mu = 24$, $\sigma = 1.5$, $n = 36$, $\bar{x} = 25$, CI = 95%, $\alpha = 0.05$

Z-test (✓)

- When they give population SD do Z-test.

T-Test

- When they give sample SD do T-test.

Ho: Age = 24
H1: Age ≠ 24
} Two-tailed test.