Analysis of Variance (ANOVA):

It is the separation of variance ascribable to one group of causes from the variance ascribable to other group. ANOVA consists in the estimation of the amount of variation due to each of the independent factors (causes) separately and then comparing these estimates due to assignable factors with the estimates due to chance factors, the later being known as experimental error.

1. ANOVA I way classification
2. ANOVA II way classification.

1. ANOVA I way classification:

Formulae:

$$RSS = \text{Raw sum of squares} = \sum_{i=1}^{k}\sum_{j=1}^{h} x_{ij}^2$$

$$G = \text{Grand Total} = \sum_{i=1}^{k}\sum_{j=1}^{h} x_{ij}$$

$$CF = \text{Correction factor} = \frac{G^2}{N}$$; N = Total no. of observations.

$$TSS = \text{Total Sum of Squares} = RSS - CF$$

$$SST = \text{Treatment sum of squares} = \sum_{i=1}^{k}\left(\frac{T_{i\cdot}^2}{n_i}\right) - CF$$

SSE = Error Sum of Squares (Sum of Squares due to error)

= TSS − SST

ANOVA I way Table

Source of Variation	Degrees of Freedom (df)	Sum of Squares	mean Sum of Squares	Variance Ratio
Treatments	K−1	SST	$MST = \frac{SST}{K-1}$	$F = \frac{MST}{MSE} \sim F_{(K-1)(n-1)}$
Error	n−K	SSE	$MSE = \frac{SSE}{n-K}$	
TSS	n−1			

Problem:

The following table shows the time (in hours) of four batches of electric lamps.

Batches (K)	Life of bulbs in hrs							
1	1600	1610	1650	1680	1700	1720	1800	–
2	1580	1640	1640	1700	1750	–	–	–
3	1460	1550	1600	1620	1640	1660	1740	1820
4	1510	1520	1530	1570	1600	1680	–	–

Perform an analysis of variance of the data and show that a significance test does not reject their homogenity.

8d) H_0:- $\mu_1 = \mu_2 = \mu_3 = \mu_4$ ——> Null Hypothesis

H_1 : $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$ ——> Alternative Hypothesis

coding method: Subtract each number from 1600 and divide by 10

(max repeated numbers)

									T_i
1	0	-1	-5	-8	-10	-12	-20	—	-56
2	2	-4	-4	-10	-15	—	—	—	-31
3	14	+5	0	-2	-4	-6	-14	-22	-29
4	9	8	7	3	0	-8	—	—	19

$G = -97$

	x_{ij}^2	n_i
1	734	7
2	361	5
3	957	8
4	267	6
	$RSS = \Sigma\Sigma x_{ij}^2$	$N = 26$

$= 2319$

$N = 26$

$RSS = 2319$

$G = \Sigma\Sigma x_{ij} = -97$

Correction factor $= \frac{G^2}{N}$

$C.F = \frac{(-97)^2}{26} = 361.8846$

Total Sum of Squares = $TSS = RSS - CF$

$= 2319 - 361.8846$

$= 1957.1154$

4)

Sum of squares due to treatments

$$SST = \sum_{i=1}^{4} \left(\frac{T_i^2}{n_i}\right) - CF$$

$$= \left[\frac{(-56)^2}{7} + \frac{(-31)^2}{5} + \frac{(-29)^2}{8} + \frac{(19)^2}{6}\right] - 361.8846$$

$$= 443.6071$$

Sum of squares due to errors

$$SSE = TSS - SST$$

$$= 1957.1154 - 443.6071$$

$$= 1513.5083$$

ANOVA I way Table :

Source of variation	Degrees of freedom df	Sum of Squares	mean sum of squares	Variance ratio
Treatments	K-1 $\Rightarrow 4-1=3$	SST = 443.6071	$MST = \frac{SST}{K-1}$ $= 433.6071/3$ $= 147.8690$	$F = \frac{MST}{MSE}$ $= \frac{147.8690}{68.7952}$ $= 2.1494$
Error	N-K $\Rightarrow 26-4=22$	SSE = 1513.5083	$MSE = \frac{SSE}{N-K}$ $= 68.7952$	
TSS	N-1 $\Rightarrow 26-1=25$			

Table value of F at 5% level of significance at (3,22) df.

is 3.0491

Calculated value < Table value

2.1493 3.0491

Accept H_0

Problem:

4 salesman were posted in different areas by a company. The no of units of commodity x sold by them are as follows. Is there a significant difference in performance of salesmen?

TB = 3.4903

(k)				
A	20	23	28	29
B	25	32	30	21
C	23	28	35	18
D	15	21	19	25

Sol

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ → Depends on no of cases (i.e A, B, C, D)

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$

6)

Coding Method: Subtract each number from 2

					T_i	Σn_{ij}^2	n_i
A	+8	+5	0	-1	12	90	4
B	3	-4	-2	7	4	78	4
C	5	0	-7	10	8	174	4
D	13	7	9	3	32	308	4
					G=56	RSS = 650	N=16

Here $N=16$, $G=56$, $RSS=650$

$$CF = \frac{G^2}{N} = \frac{(56)^2}{16} = 196$$

Total sum of squares = $TSS = RSS - CF$

$$TSS = 650 - 196$$

$$\rightarrow TSS = 454$$

Sum of squares due to treatments

$$SST = \sum_{i=1}^{4} \left(\frac{T_i^2}{n_i}\right) - CF$$

$$= \left[\frac{(12)^2}{4} + \frac{(4)^2}{4} + \frac{(8)^2}{4} + \frac{(32)^2}{4}\right] - 196$$

$$= 312 - 196$$

$$SST = 116$$

Sum of squares due to errors

$SSE = TSS - SST$

$= 454 - 116$

$SSE = 338$

ANOVA I way table:

Source of variation	Degrees of freedom df	Sum of Squares	Mean Sum of Squares	Variance ratio
Treatments	$K-1$ $\Rightarrow 4-1 = 3$	$SST = 116$	$MST = SST/K-1$ $= 116/3$ $= 38.6666$	$F = \frac{MST}{MSE}$ $= \frac{38.6666}{28.1666}$ $= 1.3727$
Error	$N-K$ $\Rightarrow 16-4 = 12$	$SSE = 338$	$MSE = \frac{SSE}{N-K}$ $= \frac{338}{12} = 28.1666$	
Total	$N-1$ $\Rightarrow 16-1 = 15$	$TSS = 454$		

Table value of F at 5% level of significance at $(3, 12)$ df is 3.4903

($_N$ 3, $_D$ 12)

Calculated value < Table value

1.3727 < 3.4903

Accept H_0

Inference: There is no significant difference in the performance of the 4 salesman.

2. ANOVA II way classification:

Rows $\longrightarrow$ K

Columns $\longrightarrow$ h

Formulae:

$$CF = \frac{G^2}{N} \quad , \quad G = \sum_{i=1}^{K} \sum_{j=1}^{h} y_{ij}$$

$$TSS = RSS - CF \quad , \quad RSS = \sum_{i=1}^{K} \sum_{j=1}^{h} y_{ij}^2$$

$$SST = \frac{1}{h} \sum T_{i.}^2 - CF \quad ; \quad T_{i.} = \sum_{i=1}^{K} y_{ij}$$

Sum of squares due to varieties SSV

$$SSV = \frac{1}{K} \sum T_{.j}^2 - CF \quad , \quad T_{.j} = \sum_{j=1}^{h} y_{ij}$$

$$SSE = TSS - SST - SSV$$

ANOVA II way table:

Source of Variation	Degrees of freedom(df)	Sum of Squares	Mean Sum of Squares	Variance ratio
Treatments	K-1	SST	$MST = \frac{SST}{K-1}$	$F = \frac{MST}{MSE}$ ~
Varieties	h-1	SSV	$MSV = \frac{SSV}{h-1}$	$F_{(K-1),(K-1)(h-1)}$
Errors	(K-1)x(h-1)	SSE	$MSE = \frac{SSE}{(K-1)(h-1)}$	$F = \frac{MSV}{MSE}$ ~ $F_{(h-1),(K-1)(h-1)}$
Total	hK-1			

Problem:

A tea company appoints 4 salesmen A, B, C & D and observes that sales in 3 seasons - summer, winter and monsoon. The figures (in lakhs) are given in the following table.

Seasons	A	B	C	D	Seasons Total
summer	36	36	21	35	128
winter	28	29	31	32	120
monsoon	26	28	29	29	112
Salesmen Total	90	93	81	96	360

Sol

$H_{01}: \mu_1 = \mu_2 = \mu_3$ $\quad H_{02}: \mu_1' = \mu_2' = \mu_3' = \mu_4'$

$H_{11}: \mu_1 \neq \mu_2 \neq \mu_3$ $\quad H_{12}: \mu_1' \neq \mu_2' \neq \mu_3' \neq \mu_4'$

Coding Method: Subtract each value from 29

	A	B	C	D	$T_{i.}$	$T_{i.}^2$	n_i
s	-7	-7	-8	-6	-12	144	4
w	1	0	-2	-3	-4	16	4
m	3	1	0	0	4	16	4
$T_{.j}$	-3	-6	6	9	$G = -12$	$\Sigma T_{i.}^2 = 176$	$N = 12$
$T_{.j}^2$	9	36	36	81	$\Sigma T_{.j}^2 = 162$		

Rows = K = 3

Columns = h = 4

$N = 12$

$G = -12$

$CF = \frac{G^2}{N} = \frac{(-12)^2}{12} = 12$

$RSS = (-7)^2 + (-7)^2 + 8^2 + (-6)^2 + (1)^2 + (0)^2 + (-2)^2 + (-3)^2 + (3)^2 + 1^2 + 0^2 + 0^2$

$RSS = 222$

$SST = \frac{1}{h} \Sigma T_i^2 - CF = \frac{1}{4}(176) - 12 = 32$

$SSV = \frac{1}{K} \Sigma T_{.j}^2 - CF = \frac{1}{3}(162) - 12 = 42$

$TSS = RSS - CF$

$= 222 - 12 = 210$

$SSE = TSS - SST - SSV = 210 - 32 - 42 = 136$

ANOVA II way Table:

Source of variation	df	Sum of Squares	Mean sum of Squares	Variance (F) ratio
Treatments	K-1 3-1 = 2	SST = 32	MST = SST/K-1 = 32/2 = 16	$F = \frac{MST}{MSE} = \frac{16}{22.6667}$ = 1.4167
Varieties	h-1 4-1 = 3	SSV = 42	MSV = SSV/h-1 = 42/3 = 14	$F = \frac{MSV}{MSE} = \frac{14}{22.6667}$ = 1.619
Error	(K-1)(h-1) (2)(3) = 6	SSE = 136	MSE = SSE/(K-1)(h-1) = 136/6 = 22.6667	
Total	hK-1 = 12-1 = 11			

Here numerator and denominator are reversed because it should be → N > D

Since the F ratios are < 1

Accept H_0

Since Num and Deno are reversed.

look up for → Num
F at (6, 2) → Deno
F at (6, 3) in table values.

Cal val < Tab val

$1.4167 < 19.33$ Accept H_{01}

Cal val < Tab Val

$1.619 < 8.94$ Accept H_{02}

→ there is no significant difference between the seasons.

→ there is no significant difference between the performance of salesmen.

Problem:

The following table represents the sales (Rs 1000) per month of 3 brands of soaps allocated among three cities.

	cities		
Brands	A	B	C
I	12	48	30
II	42	54	57
III	9	42	21

Test whether ① the mean sales of the three brands are equal.

② the mean sales of soaps in each city are equal.

Sol $H_{01}: \mu_1 = \mu_2 = \mu_3$ $\quad H_{02}: \mu_1' = \mu_2' = \mu_3'$

$H_{11}: \mu_1 \neq \mu_2 \neq \mu_3$ $\quad H_{12}: \mu_1' \neq \mu_2' \neq \mu_3'$

Coding method: Subtract each number from 42

	A	B	C	$T_{i.}$	$T_{i.}^2$	n_i
I	30	-6	12	36	1296	3
II	0	-12	-15	-27	729	3
III	33	0	21	54	2916	3
$T_{.j}$	63	-18	+18	$G = 63$	$\Sigma T_{i.}^2 = 4941$	$N = 9$
$T_{.j}^2$	3969	324	324	$\Sigma T_{.j}^2 = 4617$		

Row $\rightarrow k = 3$ $\quad N = 9$

Columns $\rightarrow h = 3$ $\quad G = 63$

$$RSS = (30)^2 + (-6)^2 + (12)^2 + 0^2 + (-12)^2 + (-15)^2 + (33)^2 + 0^2 + (21)^2$$

$$= 2979$$

$$CF = \frac{G^2}{N} = \frac{(63)^2}{9} = 441$$

$$SST = \frac{1}{h}\Sigma T_{i.}^2 - CF = \frac{1}{3}(4941) - 441 = 1206$$

$$SSV = \frac{1}{k}\Sigma T_{.j}^2 - CF = \frac{1}{3}(4617) - 441 = 1098$$

$$TSS = RSS - CF$$

$$= 2979 - 441 = 2538$$

$$SSE = TSS - SST - SSV$$
$$= 2538 - 1206 - 1098$$
$$= 234$$

ANOVA II way Table:

Source of variation	df	Sum of Squares	Mean Sum of Squares	Variance ratio (F)
Treatments	$k-1$ $3-1=2$	$SST = 1206$	$MST = SST/k-1$ $= 1206/2 = 603$	$F = \frac{MST}{MSE} = \frac{603}{58.5}$ $= 10.3077$
varieties	$h-1$ $3-1=2$	$SSV = 1098$	$MSV = SSV/h-1$ $= 1098/2 = 549$	$F = \frac{MSV}{MSE} = \frac{549}{58.5}$ $= 9.3846$
Error	$(k-1)(h-1)$ $(2)(2) = 4$	$SSE = 234$	$MSE = SSE/(k-1)(h-1)$ $= 234/4$ $= 58.5$	
Total	$hk-1$ $(3)(3)-1=8$			

→ Table value of F at (2,4) df is at 5% level of significance is 6.94

Cal val > Tab val
10.3077 > 6.94
Reject H_{01}

Cal val > Tab val
9.3846 > 9.3846
Reject H_{02}

Inference: ① the mean sales of three brands significantly differ.

② the mean sales of soaps in each city are not equal.

Problem:

A trucking company wishes to test the average life of each of the four brands of tyres. The company uses all brands on randomly selected trucks. The records showing the lifes (1000 of miles) of tyres are given in the table.

Test the hypothesis that the average life of each brand of tyres is the same. Assume $\alpha = 0.01$ (l.l. table value → level of significance)

Brand I	II	III	IV
20	19	21	15
23	15	19	17
18	17	20	16
17	20	17	18
–	16	16	–

Sol: $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$

→ ANOVA I way method is used

Brand I	20	23	18	17	–
Brand II	19	15	17	20	16
Brand III	21	19	20	17	16
Brand IV	15	17	16	18	–

Coding Method: Subtract each number from 17

						T_i	Σx_{ij}^2	n_i
I	-3	-6	-1	0	-	-10	46	4
II	-2	2	0	-3	1	-2	18	5
III	-4	-2	-3	0	1	-8	30	5
IV	2	0	1	-1	-	2	6	4
						$G=-18$	RSS = 100	$N=18$

Here $N=18$, $G=-18$, $RSS=100$

$$CF = \frac{G^2}{N} = \frac{(-18)^2}{18} = 18$$

$$TSS = RSS - CF = 100 - 18 = 82$$

$$SST = \sum_{i=1}^{4} \left(\frac{T_i^2}{n_i}\right) - CF$$

$$= \left[\frac{(10)^2}{4} + \frac{(-2)^2}{5} + \frac{(-8)^2}{5} + \frac{(2)^2}{4}\right] - 18$$

$$= 39.6 - 18$$

$$= 21.6$$

$$SSE = TSS - SST$$

$$= 82 - 21.6$$

$$= 60.4$$

ANOVA I way Table:

Source of Variation	df	Sum of Squares	Mean Sum of Squares	Variance ratio (F)
Treatment	K-1 4-1 = 3	SST = 21.6	MST = $\frac{SST}{K-1}$ = $\frac{21.6}{3}$ = 7.2	F = $\frac{MST}{MSE}$ = $\frac{7.2}{4.3143}$ = 1.6689
Error	N-K 18-4 = 14	SSE = 60.4	MSE = SSE/N-K = 60.4/14 = 4.3143	
Total	N-1 18-1 = 17	TSS = &2		

Table value of F at 1% level of significance at (3, 14) df is 5.56

Cal Val < Tab Val

1.6689 < 5.56

Accept H_0

Inference:

Avg life of each brand of tyres is the same.

Estimation:

The theory of statistical inference can be divided into two major areas

1. Estimation of parameters.
2. Testing of Hypothesis.

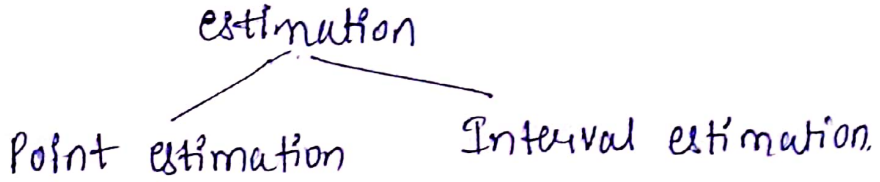
Parameter: A statistical constant derived from the population values is called 'parameter'

ex: population mean, μ

population standard deviation, σ

Statistic: A statistical constant derived from the sample values is called a 'statistic'

Estimation: It is a procedure of estimating a population parameter by using sample information or observations.



Point Estimate: A point estimate of some population parameter θ is a single numerical value.

Point Estimator: A point estimator is a statistic for estimating a population parameter θ and is denoted by $\hat{\theta}$.

Properties of estimator:

An estimator is said to be a good estimator if it is

- unbiased
- consistent
- Efficient
- Sufficient

Unbiased Estimator: A statistic $\hat{\theta}$ is said to be an unbiased estimator or its value an unbiased estimate if and only if the mean of the sampling distribution of the estimator equals θ i.e

$$\text{Mean of } \hat{\theta} = E(\hat{\theta}) = \theta$$

→ If the estimator is not unbiased, the difference $E(\hat{\theta}) - \theta$ is called the bias of the estimate $\hat{\theta}$.

When the estimator is unbiased,

$$E(\hat{\theta}) - \theta = 0$$

i.e the bias is zero.

positive bias : If $E(\hat{\theta}) > \theta$, then $\hat{\theta}$ is called positively biased.

Negative bias : If $E(\hat{\theta}) < \theta$, then $\hat{\theta}$ is called negatively biased.

Problem:

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators of θ, then $w_1\hat{\theta}_1 + w_2\hat{\theta}_2$ is an unbiased estimator of θ. when $w_1 + w_2 = 1$

Sol: Given that $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimator of θ,

i.e $E(\hat{\theta}_1) = \theta$ — ①

$E(\hat{\theta}_2) = \theta$ — ②

Now $E(w_1\hat{\theta}_1 + w_2\hat{\theta}_2) = w_1 E(\hat{\theta}_1) + w_2 E(\hat{\theta}_2)$

$= w_1\theta + w_2\theta$ (from ① & ②)

$= (w_1 + w_2)\theta$

$= \theta$ ($\because w_1 + w_2 = 1$)

$\therefore w_1\hat{\theta}_1 + w_2\hat{\theta}_2$ is an unbiased estimator of θ.

Theoran: Let $x_1, x_2, \ldots x_n$ be a random sample drawn from a given population with the mean μ and variance σ^2, show that the sample mean

is an unbiased estimator of population mean μ
i.e $E(\bar{x}) = \mu$

Proof: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

taking expectations on both sides

$$E(\bar{x}) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)$$

$$= \frac{1}{n} E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n} E[x_1 + x_2 + \cdots + x_n]$$

$$= \frac{1}{n}\{E(x_1) + E(x_2) + \cdots + E(x_n)\}$$

$$= \frac{1}{n}[\mu + \mu + \cdots \mu] \quad (\because E(x_i) = \mu)$$

$$= \frac{1}{n}[n\mu]$$

$$= \mu$$

Theorem:

For a random sample of size n, $x_1, x_2 \ldots x_n$ taken from a finite population $s^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$ is not an unbiased estimator of σ^2 but $s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$ is an unbiased estimator of σ^2

* MVUE (Minimum Variance Unbiased Estimate) or most efficient unbiased estimator:

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators

of θ and if σ_1^2 and σ_2^2 are variances of their sampling distributions and $\sigma_1^2 < \sigma_2^2$, $\hat{\theta}_1$ is said to be more unbiased estimator of θ.

Note: If $\hat{\theta}_1, \hat{\theta}_2, \ldots \hat{\theta}_k$ are unbiased estimators of the parameter θ, $\sigma_1^2, \sigma_2^2, \ldots \sigma_k^2$ are variances of their sampling distributions, then $\hat{\theta}_1$ is the most efficient estimator of θ if $\sigma_1^2 < \sigma_2^2 < \ldots \sigma_k^2$

e = efficiency of an estimator

$$= \frac{\text{var } \theta_1}{\text{var } \theta_i} \quad \text{where } i = 2, 3, \ldots k$$

$$e < 1$$

- Consistency: An estimator $T_n = T(x_1, x_2, \ldots x_n)$ based on a random sample of size n, is said to be a consistent estimator of $\gamma(\theta)$, $\theta \in \Theta$ → the parameter space, if T_n converges to $\gamma(\theta)$ in probability

i.e if $T_n \xrightarrow[\text{converges}]{P} \gamma(\theta)$ as $n \to \infty$

In other words, T_n is a consistent estimator of $\gamma(\theta)$ if for every $\varepsilon > 0$, $\eta > 0$, there exists a positive integer $n \geq m(\varepsilon, \eta)$ such that

$$P\{|T_n - \gamma(\theta)| < \varepsilon\} \to 1 \text{ as } n \to \infty$$

$$\Rightarrow P\{|T_n - \gamma(\theta)| < \varepsilon\} > 1 - \eta \quad \forall\ n \geq m$$

where m is some very large value of n

. Sufficiency: An estimator is said to be sufficient for a parameter, if it contains all the information in the sample regarding the parameter.

If $T = t(x_1, x_2 \ldots x_n)$ is an estimator of a parameter θ based on a sample - $x_1, x_2 \ldots x_n$ of size n from the population with density $f(x, \theta)$ such that the conditional distribution of $x_1, x_2 \ldots x_n$ given T, is independent of θ, then T is sufficient estimator for θ

Methods of estimation:

1. Maximum likelihood method
2. Method of moments
3. Method of minimum chi-square
4. Method of minimum variance
5. Method of least squares
6. Method of Inverse Probability.

Maximum likelihood Estimation: It was formulated by C.F. Gauss but as a general method of estimation was first introduced by Professor R.A. Fisher

Likelihood function: Let $x_1, x_2 \cdots x_n$ be a random sample of size n from a population with density function $f(x, \theta)$. Then the likelihood function of the sample values $x_1, x_2 \cdots x_n$ usually denoted by $L = L(\theta)$ is their joint density function.

$$L = f(x_1, \theta)\, f(x_2, \theta) \cdots f(x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

L gives the relative likelihood that the random variables assume a particular set of values $x_1, x_2, \cdots x_n$. For a given sample $x_1, x_2 \cdots x_n$ L becomes a function of the variable θ, the parameter.

The principle of maximum likelihood estimation in finding an estimator for the unknown parameter $\theta = (\theta_1, \theta_2, \cdots \theta_k)$ which maximises the likelihood function $L(\theta)$ for variations in parameter.

If there exists a function $\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots x_n)$ of the sample values which maximises L for variations in θ, then $\hat{\theta}$ is to be taken as an estimator of θ. $\hat{\theta}$ is usually called maximum likelihood estimator.

Thus $\hat{\theta}$ is the solution of

$$\frac{\partial L}{\partial \theta} = 0 \quad \text{and} \quad \frac{\partial^2 L}{\partial \theta^2} < 0 \quad -②$$

L ①

Since $L > 0$ and $\log L$ is non decreasing function of L, L and $\log L$ attain their extreme values (maxima or minima) at the same value of $\hat{\theta}$

① can be rewritten as

$$\frac{1}{L}\frac{\partial L}{\partial \theta} = 0 \Rightarrow \frac{\partial \log L}{\partial \theta} = 0$$

this is a convenient form from practical view. If θ is a vector valued parameter, then $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3 \cdots \hat{\theta}_k)$ is given by the solution of simultaneous equations

$$\frac{\partial}{\partial \theta_i} \log L = \frac{\partial}{\partial \theta_i} \log L(\theta_1, \theta_2 \cdots \theta_k) = 0 \quad -③$$

$$i = 1, 2 \cdots k$$

eqns ①, ② & ③ are usually referred to as likelihood eqns for estimating the parameters

Date: 12/2020

UNIT 2 : Testing of Hypothesis

. Hypothesis is a statement regarding the population parameter. Hypothesis is of two types

1. Null Hypothesis
2. Alternative Hypothesis

Null Hypothesis: It is a no difference statement. It is denoted by H_0.

ex: $H_0 : \mu = 3$

$H_0 : \mu_1 = \mu_2$

$H_0 : \sigma^2 = 2$

Alternative Hypothesis: It is complimentary to Null Hypothesis. It is denoted by H_1. It is of two types. ex: $\mu_1 \neq \mu_2$; $\sigma_1 \neq \sigma_2$; $\mu > 2 \Rightarrow H_1$

1. Two Tailed Alternative
2. Single Tailed alternative

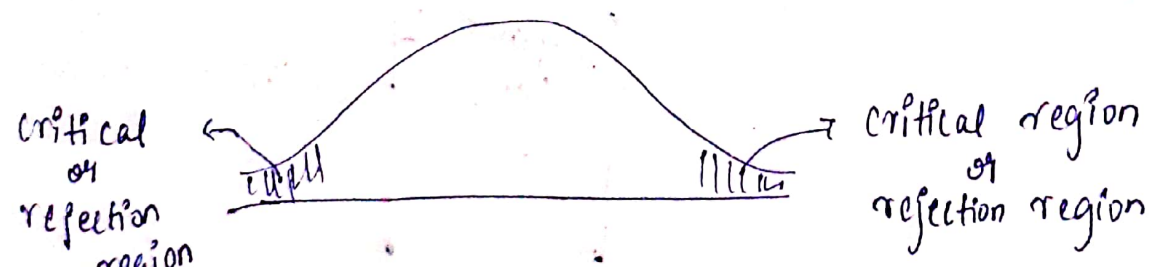
Two Tailed Alternative:

Ex: $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Ex: $H_0 : \sigma^2 = 30$

$H_1 : \sigma^2 \neq 30$



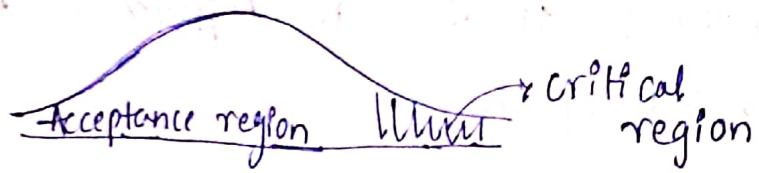
Single Tailed Alternative:

It is divided into two types. They are

1. Right Tailed Alternative

ex: $H_0: \mu_1 = \mu_2$ Ex: $H_0: \sigma^2 = 30$

$H_1: \mu_1 > \mu_2$ $H_1: \sigma^2 > 30$

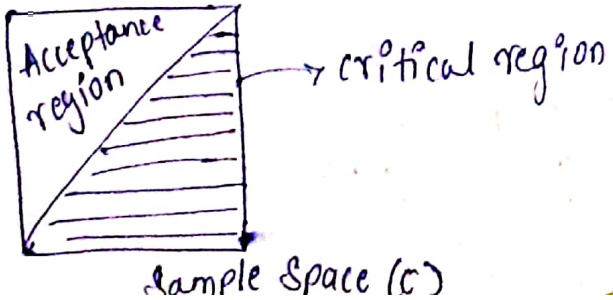


2. left Tailed Alternative

ex: $H_0: \mu_1 = \mu_2$ $H_0: \sigma^2 = 30$

$H_1: \mu_1 < \mu_2$ $H_1: \sigma^2 < 30$

- Critical region: A region in the sample space 'S' which amounts to rejection of H_0 is called the critical region or the rejection region. It is denoted by w.



. Level of Significance: If w is the critical region and $t = t(x_1, x_2, x_3 \cdots x_n)$ is the value of the statistic based on random sample of size 'n', then

$$P(t \in w / H_0) = \alpha$$

Here α is the ~~critical region~~ level of significance and

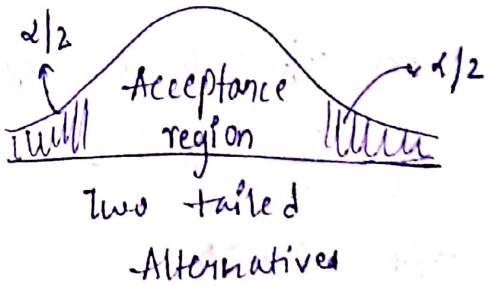
$$P(t \in \bar{w} / H_1) = \beta$$

$$w \cup \bar{w} = S$$

$$w \cap \bar{w} = \phi$$

· Critical value or significant value: The value of the test statistic which separates the critical region and the acceptance region is called critical value or significant value. It depends upon:

1. Level of significance
2. Alternative Hypothesis



Two tailed Alternatives

* Types of Errors: there are two types of errors

1. Type I error
2. Type II error

1. Type I error: Probability of rejecting a good lot is called Type I error.

P(rejecting a good lot) = α

This is called producers risk

2. Type II error: Probability of accepting a bad lot is Type II error.

P(accepting a bad lot) = β

This is called consumers risk

$1-\beta$ is called Power of the test.

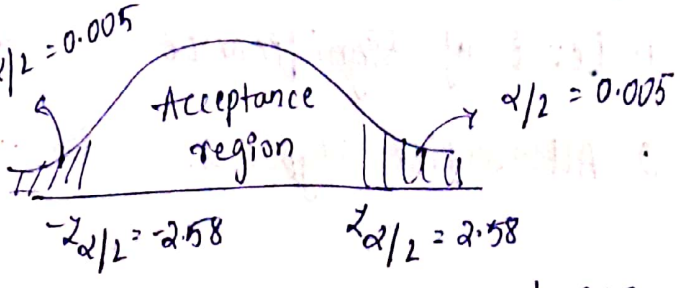
	1%	5%	10%
Two Tailed test	$Z_{\alpha/2} = 2.58$	$Z_{\alpha/2} = 1.96$	$Z_{\alpha/2} = 1.65$
Right Tailed test	$Z_{\alpha} = 2.33$	$Z_{\alpha} = 1.65$	$Z_{\alpha} = 1.29$
Left Tailed test	$Z_{\alpha} = -2.33$	$Z_{\alpha} = -1.65$	$Z_{\alpha} = -1.29$

- Two tailed 1% :

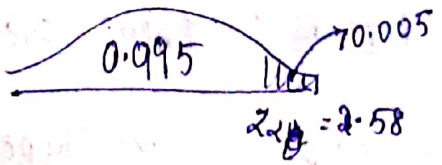
$\alpha = 1\% = 0.01$

$\frac{\alpha}{2} = \frac{0.01}{2} = 0.005$

$\alpha/2 = 0.005$



1.000
0.005
0.995



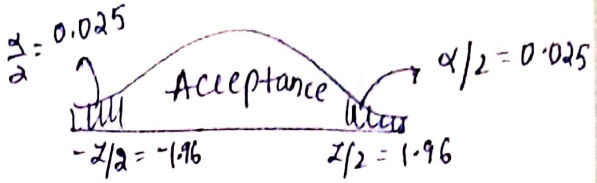
Normal table
Horizontal 2.5
Vertical 0.08
2.58

1% two tailed $|Z_{\alpha/2}| = 2.58$

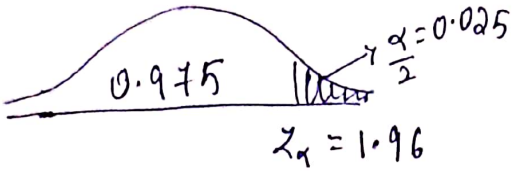
• 5·/· Two Tailed:

$\alpha = 0.05 = 5\%$

$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$



Total area under Normal curve = 1



From $-\infty$ upto Z_α area is

$1.000 - 0.025 = 0.975$

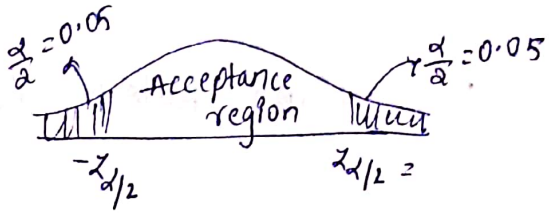
So two tailed test 5·/· los,

$|Z_{\alpha/2}| = 1.96$

• 10·/· Two Tailed:

$\alpha = 10\% = \frac{10}{100} = 0.1$

$\frac{\alpha}{2} = \frac{0.1}{2} = 0.05$

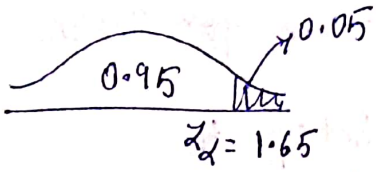


Nrml table

1.6 – horizontal

0.05 – Vertical

1.65



$1.00 - 0.05 = 0.95$

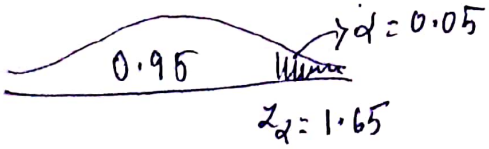
10·/· two tailed

$|Z_{\alpha/2}| = 1.65$

• 5·/· right tailed:

$\alpha = 5\% = 0.05$

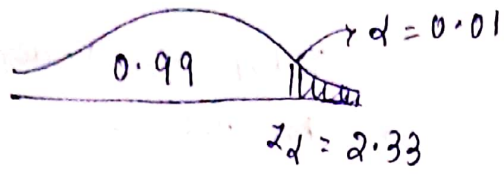
$1.00 - 0.05 = 0.95$



• 5·/· right tailed $Z_\alpha = 1.65$

1% right tailed:

$\alpha = 1\% = 0.01$

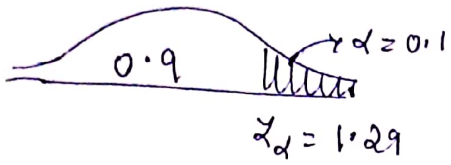


$1.00 - 0.01 = 0.99$

1% right tailed, $Z_\alpha = 2.33$

10% right tailed:

$\alpha = \frac{10}{100} = 0.1$



$1.0 - 0.1 = 0.9$

1.2 hori + 0.09 vert = 1.29

10% right tailed, $Z_\alpha = 1.29$

Procedure for Testing of Hypothesis:

1. Setup H_0
2. Setup H_1
3. Choose the level of significance, α
4. Calculate the test statistic under H_0.
5. Compare the calculated value of the test statistic with table value.

For two tailed and right tailed test

Calculated value < Table value, Accept H_0

Calculated value > Table value, Reject H_0

Left Tailed test

Calculated value < Table value, Reject H_0

Calculated value > Table value, Accept H_0