⊛ Hypothesis testing and Hypothesting testing mechanism

⇒ Hypothesis testing is a core part of inferential statistics that uses data from the sample to make conclusions about a larger population.

⇒ Hypothesis testing is a process to evaluate whether a claim or an assumption about a population. It allows researchers to determine if a specific observation is statistically significant or could have happened by random choice.

(A) Hypothesis testing mechanism :-

(i) NULL Hypothesis (H0):-

This is the default position. It is the statement the researcher is trying to disprove.

Example:- If we are testing a new drug, the null hypothesis would be that the drug has no effect on the diseases.

(ii) Alternate Hypothesis (H1) :-

This is the claim that we are trying to prove. It is the opposite of null hypothesis.

Example:- If we are testing a new drug then alternate hypothesis would be that the drug does have effect on the disease.

(iii) Experiment :-

After setting up the hypothesis, a statistical experiment is performed, which involves collection of a sample data from the population we are studying.

Example :- If we were testing a new drug, then we would give this to a sample of few patients and then we a test statistic and a p-value.

p-value ⇒ The probability of observing your data / something more extreme if the null hypothesis is true.

(iv) Action :-

In the final step, we compare the p-value and the pre-determined significance level (α).

If :-

$p \leq \alpha$ ⇒ Observed results are highly unlikely and hence we can reject the null hypothesis.

$p > \alpha$ ⇒ observed results are likely to have occured and hence we fail to reject the null hypothesis

P-value

⇒ The p-value is the probability of observing a result as exterme as/ extreme than the one we obtain in our experiment, assuming null Hypothesis (Ho) is true.

It is a measure of how surprising the data is provided that nothing special is going on.

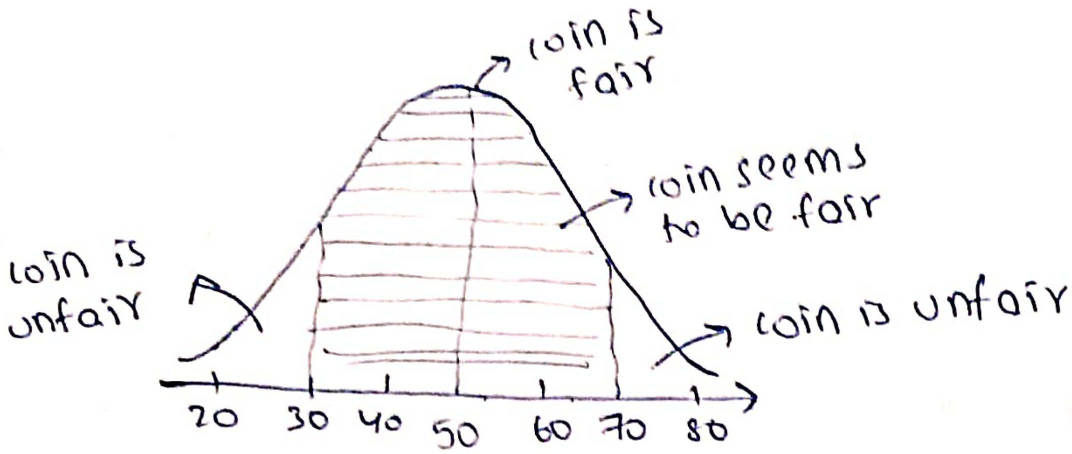
Small p-value ⇒ Results are surprising
Large p-value ⇒ Results are not surprising.

Example:-

An experiment to check whether a coin is fair or not by tossing it 100 times.

If. P(Head) = 0.50 ⇒ coin seems to be fair
P(Head) = 0.60 ⇒ coin still seems to be fair
P(Head) = 0.70 ⇒ coin looks to be unfair

Here, the p-value is the probability of getting a result as extreme as/more extreme than obtaining a head 70 times out of 100 times.



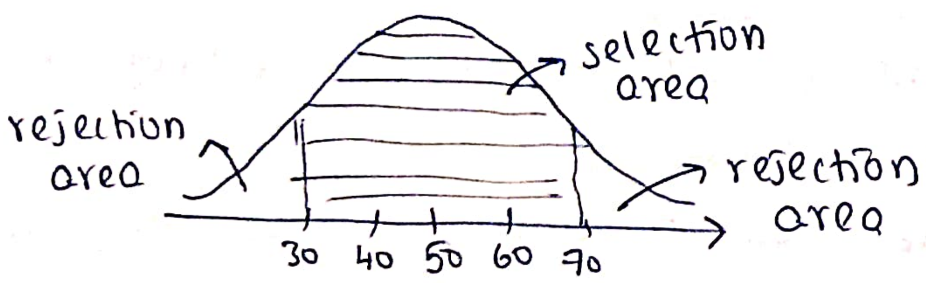
So, the p-value here will be the probability of getting 0-30 and 70-100 times heads when a coin is flipped 100 times

i.e p-value = $\sum_{k=0}^{30} P(X=k) + \sum_{l=70}^{100} P(X=l)$

<u>when compared to pre-defined significance level (α) :-</u>

If p-value $\leq \alpha \Rightarrow$ we can reject the null hypothesis

If p-value $> \alpha \Rightarrow$ we fail to reject the null hypothesis and conclude that the coin is fair.



Hypothesis testing and statistical analysis

→ Different tests for statistical analysis:-

(i) Z-test
(ii) t-test } statistical analysis for data dealing with mean

(iii) chi-square test } statistical analysis for categorical data

(iv) Annova test } statistical analysis for data dealing with variance

(A) Z-test:-

Z-test is a type of hypothesis test used to determine if a sample mean is significantly different from the population mean provided that the population standard deviation is known

when can we use Z-test:-

(i) Data is normally distributed
(ii) Population standard deviation is known
(iii) sample size is large, $(n) \geq 30$. Due to the central limit theorem, even if data is not perfectly distributed, a larger sample size will ensure that the sampling distribution of mean is approximately normal.

The Z-test is less normal in real-world applications since the t-test helps in the cause even when population standard deviation is unknown.

Formula for Z-test:-

$$\boxed{Z\text{-score} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}$$

$\bar{x} \Rightarrow$ sample mean

$\mu \Rightarrow$ Population mean

$\sigma \Rightarrow$ Population standard deviation

$\sigma/\sqrt{n} \Rightarrow$ standard error of the mean. It measures the variability of the spread of sample mean from the Population mean.

we take n samples and calculate the mean of each sample. These samples form their own distribution, which is narrower than the original population's distribution.

The central limit theorem states that the standard deviation of sample means will be equal to standard deviation of the population divided by $\sqrt{n}$.

$$\boxed{\sigma_{means} = \sigma/\sqrt{n}}$$

That is the reason why, in the denominator, we use standard error of the mean ($\sigma/\sqrt{n}$) instead of population standard deviation.

① Left - tailed test:-

This test is used when the alternate hypothesis claims that the population mean is less than the null-hypothesis value.

Example:-

ⓐ A factory manufactures bulbs with an average warranty of 5 years. A worker believes that the bulbs will malfunction is less than 5 years. He tests a sample of 40 bulbs and finds the average malfunction time to be 4.8 years, with a standard deviation of 0.5. At 2% significance level, ($\alpha = 0.02$), is there enough evidence to reject the idea that the ~~null~~ warranty should be revised?

ⓢ Null hypothesis (H_0):-

The average warranty is 5 years ($\mu = 5$)

Alternate hypothesis (H_1):-

The average warranty is less than 5 years ($\mu < 5$ years).

So, the p-value here will be getting the probability of sample mean 4.8 or below. Since, the direction is "less than", we look at the area of left tail.

The, Z-score is

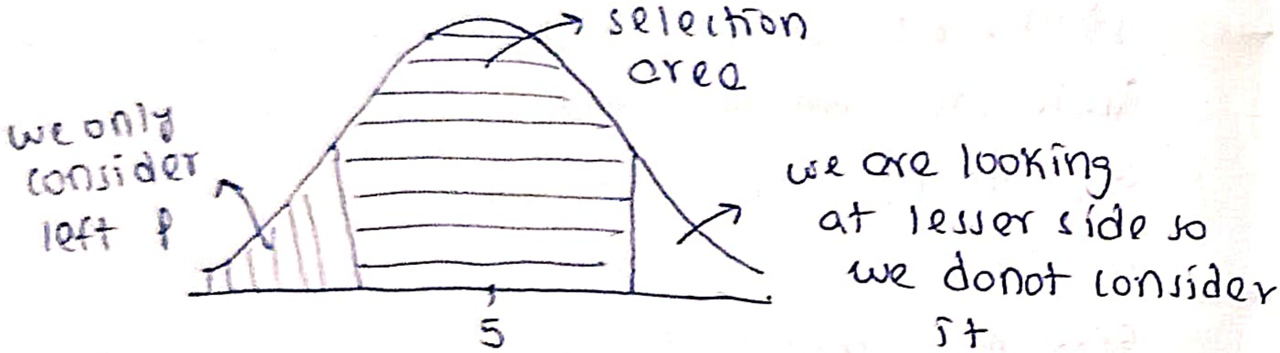
$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$\bar{x} = 4.80$ years $\quad \sigma = 0.5$ years

$\mu = 5$ years $\quad n = 40$ bulbs

$$z\text{-score} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{4.80 - 5}{0.5/\sqrt{40}} = -2.53$$

As, per Z-table, the corresponding p-value for $z = -2.53$ is 0.0057.



$p = 0.0057$, $\alpha = 0.02$ } since $p < \alpha$, we reject the null hypothesis (H_0).

It is left-tailed test because we only look at the rejection area towards the left.

(ii) Right-tailed test:-

This test is used when alternote hypothesis claims the population mean is greater than null-hypothesis value. The rejection region is entirely in the right tail.

Example:-

(6) A factory manufactures bulbs with an average warranty of 5 years. The CEO claims that with a new manufacturing process, the average warranty will be more than 5 years. A sample of 40 bulbs from the new process has a warranty of 5.2 years. If the standard deviation is 0.5, is there enough evidence at 0.02/2% significance level to support the claim of CEO?

(S) Null hypothesis (H_0):-

Average warranty is 5 years ($\mu = 5$)

Alternate hypothesis (H_1):-

The average warranty is greater than 5 years ($\mu > 5$)

Since the p-value is about probability of getting mean of 5.2 or more, the direction is "greater than" and we look at rejection area in the right tail.

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{5.20 \text{ years} - 5 \text{ years}}{0.50/\sqrt{40}}$$

$$= +2.53$$

The corresponding P-value for a z-score of +2.53 in the z-table is 0.0057.

$$\left.\begin{array}{l} \text{P-value} = 0.0057 \\ \alpha = 0.02 \end{array}\right\} P < \alpha$$

Since $P < \alpha \Rightarrow$ we will reject the H_0.

We can say that there is some evidence to claim the statement of CEO that the worranty of bulb is greater than 5 years.

(iii) Two-tailed test:-

This test is used when the alternative hypothesis claims the population mean is different (either greater or lesser) than the hypothesis value. The rejection area is both at left and right tail.

Example:-

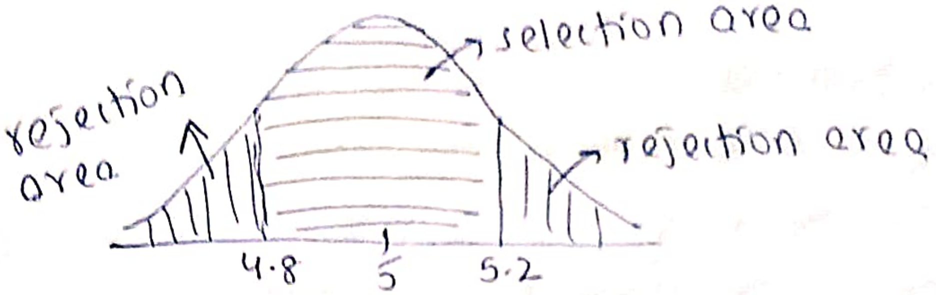
(Q) A factory manufacturers manufacture bulb with an average worranty of 5 years. The quality control manager checks if the average worranty has changed. A sample of 40 bulbs has an average malfunction time of 4.8 years. The population standard deviation is 0.50. At 2% significance level, is there any evidence to suggest that average worranty is no longer 5 years.

(S) Null hypothesis:- (H_0)

The average worranty is 5 years ($\mu = 5$)

Alternate hypothesis (H_1):-

The average worranty is not 5 years ($\mu \neq 5$).



Here, the p-value is the probability of sample mean to be less than 4.8 years (as per question) or probability of sample mean to be greater than 5.2 years (symmetrical difference). So, rejection area of both the ends should be considered

$$z\text{-score} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{4.80-5}{0.5/\sqrt{40}} = -2.53$$

z-score of right extreme would be +2.53.

So, as per the z-table, the p-value for left tail at $z = -2.53$ is 0.0057 and the p-value for right tail at $z = +2.53$ is 0.0057. The total p-value will be:-

$$P = \text{p-value (left)} + \text{p-value (right)}$$
$$= 0.0057 + 0.0057 = 0.0114$$

$$\left.\begin{aligned} \text{p-value} &= 0.0114 \\ \alpha\text{-value} &= 0.02 \end{aligned}\right\} \text{p-value} < \alpha$$

we reject the null hypothesis that the warranty of bulbs is 5 years.

Student t- distribution

⇒ A t-test/ student's t-test is a statistical hypothesis test used to determine if there is a significant difference between means of two groups. It is particularly used when sample size is small ($n<30$) and population standard deviation is unknown

⇒ Formula to obtain t-value is

$$\boxed{t = \frac{\bar{x} - \mu}{s/\sqrt{n}}}$$

$\bar{x}$ ⇒ sample mean
μ ⇒ population mean
s ⇒ sample standard deviation
n ⇒ sample size.

⇒ Like we have a z-table when performing analysis using z-value, we have a t-table too.

There is an additional parameter in t-table:-

Degree of freedom (dof):-

$$\boxed{dof = n-1}$$

Example of why degree of freedom is n-1:-

If 3 people are given choice of chosing a chair from three chairs:-

First person ⇒ ☐ ☐ ☐ } 3 options
Second person ⇒ ☒ ☐ ☐ } 2 choices
Third person ⇒ ☒ ☒ ☐ } no choice

only 2 people have choices while the last person doesn't

Example problem of t-stats with t-test hypothesis testing:-

(Q) In the population, the average IQ is 100. A team of researchers want to test a new medication to see if it has either a positive/negative effect on the intelligence, or no effect at all. A sample of 30 participants who have taken medication have a mean IQ of 140 with a standard deviation of 20. Did the medication affect intelligence at 95% confidence interval?

(S) confidence interval (C.I) = 95% = 0.95

significance level (α) = 1 − C.I = 0.05

Population mean (μ) = 100

sample mean ($\bar{x}$) = 140

sample size (n) = 30

sample standard deviation (s) = 20

(i) NULL hypothesis:- (H_0)

The population mean IQ is 100, $\mu = 100$

(ii) Alternate hypothesis:- (H_1)

The population mean IQ is not 100, $\mu \neq 100$

$\mu \neq 100 \Rightarrow$ Two tail test since it can either be greater or lesser.

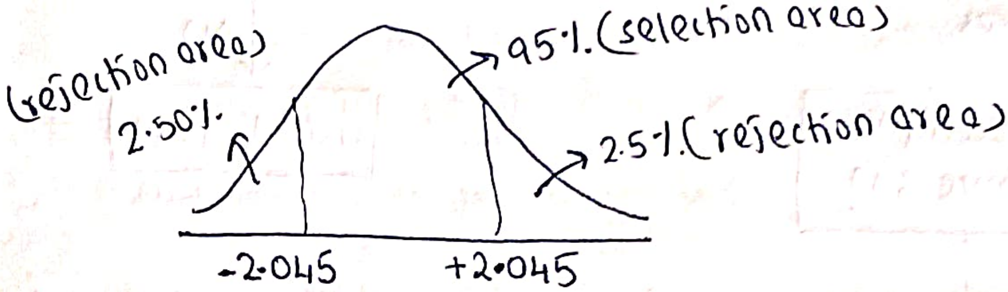
(iii) significance level (α) = 0.05

(iv) Finding degree of freedom,

dof = n − 1 = 30 − 1 = 29

(v) Decision rule

From the t-table,
with dof = 29, two tail test at a significance level of 0.05, the corresponding value will be ± 2.045



If the t-value is either less than -2.045 or greater than +2.045, we can reject the null hypothesis.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{140-100}{20/\sqrt{30}} = 2\sqrt{30} = 10.96$$

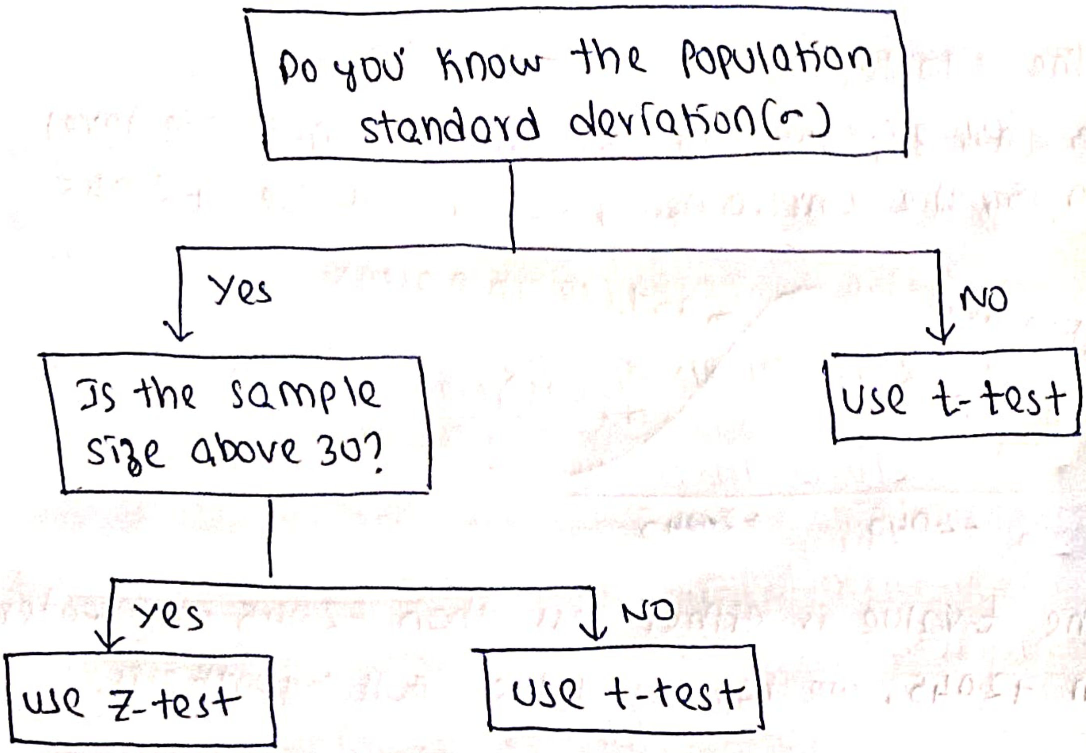
Since t-value is greater than +2.045, it falls in the right tail (rejection area) and hence we can reject the null hypothesis.

We can conclude that the medication use did effect on the intelligence and since the t-value fell in the right tail area, we can say that the medication has increased the average IQ.

④ When to use t-test and Z-test?

A t-test is used when:-
(i) sample size (n) is low (n<30)
(ii) Population standard deviation is not known.

A z-test is used when:-
(i) sample size (n) is high (n>30)
(ii) Population standard deviation is known.



Do you know the population standard deviation(σ)
yes
NO
use t-test
Is the sample size above 30?
yes
NO
use Z-test
use t-test

Type1 and Type2 Errors:-

⇒ The possibilities of NULL hypothesis are:-

In reality ⇒ NULL hypothesis can be True/False

Decision made ⇒ NULL hypothesis can be True/False

when we take all possibilities of combinations:-

NULL hypothesis is false in reality & we reject NULL hypothesis } Good/correct

NULL hypothesis is true in reality & we reject the NULL hypothesis } Type 1 Error

NULL hypothesis is false in reality & we retain the NULL hypothesis } Type 2 Error

NULL hypothesis is true in reality & we retain the NULL hypothesis } Good/correct

Bayes Statistics

Bayesian statistics is an approach to data analysis and parameter estimation based on Bayes theorem.

Bayes theorem :-

Probability → Independent events
→ Dependent events

⇒ Independent events :-

(i) Tossing a coin

$Pr(Head) = 1/2$, $Pr(Tail) = 1/2$

occurence of one event will not effect other

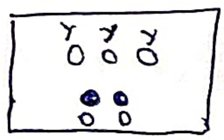
(ii) Rolling a fair dice

$Pr(1) = Pr(2) = Pr(3) = Pr(4) = Pr(5) = Pr(6) = 1/6$

occurence of one events will not effect others.

⇒ Dependent events :-

Taking out marbles from box



$Pr(\text{Dark marble}) = 2/5$

After removing a dark marble,

$Pr(\text{yellow marble}) = 3/4$

Removing of dark marble is effecting the probability of removing the yellow marble

$$\boxed{Pr(\text{Dark and yellow}) = Pr(\text{Dark}) \times Pr(\text{yellow}/\text{Dark})}$$

since, the probability is commutative,

$Pr(\text{Dark and Yellow}) = Pr(\text{yellow and Dark})$

$\rightarrow Pr(Yellow) \times Pr(Dark/Yellow) = Pr(Dark) \times Pr(Yellow/Dark)$

$$\boxed{Pr(Yellow) = \frac{Pr(Dark) \times Pr(Yellow/Dark)}{Pr(Dark/Yellow)}}$$

$$\left.\begin{matrix} Pr(Yellow/Dark) \\ Pr(Dark/Yellow) \end{matrix}\right\} \text{It is called conditional probability.}$$

$$\left.\boxed{Pr(B/A) = \frac{Pr(B) \times Pr(A/B)}{Pr(A)}}\right\} \text{Bayes theorem}$$

A, B $\Rightarrow$ Events

$Pr(A/B) \Rightarrow$ Probability of A provided B has occured

$Pr(B/A) \Rightarrow$ Probability of B provided A has already occured.

$Pr(A), Pr(B) \Rightarrow$ Independent probabilities of A and B

Importance of Bayes theorem

Suppose we have a dataset

Size of house (x_1)	No. of rooms (x_2)	location (x_3)	Price (y)

Independent variables — Dependent variable

As per Baye's theorem,

$$\boxed{Pr(y/x_1, x_2, x_3) = \frac{Pr(y) \times P(x_1, x_2, x_3/y)}{Pr(x_1, x_2, x_3)}}$$

This is called Naive Bayes theorem

Confidence interval and Margin of error

⇒ we generally use point estimate to estimate the actual population mean

Point estimate $(\bar{x}) \rightarrow$ Population mean (μ)

The point estimate can either be smaller/equal/greater than the actual population mean

⇒ **Margin of error:-**

margin of error for Z-test = $Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

Margin of error for t-test = $t_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

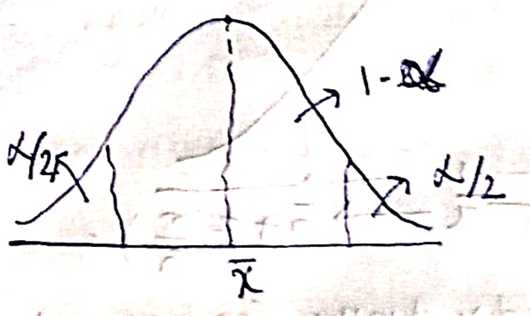
⇒ **confidence interval range:-**

C.I = Point estimate $(\bar{x})$ ± Margin of error

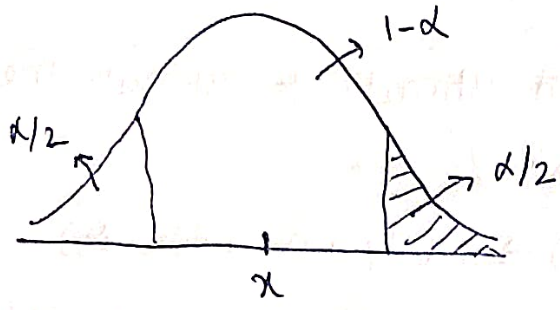
$$C.I = \begin{cases} \bar{x} \pm Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right), & \text{for Z-test} \\ \bar{x} \pm t_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right), & \text{for t-test} \end{cases}$$

(*) **understanding why/how it can be calculated:-**

For a significance level, (α)

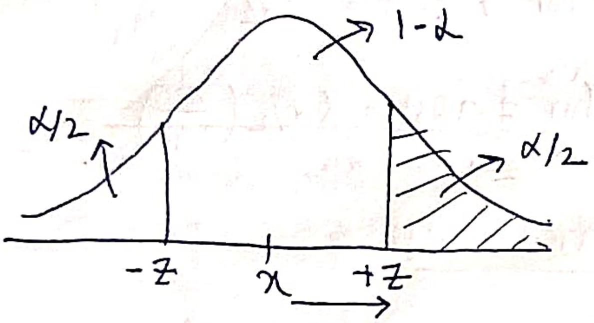


considering for higher confidence interval, we only look at the right tail



Total area we are interested in = $(1-\alpha) + \alpha/2 = 1 - \alpha/2$

Let the value corresponding to area $1-\alpha/2$ be Z, then since normal distribution is symmetric,



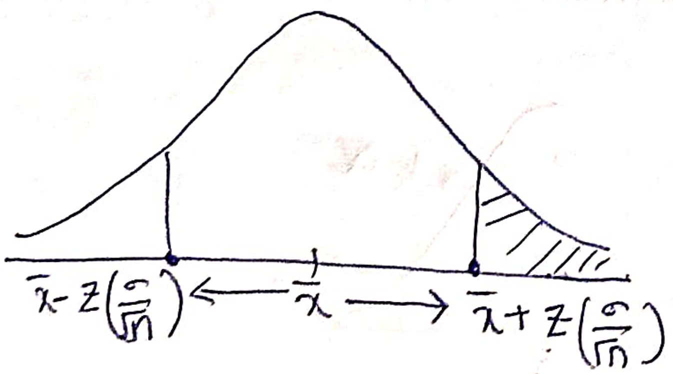
Margin of error towards higher side/right side

$$= Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) = (+Z)\left(\frac{\sigma}{\sqrt{n}}\right)$$

margin of error towards lower side/left side

$$= Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) = (-Z)\left(\frac{\sigma}{\sqrt{n}}\right)$$

So, the range will be



So, instead of individually finding out Z-values for left tail to obtain lower confidence interval and right tail for higher confidence interval, we always only consider margin of error of right / higher side $(+z)\left(\frac{\sigma}{\sqrt{n}}\right)$ and deduct the same to obtain lower confidence interval range $(-z)\left(\frac{\sigma}{\sqrt{n}}\right)$.

This is true because normal distribution is symmetric.

Ⓠ An Example Question:-

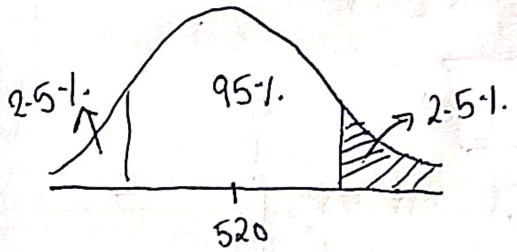
on the verbal section of CAT exam, the standard deviation is known to be 100. A sample of 25 students have a mean score of 520. construct 95% confidence interval about the mean

Ⓢ $\alpha = 1 -$ confidence interval $= 1 - 0.95 = 0.05$

$\sigma = 100$ $\quad\quad \bar{x} = 520$

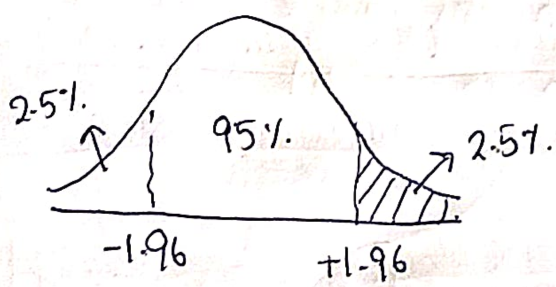
$n = 25$

Area = 97.5%.



corresponding value for 97.5% area covered is +1.96

so, $Z_{0.025} = +1.96$



Higher confidence interval $= \bar{x} + Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

$= 520 + (+1.96)\left(\frac{100}{\sqrt{25}}\right)$

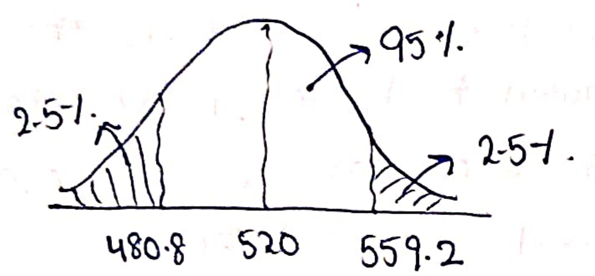
$= 520 + 39.20 = 559.20$

Lower confidence interval $= \bar{x} - Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

$$= 520 - (1.96)\left(\frac{100}{\sqrt{25}}\right) = 520 - 39.20$$

$$= 480.80$$

so, the range of mean at 95% confidence interval is [480.80, 559.20]



CHI SQUARE TEST

→ It is also known as chi square test for goodness of fit. It claims about population proportions.

→ It is a non-parametric test, that is performed on categories. i.e nominal and ordinal data.

⊗ Example of goodness of fit test:-

In a science class of 75 students, 11 are left handed. Does this class fit the theory that 12% of the people are left handed?

S

Category	Theory categorical distribution	Observed categorical distribution
Left handed	12% of 75 (9)	11
Right handed	88% of 75 (66)	64
	75	75

NULL hypothesis (H_0):-

The class data fits the theory. The proportion of left handed students is 12%.

Alternate hypothesis (H_1):-

The class data doesnot fit the theory. The proportion of left handed students is not 12%.

Category	Observed (O)	Expected (E)
Left	11	9
Right	64	66

The chi square test statistic (χ^2) is:-

$$\boxed{\chi^2 = \sum \left[(O-E)^2/E\right]}$$

For left handed, $(O_L - E_L)^2/E_L = (11-9)^2/9 = 0.444$

For right handed, $(O_R - E_R)^2/E_R = (64-66)^2/66 = 0.061$

Total chi-square test is:-

$$\chi^2 = \sum \left[(O-E)^2/E\right] = \frac{(O_L - E_L)^2}{E_L} + \frac{(O_R - E_R)^2}{E_R}$$

$$= 0.444 + 0.061 = 0.505$$

Let us consider at 95% confidence level.

so, α (significance level) $= 0.05$

$$\boxed{\text{Degree of freedom} = K-1}$$

where K = number of categories.

dof $= K-1 = 2-1 = 1$

The critical value in chi-square table for dof $=1$ and $\alpha = 0.05$ is 3.841

Decision rule:-

$\chi^2 <$ critical value $\Rightarrow$ Reject null hypothesis

$\chi^2 >$ critical value $\rightarrow$ Retain null hypothesis / fail to reject null hypothesis since $X^2 = 0.505$ is less than critical value $= 3.841$, we fail to reject the nulls hypothesis.

So, based on chi-square goodness of fit test, there is no strong statistical evidence to say that number of left-handed students in the class is different from theoretical 12%.

(4) Another example:-

In 2010 census, in a city, the weight of the individuals in a small area were found to be the following

<50Kg	50-75kg	>75Kg
20%	30%	50%

In 2020, weight of 500 individuals were sampled and the results are

<50Kg	50-75Kg	>75Kg
140	160	200

using $\alpha = 0.05$, would you conclude that the population difference of weights has been changed in the last 10 years?

(5)

category	Expected (E)	Observed (O)
<50Kg	20% of 500 = 100	140
50-75 Kg	30% of 500 = 150	160
>50Kg	50% of 500 = 250	200

NULL hypothesis :- (H_0)

The data meets the expectation and hence there is no population change in weight.

Alternate Hypothesis :- (H_1)

The data doesn't meet the expectation and hence there is population change in weight.

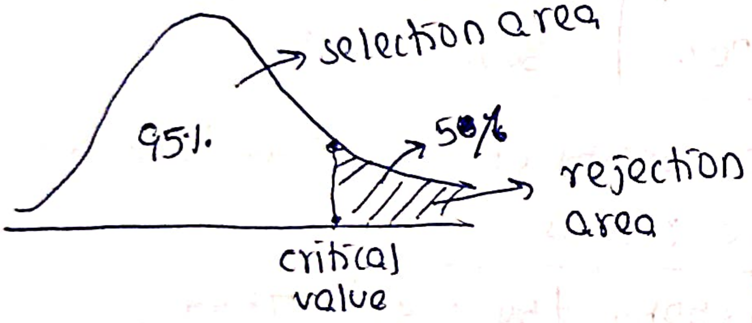
Finding chi-square value:-

$\alpha = 0.05$

Degree of freedom = $K-1$ = No. of categories -1

$= 3-1 = 2$

The decision boundary is:-



The critical value for $\alpha = 0.05$ and dof = 2 is 5.991

So, if

$\chi^2 > 5.991 \Rightarrow$ we reject the null hypothesis

$\chi^2 \leq 5.991 \Rightarrow$ we fail to reject the null hypothesis

Let us calculate χ^2 (chi-square test statistic) :-

$$\boxed{\chi^2 = \sum \left[(O-E)^2 / E \right]}$$

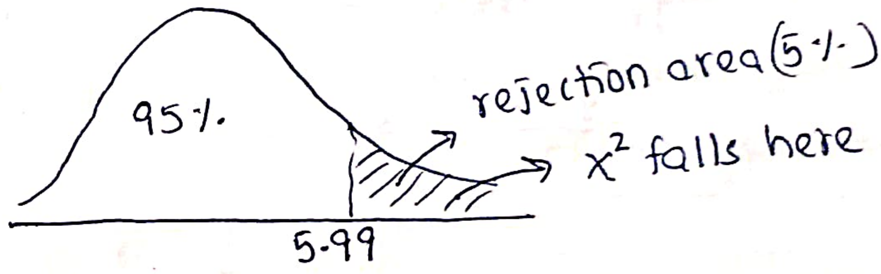
For our data,

$$\chi^2 = \sum \left[(O-E)^2 / E \right] = \frac{(140-100)^2}{100} + \frac{(160-150)^2}{150} + \frac{(200-250)^2}{250}$$

$$= \frac{1600}{100} + \frac{100}{150} + \frac{2500}{250}$$

$$= 16 + 0.667 + 10 = 26.667$$

$\chi^2 (26.667) >$ critical value (5.99), so we reject the NULL hypothesis



So, we can reject the NULL hypothesis.

At 95% confidence level using the chi-square goodness of fit test, there is an evidence to say that the weights of population in 2020 are different than expected in 2010 population.

Analysis of Variance (ANOVA)

ANOVA is a statistical method used to compare the means of two or more groups

(A) Components of ANOVA:-

(a) Factors ⇒ Factor is an independent variable that we are investigating in an experiment.

(b) Levels ⇒ Levels are different values/groups/conditions within a factor that represent the variations of the independent variable that we will be comparing.

Examples of factors and levels:-

(i) Factor :- Dosage of medicine
Levels :- 5mg, 10mg, 15mg

(ii) Factor :- Mode of Payment
Levels :- GooglePay, PhonePe, cash

(A) ASSUMPTIONS of ANOVA:-

(i) The distribution of sample means should be normal/Gaussian distribution.

(ii) Outliers should be removed from the dataset.

(iii) Population variance of different levels within the factor should be equal.

(iv) Samples should be independent and randomly selected.

(A) Types of ANOVA:-

(i) One way ANOVA:- It contains one factor with atleast two levels and these levels will be independent

Example of one way ANOVA:-

Doctor wants to test a new medicine to decrease headache. Doctor splits the participants in three groups in 3 different groups consuming 10mg, 20mg and 30mg dosages. Doctors ask the participants to rate the medicine on the scale of 1-10 about effectiveness of medicine.

Factor ⇒ medicine Dosage

Levels ⇒ 10mg, 20mg, 30mg

One factor (medicine dosage) has three levels (10mg, 20mg, 30mg) and all these levels are independent. So, this is a perfect example of one-way ANOVA.

(ii) **Repeated measures ANOVA**:- One factor with atleast two levels and levels are dependent.

Example of typing speed test:-

Factor ⇒ Typing speed

Levels ⇒ 10days practice, 20days and 30 days

One factor (typing speed) has three levels (practie of 10days, 20 days and 30 days) and levels are dependent. So, this is an example of repeated measures ANOVA.

Here, the levels are dependent because 20days of practice includes first 10days too which itself is a level for the factor, for the same person.
i.e The levels are dependent because the same individuals are being measured at every level. The dependency is from the person and not the time line itself.

(iii) <u>Factorial ANOVA</u>:- Two or more factors, each of which with atleast two levels and levels can be dependent / independent.

<u>Example of Factorial ANOVA</u>:-

Amount of / Distance covered by running for each individual day in 3 days by male and female.

Factor1 ⇒ Distance covered by running
Levels for factor1 ⇒ Day1, Day2, Day3
Since we are measuring the same individuals at every level, the levels are dependent here.

Factor 2 ⇒ Gender
Levels for factor2 ⇒ male and Female
The levels here are independent.

Since this example has two factors and atleast two levels for each factor, this is an example of factorial ANOVA.

Hypothesis testing in ANOVA / Partitioning of variance in ANOVA

⇒ A problem with one way ANOVA

Doctor wants to test a new medication which reduces headache. They split the participants into 3 conditions of 15mg, 30mg, 45mg dosages. Later on, the doctor asks the participants to rate the effectiveness of medication on the scale of 1 to 10. Are there any differences in 3 conditions at 95% confidence interval?

15mg	30mg	45mg
9	7	4
8	6	3
7	6	2
8	7	3
8	8	4
9	7	3
9	6	2

(S) Step1 ⇒ Defining NULL and alternate hypothesis

$H_0 \Rightarrow \mu_{15} = \mu_{30} = \mu_{45}$

$H_1 \Rightarrow$ Atleast one mean is not equal

Significance level $(\alpha) = 1 - C.I = 1 - 0.95 = 0.05$

Degree of freedom (dof) = 21 samples of ratings − 1
= 20 samples of ratings in total

15mg	30mg	45mg
7 samples	7 samples	7 samples

Degree of freedom within = Total number of samples (N) - Number of categories (a)

$= 21 - 3 = 18$

Degree of freedom between = Number of categories (a) - 1

$= 3 - 1 = 2$

$dof_{between} = 2$

$dof_{between} = 2$

$dof_{within} = 18$

$dof_{total} = dof_{between} + dof_{within}$

$= 2 + 18 = 20$

Simply, we can instead calculate as.

$$\boxed{\begin{aligned} dof_{total} &= dof_{between} + dof_{within} \\ &= (a-1) + (N-a) = N-1 \end{aligned}}$$

In f-table, we will be using the values of df between, df within and α to obtain the critical value.

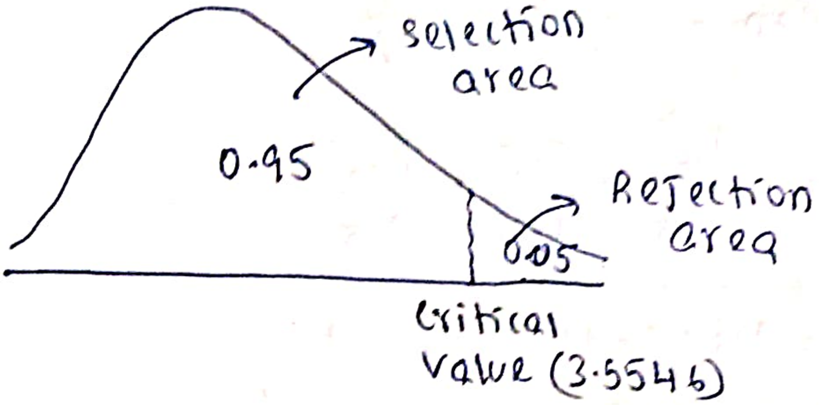
<u>Step2</u> :- obtain critical value and decision rule

$dof_{between} = 2 \Rightarrow (a-1)$

$dof_{within} = 18 \Rightarrow (N-a)$

α (significance level) $\Rightarrow 0.05$

As per the f-table, the critical value corresponding to df between, df within and α is 3.5546



As per the decision rule, if the f-value is greater than the critical value of 3.5546, then we can reject the NULL hypothesis

<u>Step 3</u> :- Find f-value / f-test statistics

$$\boxed{F = \frac{\text{variance between samples}}{\text{variance within samples}}}$$

Finding SS between first,

$$\boxed{\begin{aligned} \text{SS-between} &= \text{sum of squares between} \\ &= \sum \left(T_i^2 / n_i \right) - \left(T_G^2 / N \right) \end{aligned}}$$

$T_i \Rightarrow$ Total sum of all scores in group i

$n_i \Rightarrow$ Total number of scores in group i

$T_G \Rightarrow$ The sum of all scores in all groups

$N \Rightarrow$ Total number of participants across all groups.

Finding sum of square betweens,

$$SS\text{-}between = \sum \frac{T_i^2}{n_i} - \frac{T_G^2}{N}$$

$$= \frac{T_{15}^2}{n_{15}} + \frac{T_{30}^2}{n_{30}} + \frac{T_{45}^2}{n_{45}} - \frac{T_G^2}{N}$$

$$\left\{ \begin{aligned} T_{15} &= 9+8+7+8+8+9+8 = 57 \\ T_{30} &= 7+6+6+7+8+7+6 = 47 \\ T_{45} &= 4+3+2+3+4+3+2 = 21 \end{aligned} \right\}$$

$$SS\text{-}between = \frac{T_{15}^2}{n_{15}} + \frac{T_{30}^2}{n_{30}} + \frac{T_{45}^2}{n_{45}} - \frac{(T_{15}+T_{30}+T_{45})^2}{N}$$

$$= \frac{57^2}{7} + \frac{47^2}{7} + \frac{21^2}{7} - \frac{(57+47+21)^2}{21}$$

$$= \frac{3249 + 2209 + 441}{7} - \frac{15625}{21}$$

$$= 842.714 - 744.04 = 98.67$$

Finding SS-within :-

$$\boxed{\begin{aligned} SS\text{-}within &= \text{sum of squares within} \\ &= \sum \left[\sum x_i^2 - (T_i^2/n_i) \right] \end{aligned}}$$

$$SS\text{-}within \text{ for } 15mg = \left[9^2 + 8^2 + 7^2 + 8^2 + 8^2 + 9^2 + 8^2 - \left(\frac{57^2}{7}\right) \right]$$

$$= 162 + 256 + 49 - \left(\frac{57^2}{7}\right)$$

$$= 467 - \frac{3249}{7} = 467 - 464.142$$

$$= 2.85$$

SS-within for 30mg $= 7^2+6^2+6^2+7^2+8^2+7^2+6^2-\left(\frac{47^2}{7}\right)$

$= 147+108+64-\frac{2209}{7}$

$= 319-315.57 = 3.428$

SS-within for 45mg $= 4^2+3^2+2^2+3^2+4^2+3^2+2^2-\left(\frac{21^2}{7}\right)$

$= 32+27+8-\frac{441}{7} = 67-63 = 4$

SS-within $= (\text{SS-with})_{15mg} + (\text{SS-within})_{30mg} + (\text{SS-within})_{45}$

$= 2.85+3.428+4 = 10.29$

For F-test

	S.S	df	M.S	F
Between	98.67	2	49.34	
within	10.29	18	0.54	
Total	108.8	20	54.4	

$df_{between} = a-1$

$df_{within} = N-a$

$\boxed{M.S = S.S/df}$

$$F = \frac{\text{Variance between samples}}{\text{variance within sample}}$$

$$= \frac{\text{M.S between}}{\text{M.S within}} = \frac{49.34}{0.54} = 86.56$$

Since F-value = 86.56 is greater than the critical value of 3.5546, we can reject the NULL hypothesis that the sample means are equal for all samples.