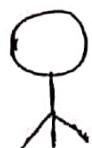


Statistics

collection, organising, collecting the data (statistical operation)

Example



$\mu_E = 85 < \text{English} - 88/100$

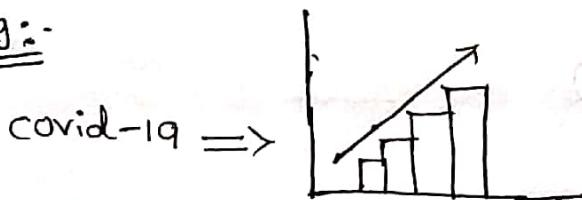
$M_m = 98 > \text{maths} - 95/100$

Types

Descriptive Statistics

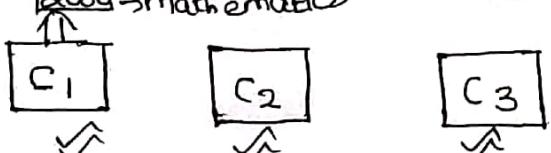
→ collecting, organising and representing the data in a better way

Eg :-



Inferential Statistics

① Eg: People



A hand-drawn diagram consisting of a rectangular box containing the letters 'C4'. Three arrows point from this box to the words 'NMR', 'norm', and 'maths' arranged below it.

② Population

True population

Sample

Doing calculation by collecting data from various groups.

Various groups.



with - 18 to 25 H

Variables in Statistics

Based on Descriptive Stats

Categorical

* Rating \rightarrow 1, 2, 3, 4, 5 [Eg:- Zomato Delivery]

* Decision \rightarrow yes or no [Switching towards Data Science \rightarrow yes or no]

Numerical

* Eg:- 800.91kg (Weight)



Level of measurement

Datatype

Qualitative

Variable determines the quality of the data

(i) Nominal :-

Categorize the data based on the name

Eg:- Season:- winter, Summer like that

1, 2, 3, 4, 5

ordinal :- (i)

OK, good, bad, Excellent, Amazing (order pair)

Quantitative

Describe the quantity of the dataset

(i) Discrete (It ~~doesn't~~ give the absolute number)

Eg:- 1, 2, 3, 4, 5

(ii) Continuous

(It doesn't give the absolute number)

Eg:- 75.5234110 (Weight)

① mean

Sample

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

Population

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

Eg :-

① 3, 9, 5, 10, 10, 1, 7, 8, 5, 5

$$\frac{3+9+5+10+10+1+7+8+5+5}{10}$$

$$= 6.30$$

② median

Eg :-

1, 5, 5, 5, 7, 8, 9, 10, 10 (Arrange in ascending order)

$$\frac{7+5}{2} = 6$$

③ mode :-

frequency of ^{highest} occurrence of the data

Eg :- (We can take mode upto 2 and 3 only)

3, 9, 5, 10, 10, 1, 7, 8, 5, 5

5 → 3 Occurrence of 5 will be more
10 → 2

Skewness

Eq:

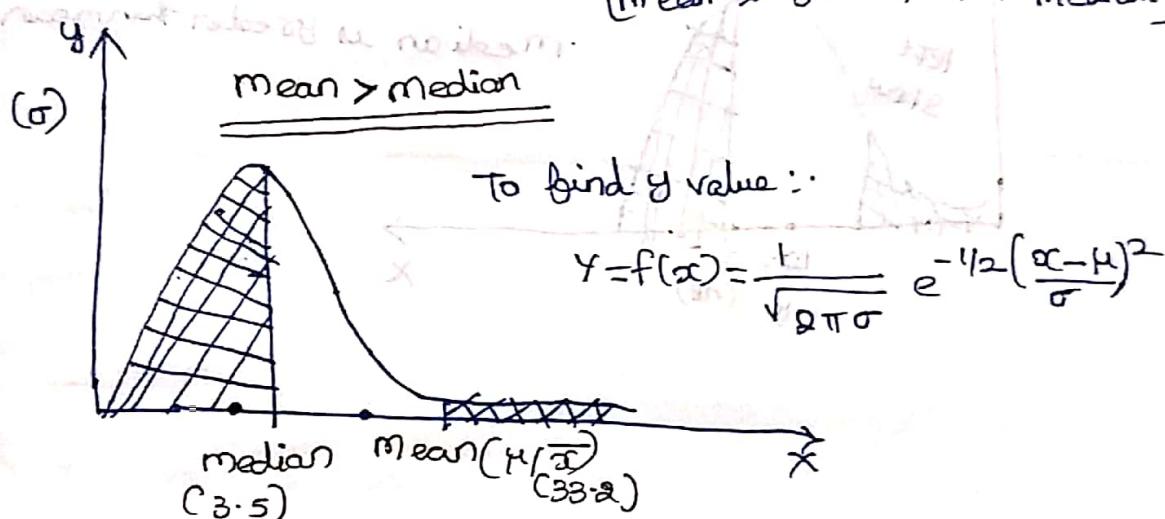
$D = 1, 2, 3, 4, 1, 2, 8, \boxed{100, 101, 110} \rightarrow$ frequency is high

$$\bar{x} \text{ (Sample)} = 33.2 \quad (\text{Outliers in Positive side})$$

$$\text{median} = 3.5$$

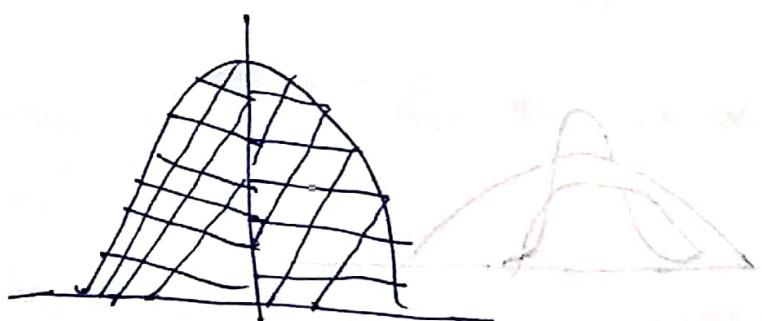
Data in Ascending order: *Arranging numbers*

1, 1, 2, 2, 3, 4, 8, 100, 101, 110 [Mean is greater than medium]



~~Positive Skewness~~ \rightarrow ~~positive~~ \rightarrow ~~positive~~

Normal Distribution



2000 वर्ष की अवधि में इन्होंने एक नियमित विद्युत उपलब्धि का बनाया।

W. J. Englehardt

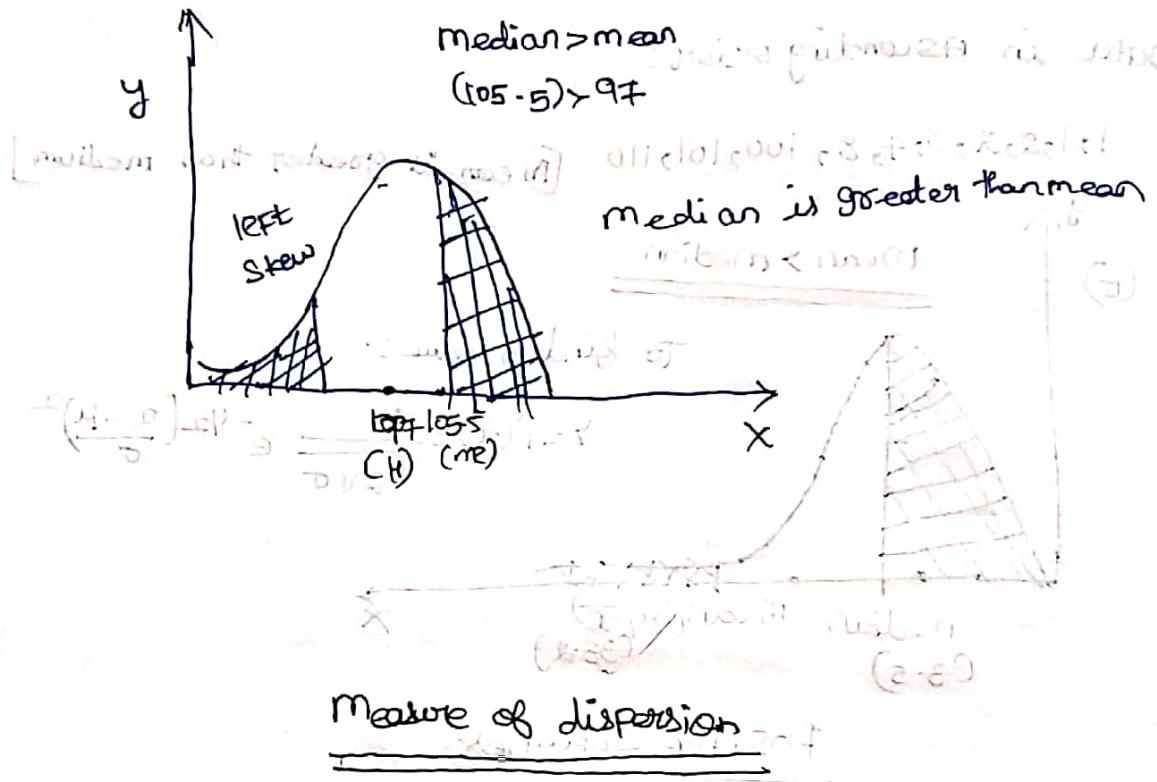
১৭। কৃষ্ণ হাতা পদ্মা পদ্মনাভ

Negative skewness [Outliers in negative side]

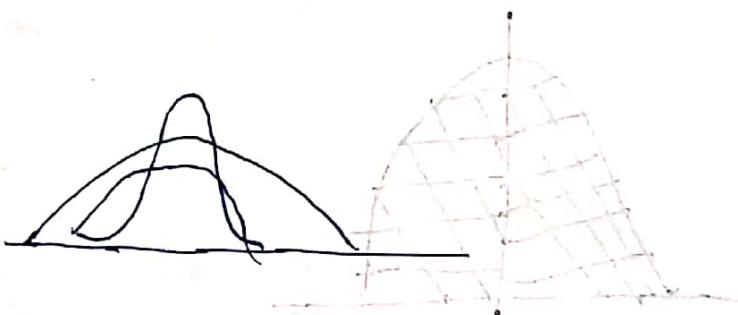
$D = 2, 3, 4, 100, 101, 110, 120, 150, 160, 140, 110$

$$\bar{x} = 97 \text{ (mean)}$$

$$\text{median} = 105.5$$



① Range ::



Eg::

Laptop price varies from 60k to 90k

1, 3, 100, 110, 101, 2, 8, 100

$$= 109$$

↓
dispersion

(Summation of data from mean)

② Variance or Standard deviation

Population Dataset:

(Variation of data from the actual

-mean)

$$(4, 5, 10, 8, 2, 10) \rightarrow 100, 110$$

$$\hookrightarrow \bar{x} = 6.5$$

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

$$= (4-6.5)^2 + (5-6.5)^2 + (10-6.5)^2 + (8-6.5)^2 + (2-6.5)^2$$

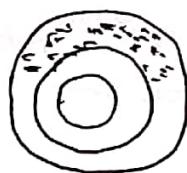
$$+ (10-6.5)^2$$

$$6$$

$$\sigma^2 = 11.1$$

* If variance is high then this will happen,

Eg:-

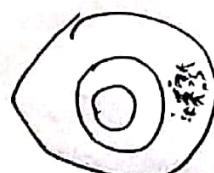


[Driver is not steady, he will apply many brakes]

$$[0 - 140]$$

* If variance is less then this will happen

Eg:-



$$[1, 1, 1, 1] \rightarrow \text{no distance}$$

opposite

Variance:
Standard deviation (It is used to check the fluctuation)

standard deviation
of the data)

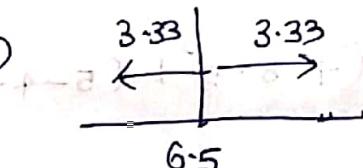
$$\sigma = \sqrt{\left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N} \right)}$$

4, 5, 10, 8, 2, 10

$$\mu = 6.5$$

$$\sigma^2 = 11.1 \text{ (variance)}$$

$$\sigma = 3.33 \text{ (standard deviation)}$$



Eg.:

4, 5, 10, 8, 2, 10

$$6.5 \pm 3.33$$

$$(i) 6.5 + 3.33 = 9.83$$

$$(ii) 6.5 - 3.33 = 3.17$$

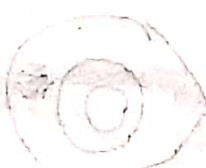
For $3.17 \leq x \leq 9.83$ ^{for} _{for}

$$\begin{aligned} 1\sigma &\rightarrow 68\% \\ 2\sigma &\rightarrow 95\% \\ 3\sigma &\rightarrow 99\% \end{aligned}$$

Normal distribution



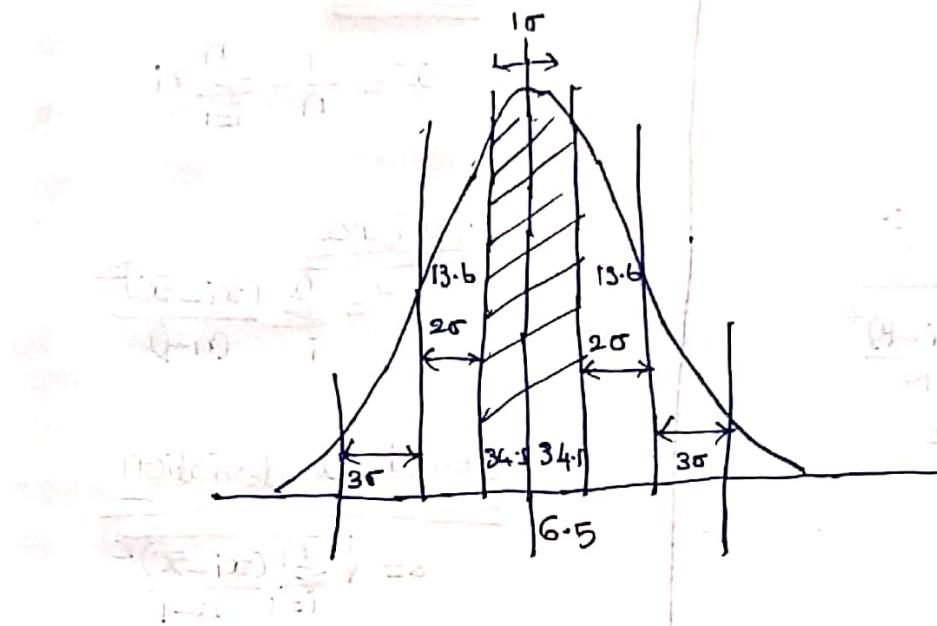
Normal distribution



working of standard deviation

Diagram

$$\sigma = 4, 5, 10, 8, 10$$



1st case:-

$$6.5 \pm 3.33 (\mu \pm \sigma)$$

68.3.

2nd case:-

$$6.5 \pm 2 \times 3.33 (\mu \pm 2\sigma)$$

95.4.

3rd case:-

$$6.5 \pm 3 \times 3.33 (\mu \pm 3\sigma)$$

99.7.

Population

Mean :-

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

Standard deviation :-

$$\sigma^2 = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

Covariance

$$S.d = \sqrt{\sigma^2}$$

Samples

mean :-

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

Variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Standard deviation

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Probability Density Function

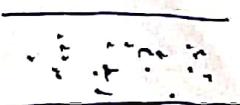
Probability of occurrence of the particular outcome

Binomial Distribution

Eg :-

(i) 1, 2, 3, 4, 5, 6 (1-Dice)

(ii) H-T (Coin)

(iii) 

{ EMS
Fire }

(How many vehicle
that we can deploy)

PDF

- * Normal Distribution
- * Bernoulli Distribution
- * Binomial Distribution
- * Uniform Distribution
- * Student T Distribution
- * Poisson Distribution

Binomial Distribution

formula ::

$$P(n/r) = \binom{N}{n} p^n q^{N-n}$$

$$= \left(\frac{N!}{n!(N-n)!} \right) p^n (1-p)^{N-n}$$

$n \rightarrow$ Successes

$N \rightarrow$ trial

Eg \uparrow can col

① 100 - 75 die (Population)

(6 person) \rightarrow (4 recover) (sample),

$$N = 6 (+\text{trial})$$

$$n = 4 (\text{success})$$

$$P(\text{recovery}) = 0.25$$

$$q = 1 - P(\text{failure}) = 0.75$$

$$= \frac{6!}{4!(6-4)!} (0.25)^4 (0.75)^{6-4}$$

$$= \frac{6!}{4! \cdot 4!} (0.25)^4 (0.75)^2$$

$$= 0.632959$$

② Die = 8 times

$N = \text{Total}$

$n = \text{Success}$

(i) P_5 (Probability of not getting any five)

(ii) P_5 (Probability of getting only five)

(iii) P_5 (Probability of getting five all the time)

(i)

$N = 3$

$n = 0$

$P_5(1/6)$ (Success)

$P_5^1 = 1 - \frac{1}{6} (5/6)$ (Failure)

$P(n=0) = C_n^N p^n q^{N-n}$

$\approx 3C_0 (1/6)^0 (5/6)^3$

$= 0.57$

$P(n=0) = 3C_1 (1/6)^1 (5/6)^2$

$N=3$ (Total)

$n=1$ (Success)

$= 3C_1 (1/6) \times (5/6)^2$

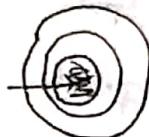
$= 0.3477$

$$P(n=3) = 3 C_3 (1/6)^3 (5/6)^{5-3}$$

$$= 4.62 \times 10^{-3} \quad N = 3 (\text{Total})$$

$$n = 3 (\text{Success})$$

$$= 0.00462$$



i) Probability of getting (hit) more than 2 times when he hit 4 times

$$\rightarrow n = 4 (\text{Total}) (4)$$

$$n = 3 (\text{Success}) (3)$$

$$N = 5 (\text{Total})$$

$$\text{Probability of getting Success (P_S)} = \frac{4}{5}$$

$$= 0.8$$

$$\text{Probability of getting failure (P_F)} = 1 - \frac{4}{5}$$

$$= \frac{5-4}{5}$$

$$= \frac{1}{5}$$

$$= 0.2$$

$$P(T) = P(3) + P(4)$$

$$= \left(\frac{4!}{3! (4-3)!} \right) (0.8)^3 (0.2)^{4-3} + \left(\frac{4!}{4! (4-4)!} \right) (0.8)^4 (0.2)^{4-4}$$

$$= 0.8192$$

(ii) Probability of getting atleast 3 times going to fail

(Failure) 3 miss 1 hit (Success) (i)

(Failure) 4 miss 0 hit (Success) (ii)



$$(i) P(T) = P(T_3) + P(T_4)$$

$$(P_{T_3}) = {}_N C^n p^n q^{N-n}$$

$N \rightarrow$ trial (4)

$n \rightarrow$ success (1)

$$(P_{T_3}) = {}_4 C^1 (0.8)^1 (0.2)^3$$

$$= \left(\frac{4!}{1! (4-1)!} \right) (0.8)^1 (0.2)^3$$

(ii)

$$(P_{T_4}) = {}_N C^n p^n q^{N-n}$$

$N \rightarrow$ trial (4)

$n \rightarrow$ success (0)

$$= {}_4 C^0 (0.8)^0 (0.2)^4$$

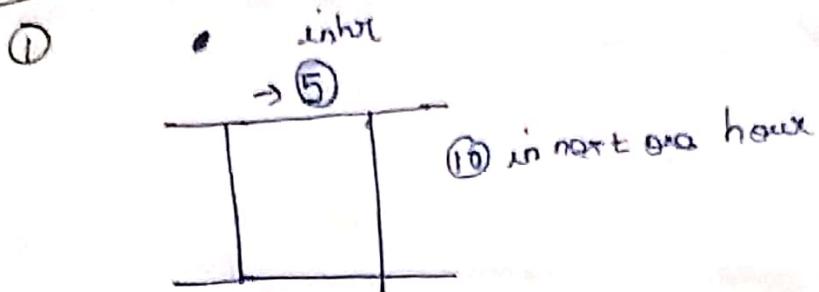
$$= \frac{4!}{0! (4-0)!} (0.8)^0 (0.2)^4$$

$$P(T) = \left(\frac{4!}{1! (4-1)!} \right) (0.8)^1 (0.2)^3 + \left(\frac{4!}{0! (4-0)!} \right) (0.8)^0 (0.2)^4$$

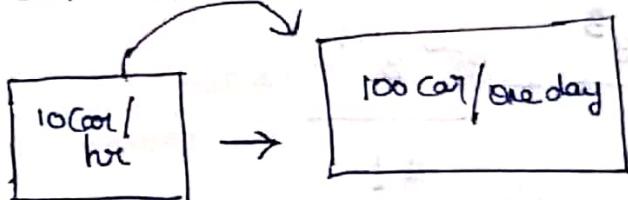
$$P(T) = 0.02$$

Poisson Distribution

eg :-



② In car wash company



Q-A

① Sales man = 3 insurance / week [find the Probability of not getting any insurance]

$$\mu = 3$$

Probability of selling insurance is (1) $[1 - P(x_0)]$ (1 is insured)
Probability of not selling any insurance is (0) $P(x_0)$ (0 is insurance)

$$P(x_0) = \frac{e^{-\mu} \mu^x}{x!}$$

$$P(x_0) = \frac{e^{-3} 3^0}{0!} \\ = 0.0497$$

$$= 1 - P(x_0)$$

$$= 1 - 0.0497$$

$$= 0.9502$$

② Probability of getting 2 or more than 2 insurance
but less than 5 insuree

$$P(2 \leq x \leq 5)$$

$$P(2 \leq x \leq 5)$$

$$= P(2) + P(3) + P(4)$$

$$= \frac{e^{-3} 3^2}{2!} + \frac{e^{-3} 3^3}{3!} + \frac{e^{-3} 3^4}{4!}$$

$$= 0.61$$

③ $100 \rightarrow 1.5$ mistake (average)

100 pages \rightarrow New Book

(i) Probability there won't be any mistake

$$P(0) = \frac{e^{-\mu} \mu^x}{x!} = \frac{e^{-1.5} (1.5)^0}{0!} = 0.223$$

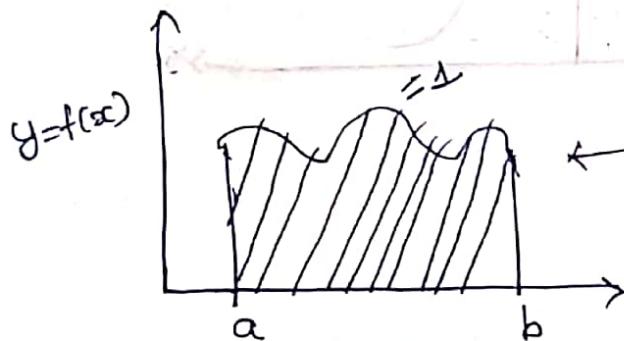
④ 400 (page) and 0 - mistake)

$$P(0) = \frac{e^{-6} (6)^0}{0!}$$

Normal Distribution

Probability Density Function (Here $\mu=?$ and $\sigma=?$)

$$f(x) \geq 0 = y \quad a \leq x \leq b$$



Here (y) value can be chosen
from a and b but it is
positive

Normal distribution

$$\text{Eq.: } \frac{\sqrt{1}}{\sqrt{1}} = \frac{3.33}{\sqrt{10}}$$

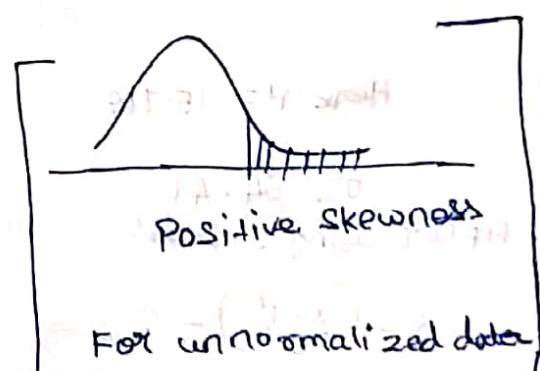
→ Applying Square rooting

Data = 1, 2, 3, 4, 10, 100

\downarrow \downarrow

$\log(1)$ $\log(100)$

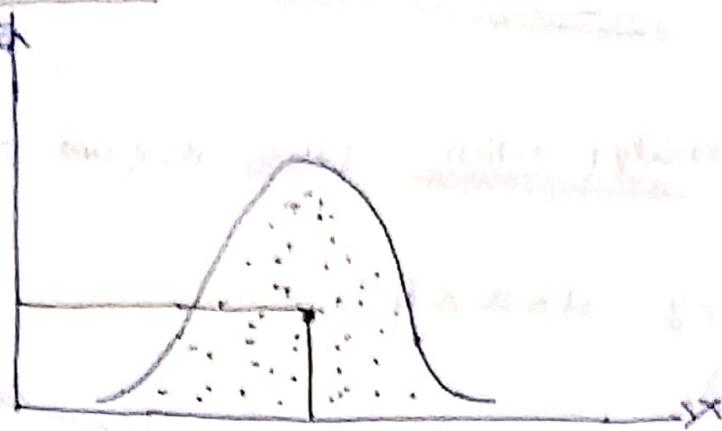
unnormalized data



For unnormalized data

Desired ref
 $f=0$

Normalized data



To find y value:-

$$y = f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty$$

Standard normal distribution

Here $\mu=0$, $\sigma=1$ (Population Data)

Data = 1, 2, 3, 1, 100

Here $\mu = 15.714$

$$\sigma = 34.41$$

After using z-Stats:-

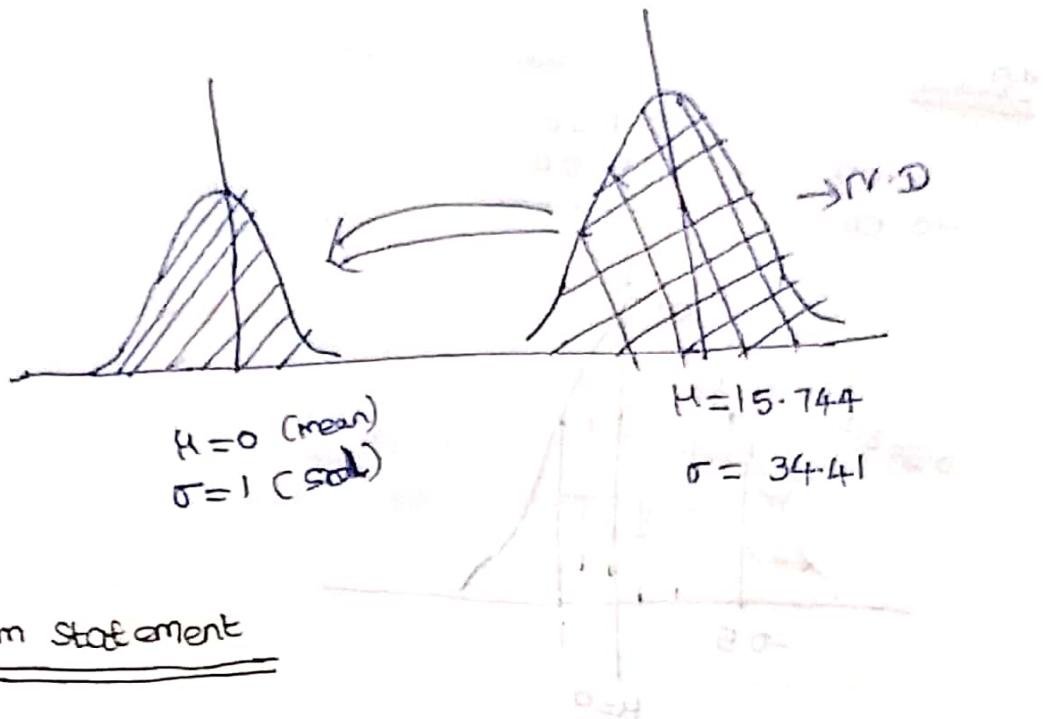
$$z = \frac{(x-\mu)}{\sigma}$$

our Data become

$$-0.42, -3.9, -3.6, -0.42, -0.39, -0.42, 2.44 \rightarrow SD$$

$$\mu = 6.334 e^{-17}$$

$$\sigma = 1$$



Problem Statement

- ① $45 \leq X \leq 60$ What is the probability of completing assignment within [45 to 60 minutes]

$$\begin{aligned} \sigma &= 10 \\ \mu &= 50 \end{aligned} \quad \left. \begin{array}{l} \text{These are normal distribution} \\ \text{Assignment time is 10 minutes} \end{array} \right\}$$

$$P\left(\frac{45-50}{10} \leq \frac{X-\mu}{\sigma} \leq \frac{60-50}{10}\right) \text{ for } N\left(\frac{\mu}{\sigma}, \frac{1}{10}\right)$$

$$Z = \frac{X-\mu}{\sigma} \quad \begin{aligned} X &= 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, \\ &\quad 55, 56, 57, 58, 59, 60 \end{aligned}$$

$$P(-0.5 \leq Z \leq 1)$$

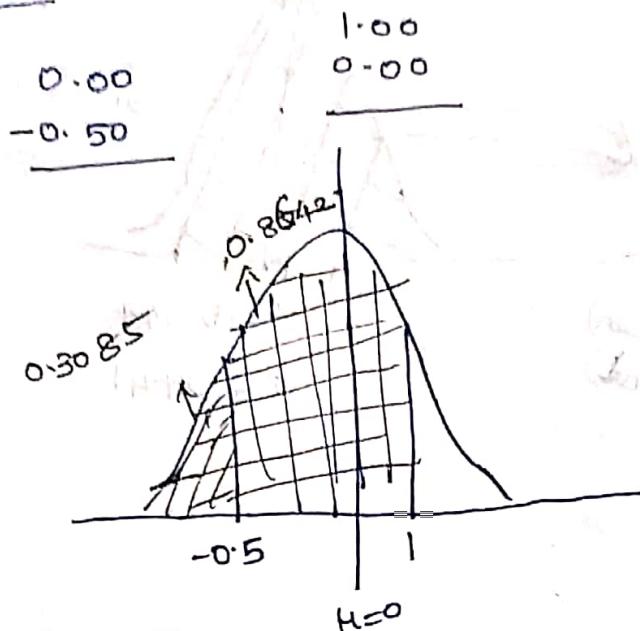
* Here we are converting normal distribution in

to standard normal distribution

$$P(-0.5 \leq Z \leq 1)$$



Eg :-



Probability to fall between -0.5 & 0.6421 is

$$\text{Probability} = 0.8642 - 0.3085$$

$$= 0.5328$$

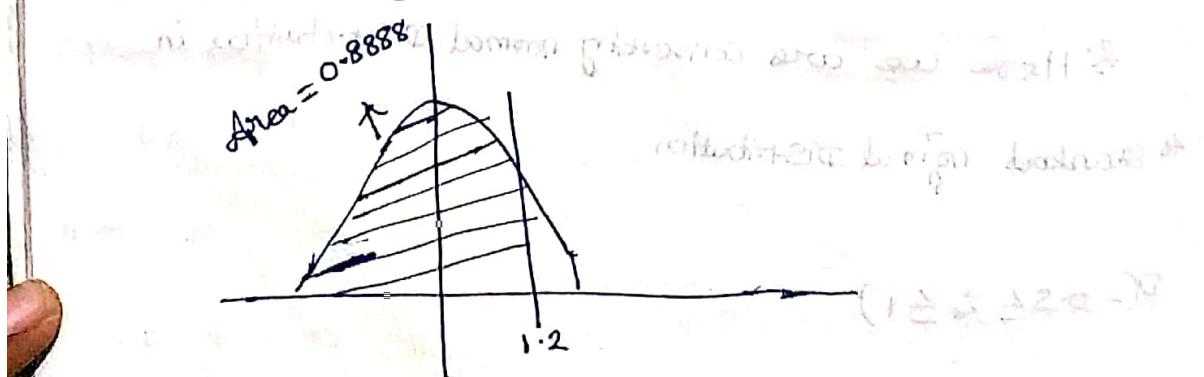
② Z statistic for 1.22 ?

$$Z = \frac{1.22 - 0.5}{0.02} = \frac{0.72}{0.02} = 36 \Rightarrow 0.8888$$

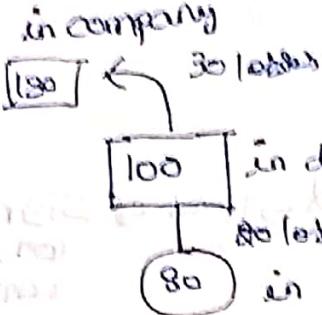
Z table is valid from 0 to 3.4 only

$$\text{Expt } 1.20 \text{ P.P. } \beta + \text{ P.P. } \alpha = 2 \\ \frac{0.02}{1.22} \quad \text{Z value} = 0.8888$$

Area enclosed by the curve



Estimation Problem

Eg:- 

outgoing
letters
arrivals

in demand
100
80 letters

in my company
80

Point estimation

Eg:- complete my task in 10 days

Interval estimation

Eg:- complete my task in 10 - 20 days

Eg:-

101, 120, 120, 130, 150, 160, 110, 120, 115, 170, 118, 120, 119

$$\bar{x} = 126.38$$

$$\sigma = 19.46$$

$$n = 13$$

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

Estimation Formula

ANSWER

(i) For 90% of level of significance

$$\alpha = 1 - 90\%$$

$$\alpha = 0.10$$

$$\alpha/2 = 0.05$$

$$Z_{0.05} = 1.645$$

$$117 \leq E \leq 135$$

↳ approx. value of Z

$\alpha/2 \rightarrow$ level of significance
(or) confidence

approx. value of Z

(ii) For 95% level of significance

$$\alpha = 1 - 95\%$$

$$\alpha = 0.05$$

$$\alpha/2 = 0.025$$

$$Z_{0.025} = 1.96$$

$$116 \leq E \leq 137$$

(iii) For 99% level of significance

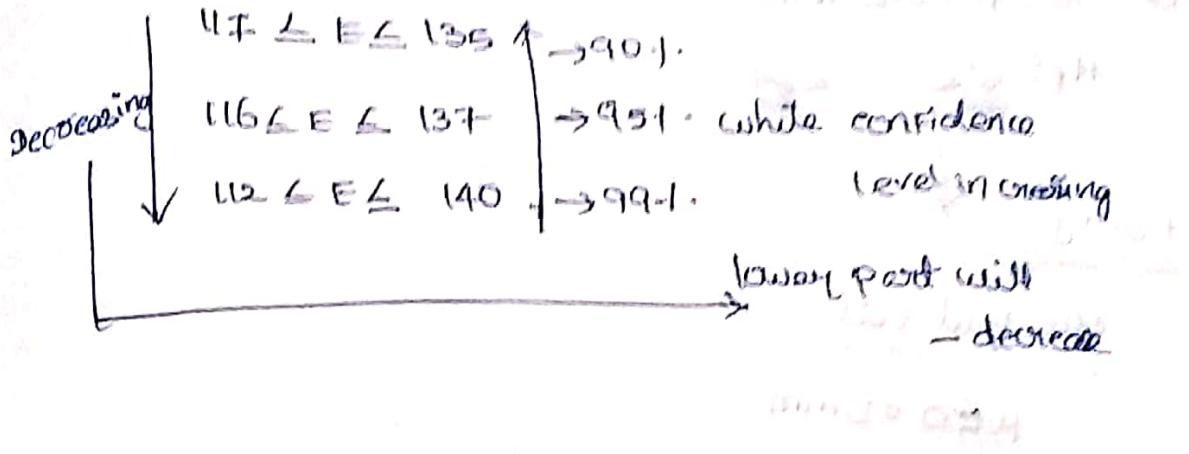
$$\alpha = 1 - 99\%$$

$$\alpha = 0.01$$

$$\alpha/2 = 0.005$$

$$Z_{0.005} = 2.525$$

$$112 \leq E \leq 140$$

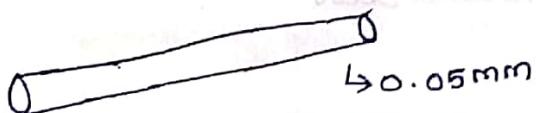


Hypothesis Testing

If we are making an assumption or fact without proving the assumption is true or false

Eg :-

①



Factory

Null hypothesis $\leftarrow H_0: \mu = 0.05$ (Correct Statement)

Alternate hypothesis $\leftarrow H_1: \mu \neq 0.05$ (We are Disapproving)

For eg.: $0.045 < 0.05$ (Left tail)
 $0.051 > 0.05$ (Right tail)

② $H_0: \mu = 0.05 \text{ mm}$

$H_1: \underline{\leq} \quad \underline{\geq}$

For eg.:

one tail test

$\mu \leq 0.05 \text{ mm}$

two tail test

$\mu < 0.05 \text{ mm} \text{ or } \mu \geq 0.05 \text{ mm}$

③ Renting an Apartment

Budget $\rightarrow 15000$

Apartement rate $\rightarrow 7000 - 80000$

$H_0: I \text{ will rent an apartment}$

$H_1: I \text{ won't rent an apartment}$

$H_0: \leq 15k$

$H_1: < 15k \text{ or } \geq 15k$

critical region

$H_0: \mu = 0.05$

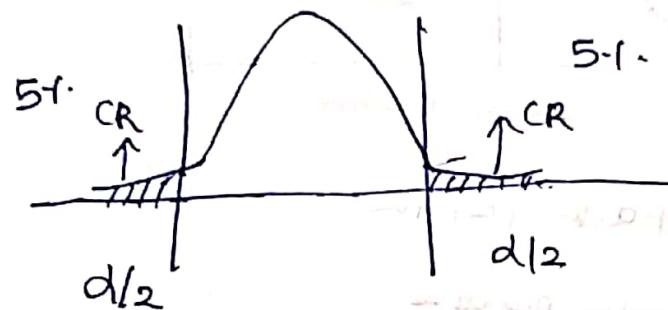
$H_1: \underline{\leq} \quad < \quad < \underline{\geq}$

$$H_0: \mu = 0.05 = 90.1$$

$$H_1: 0.05 \leq \mu \geq 0.05 = 10.1$$

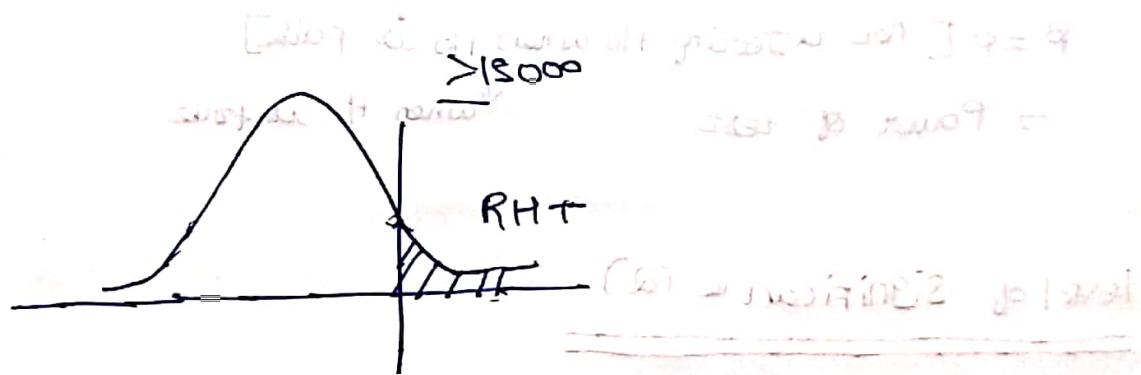
Two tail test

$$\alpha = 10.1 \text{ (critical value)}$$

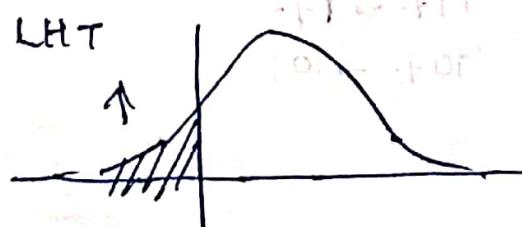


one tail test

Eğer la çok yükseliyorsa $\mu > 10.1$



$$\alpha = 10.1$$



$$< 15000$$

$$\alpha = 10.1$$

Confusion matrix

Decision	Innocent	not innocent
	H ₀ (True)	H ₀ (False)
REJECT H ₀	Type I Error	α
ACCEPT H ₀	β	Type II Error

TYPE I Error - False positive

TYPE II Error - False negative

$$\alpha = P[\text{rejecting } H_0 \text{ when } H_0 \text{ is true}]$$

$$\beta = P[\text{not rejecting } H_0 \text{ when } H_0 \text{ is false}]$$

↳ Power of test \downarrow when H_1 is true

Level of significance (α)

$$\left\{ \begin{array}{l} H_0: \text{clear exam} \quad 95\% \\ H_1: \text{Not clear} \quad 5\% \end{array} \right.$$

e.g. :- $H_0 = C$

$$\begin{aligned} 95\% &\rightarrow 5\% \\ 99\% &\rightarrow 1\% \\ 90\% &\rightarrow 10\% \end{aligned}$$

Pearquisites for Hypothesis testing

- * Null hypothesis & Alternate hypothesis
- * level of significance i.e. $\alpha = 5\%, 1\%$
- * Z test, t-test, chi square test, Anova test.
↓
For one and two tail test

Eg

① Average pointer 500 min

↓
Average

100 pointer Average
↓ ↑
sample (490 ± 30)
↓ S-deviation

$H_0: \text{Buy } 500 \geq$

$H_1: \text{Not Buy } 500 \leq$

$$\text{std} = 30$$

$$\mu = 500$$

$$\bar{x} = 490$$

$$S \cdot \text{std} = 80$$

↓
Sample

$$\alpha = 5\%$$

Sample size = 100



$$SE(\text{Standard Error}) = \left(\frac{s}{\sqrt{n}} \right)$$

$$= \frac{\text{std of Sample}}{\sqrt{\text{sample size}}}$$

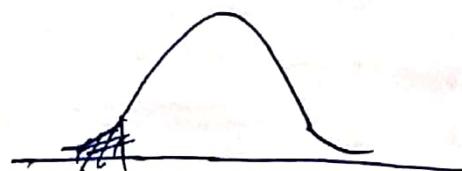
$$\text{Standard error} = \frac{30}{\sqrt{100}} = 3$$

$$z(\text{test}) = \frac{\text{sample mean} - \text{population mean}}{SE}$$

$$= \frac{490 - 500}{3}$$

$$= -\frac{20}{3}$$

$$= -3.33$$



$$\alpha = 5.1.$$

$$z(5.1) =$$

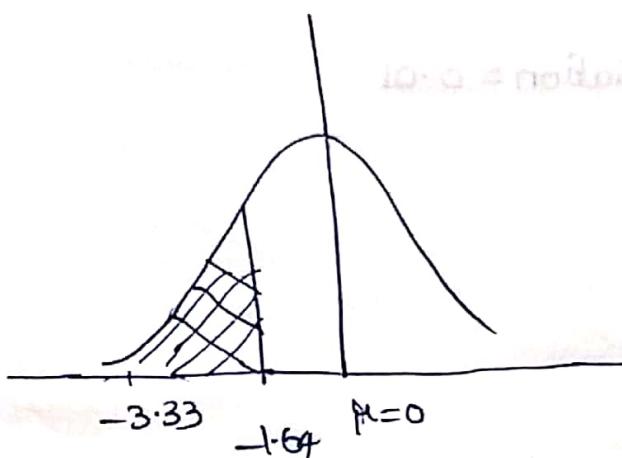
$$\begin{bmatrix} \text{From table} & -1.60 \\ -0.04 & \\ \hline -1.64 & \end{bmatrix}$$

$$z(0.05) = -1.64$$

$$Z(\text{test}) = -3.33$$

$$Z(0.05) = -1.64$$

$$Z(\text{test}) < Z(0.05)$$



IF Z test lies in between

H_0 : Reject

\approx Z table prediction

H_1 : Accept

means

we have to accept (H_1)

Problem Statement

- ① A company used a specific brand has approached the company has an average life of 1000 hours . A new brand has approached the company with new tube lights with same power at a low price . A sample of 120 light bulbs were taken for testing which yielded an average of

1100 hours with standard deviation of 90 hours.
 Should the company give the contract to this
 new company at a 1- α significance level.

$$\mu = 1000$$

$$\bar{x} = 1100$$

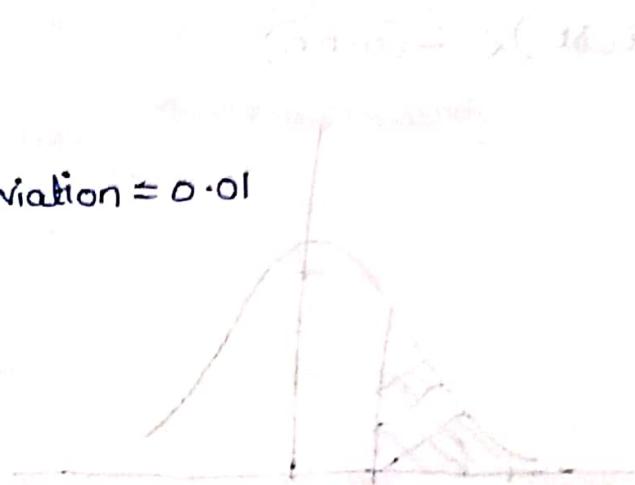
$$s = 90$$

$$\text{Size} = 120$$

$$\text{Standard deviation} = 0.01$$

$$H_0: \text{avg} \geq 1000$$

$$H_r: \text{avg} < 1000$$



$$SE = \frac{s}{\sqrt{n}}$$

$$= \frac{90}{\sqrt{120}} = 8.22$$

$$z(\text{test}) = \frac{\bar{x} - \mu}{SE}$$

$$= \frac{1100 - 1000}{8.22}$$

B-22

$$= 12.16$$

Cost of Type I error is minimum at

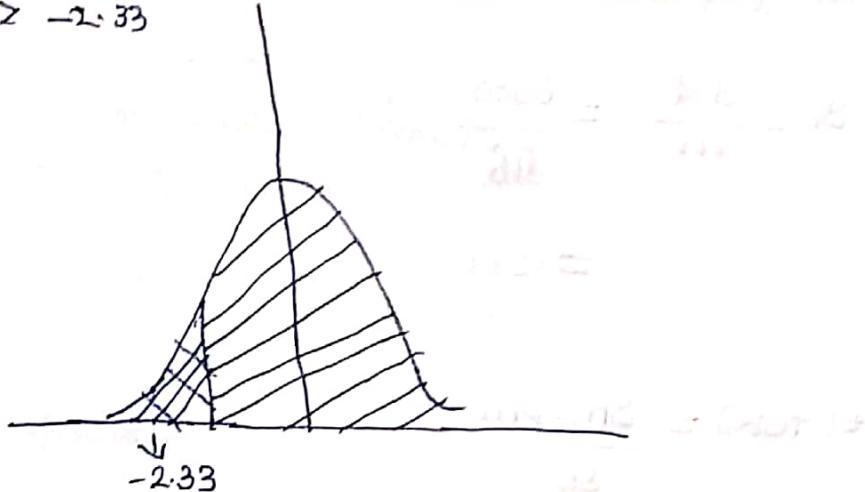
Maximum difference between \bar{x} and μ given as 12.16

$$z(0.01) = -2.33 \begin{bmatrix} -2.30 \\ -0.03 \end{bmatrix} \rightarrow 0.0099 [0.01]$$

From Z-table

$$Z_{\text{test}} > z_{0.01}$$

$$12.16 \geq -2.33$$



H_0 : Accept

H_1 : Reject

T-Statistics

multiple samples
Population is not given

small sample size ($n < 30$)
Unknown variance

- ① A tyre manufacturer claims that the average life of a particular category of its type is 18000 km when used under normal driving conditions. A random sample of 16 tyres was tested. The mean and SD of life of the tyres in the sample were 20000 km and 6000 km respectively. Assuming that the life of the tyres is normally distributed, test the claim of the manufacturer at 1% level of significance.

H_0 : Population mean = 18000 km

H_1 : Population mean $\neq 18000$ km

$$SE = \frac{s \cdot d}{\sqrt{n}} = \frac{6000}{\sqrt{16}}$$

$$= 1500$$

$$t(\text{Test}) = \frac{s_m - pm}{SE}$$

$$= \frac{20000 - 18000}{1500}$$

$$\text{Mean} = 18000 \\ \text{SD} = 6000 \\ t = \frac{20000 - 18000}{6000 / \sqrt{16}} = 3.33$$

$$t_{\text{crit}} = 2.947$$

$$s = 20000 \\ s \cdot d = 6000 \\ \text{Sample Size} = 16$$

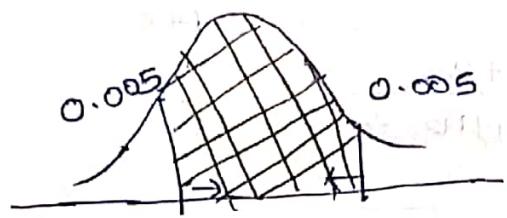
$$1.3$$

$$-2.9 2.9$$

$$df = \text{Sample Size} - 1 = 15$$

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

$$2.947$$



$$\frac{0.10}{2}$$

$$\frac{0.10}{2}$$

$$t(0.005) = 2.947$$

$$t(-0.005) = -2.947$$

$-2.947 < 1.33 < 2.947$ [t Test lie in between these regions]

Accept H_0

Reject H_1

- ② The means of two random samples of sizes 10 and 8 from two normal population is 210.40 and 208.92. The sum of squares of deviation from their means is 36.96 and 24.50 respectively.

Assuming population with equal variances can we consider the normal populations have equal mean (significance level = 5%).

$$n_1 = 10$$

$$n_2 = 8$$

$$n_1(\text{mean}) = 210.40 (\bar{x})$$

$$n_2(\text{mean}) = 208.92 (\bar{x})$$

$$\begin{aligned}
 SEd &= \sqrt{\frac{26.94 + 24.50}{10+8-2}} \\
 &= \sqrt{\frac{\mu_1 + \mu_2}{n_1+n_2-2}} = 1.745 \\
 SE &= \frac{SEd}{\sqrt{n}} = 1.745 \sqrt{\left(\frac{1}{8} + \frac{1}{10}\right)} \\
 &= \frac{1.745}{\sqrt{8+10}} \\
 &= 0.84
 \end{aligned}$$

$$\begin{aligned}
 DF &= n_1+n_2-2 \\
 &= 10+8-2
 \end{aligned}$$

$$DF = 16$$

$$t = \frac{26.94 - 24.50}{0.84} = \frac{Sm - Pm}{SE}$$

$$t = 1.76$$

$$H_0: \mu(P(n_1)) = \mu(P(n_2))$$

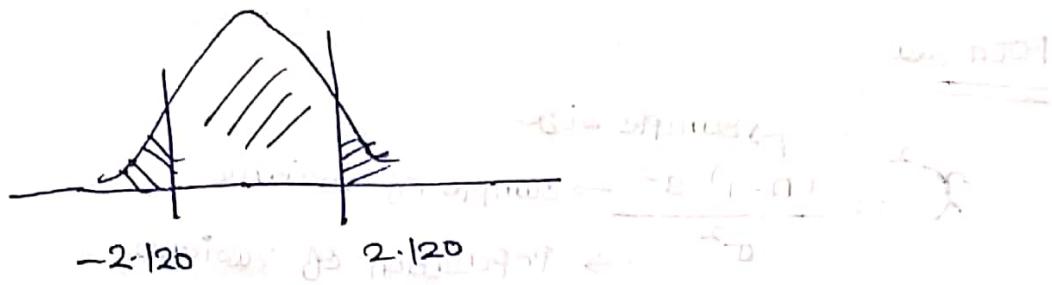
two populations are equal-mean with related

$H_1:$ Population mean are not equal

< >

Significance level = 5.1% ≈ 0.05

Decision: $\frac{0.05}{\alpha} = 0.025$ right tail test $\rightarrow 15\%$
P-value: $t_{\text{obs}} = -1.76$ \rightarrow not enough evidence to reject H_0



$$t = 1.76$$

$$t(5\%) = 2.120$$

$-2.120 \leq 1.76 \leq 2.120$ in shadow \rightarrow not enough

to reject H_0 as per significance level

Accept H_0 as best evidence of no difference

Reject H_0 in evidence of significant difference
 \rightarrow enzyme catalyzes conversion \rightarrow prime site

P-value

$$\frac{-2(1-\alpha)}{2\sigma}$$

It is less stable than z-statistic

$$\begin{aligned} & 1 - \Phi(-1.76) \\ & 1 - 0.959 \\ & 0.0409 \\ & 0.0409 \times 2 \\ & 0.0818 \end{aligned}$$

chi-squared test → For variance
→ For Expected and Actual one

- * It deals with high significant statistical analysis
- * To compare the two independent data

Formula

$$\chi^2 = \frac{(n-1) s^2}{\sigma^2} \quad \begin{array}{l} \text{→ sample size} \\ \text{→ sample of variance} \\ \text{→ population of variance} \end{array}$$

Problem statement

- ① The variance of a certain size of towel produced by a machine is 7.2 over a long period of time. A random sample of 20 towels gave a variance of 8. You need to check if the variability for towel has increased at 5% level of significance assuming a normally distributed sample.

$$\chi^2 = \frac{(n-1) s^2}{\sigma^2}$$

$$n=20$$

$$s^2=8$$

$$P.V = 7.2$$

$$= \frac{(20-1) 8}{7.2}$$

$$\chi^2 = 21.11$$

$$H_0: \chi^2 = 7.2$$

$$H_1: \chi^2 > 7.2$$

Chi-square test of independence (contd.)

$$\text{critical value (5%)} = 30.14$$

Second definition for chi-squared test

The chi-square test is widely used to estimate

how closely the distribution of a categorical variable

matches an expected distribution (the goodness-of-fit test), or to estimate whether two categorical variables

are independent of one another (the test of independence)

To calculate goodness-of-fit-test by chi-squared test

Formula

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

O → actual count in each category

E → Expected count in each category.

$H_0: O = E$

$H_A: O \neq E$

- ① A survey conducted by a pet food company determined that 60% of dog owners have only one dog, 28% have two dogs and 12% have three or more. You were not convinced by the survey and decided to conduct your own survey and have collected the data below,

Data : Out of 129 dog owners, 73 had one dog and 38 had two dogs. Determine whether your data supports the results of the survey by the Pet

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Significance level

Expected one

$$E(1\text{dog}) = 60\% = 0.60 \times 129 = 77.4$$

$$E(2\text{dog}) = 28\% = 0.28 \times 129 = 36.12$$

$$E(3\text{dog}) = 12\% = 0.12 \times 129 = 15.48$$

Actual one

$$O(1\text{dog}) = 73$$

$$O(2\text{dog}) = 38$$

$$O(3\text{dog}) = 18$$

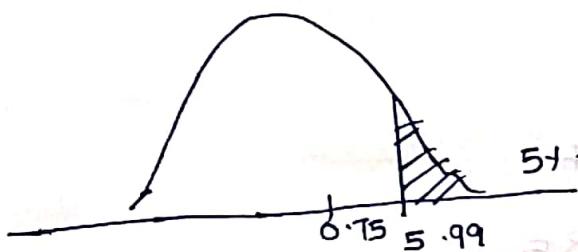
$$\chi^2 = \frac{(73 - 77.4)^2}{77.4} + \frac{(38 - 36.12)^2}{36.12} + \frac{(18 - 15.48)^2}{15.48}$$

$$= \frac{19.36}{77.4} + \frac{3.53}{36.12} + \frac{6.35}{15.48}$$

$$\chi^2 = 0.75$$

$$df = 3-1 = 2$$

$$\chi^2_{(table)} = 5.99$$

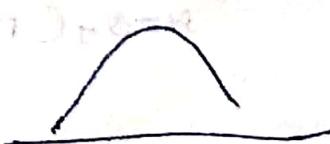


ANOVA Test (To calculate variance between and within analysis of variance)

It is used to mean of two or more samples are different from each other not.

v_1	v_2	v_3
s_1	s_2	s_3
Σ	Σ	Σ

1 factor
one-way ANOVA



Assumptions

- * Sample should be independent
- * Data should be normally distributed
- * Variance should be equal

Problem Statement

- ① In a survey conducted to test the knowledge of mathematics among 4 different schools in city - the sample data collected for the marks of students out of 10 is below.

School marks

School 1 8 6 7 5 9

School 2 2 6 4 6 5 6 7

School 3 3 6 5 5 6 7 8 5

School 3 4 5 6 6 7 6 7

H_0 : All schools are performing equally

H_1 : It's not performing equally

$k=4$ (Total no of samples)

$N=24$ (no of data)

School	Marks					
S ₁	8	6	7	5	9	
S ₂	6	4	6	5	6	7
S ₃	6	5	5	6	7	8
S ₄	5	6	6	7	6	7

Sum of Squares between the sample (S_1, S_2, S_3, S_4)
of mean

$$SS_b = \sum_{i=1}^K n_i (\bar{x}_i - \bar{x}_g)^2$$

↓
Sample
Total
no of data in a sample

\bar{x}_i - sample mean

\bar{x}_g - grand mean

mean between

$$\text{mean } b = SS_b / (k-1)$$

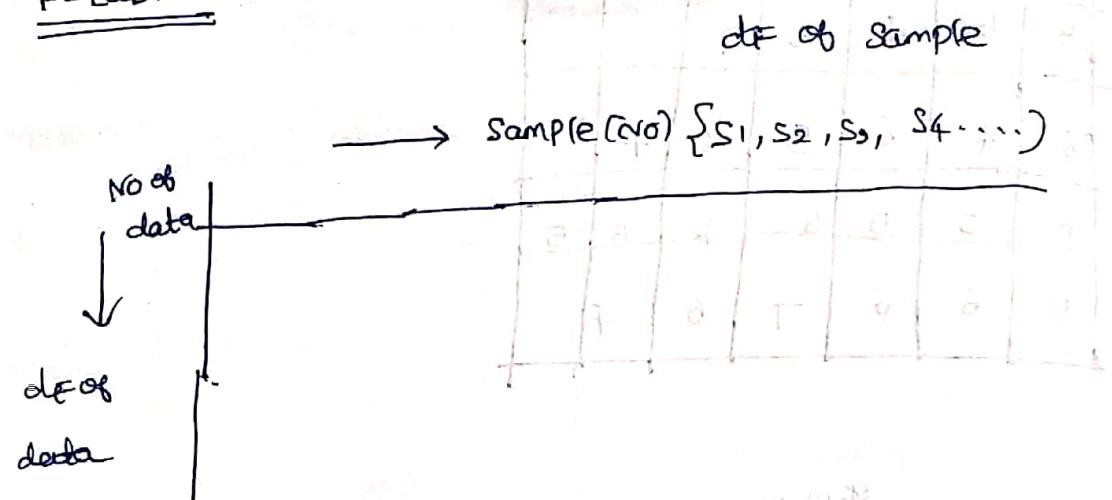
Sum of square mean within

$$SS_{\text{within}} = \sum_{i=1}^N (x_i - \bar{x}_i)^2$$

Mean within

$$\text{mean } w_i = \frac{SS_{\text{within}}}{24-4} = \frac{SS_{\text{within}}}{N-K}$$

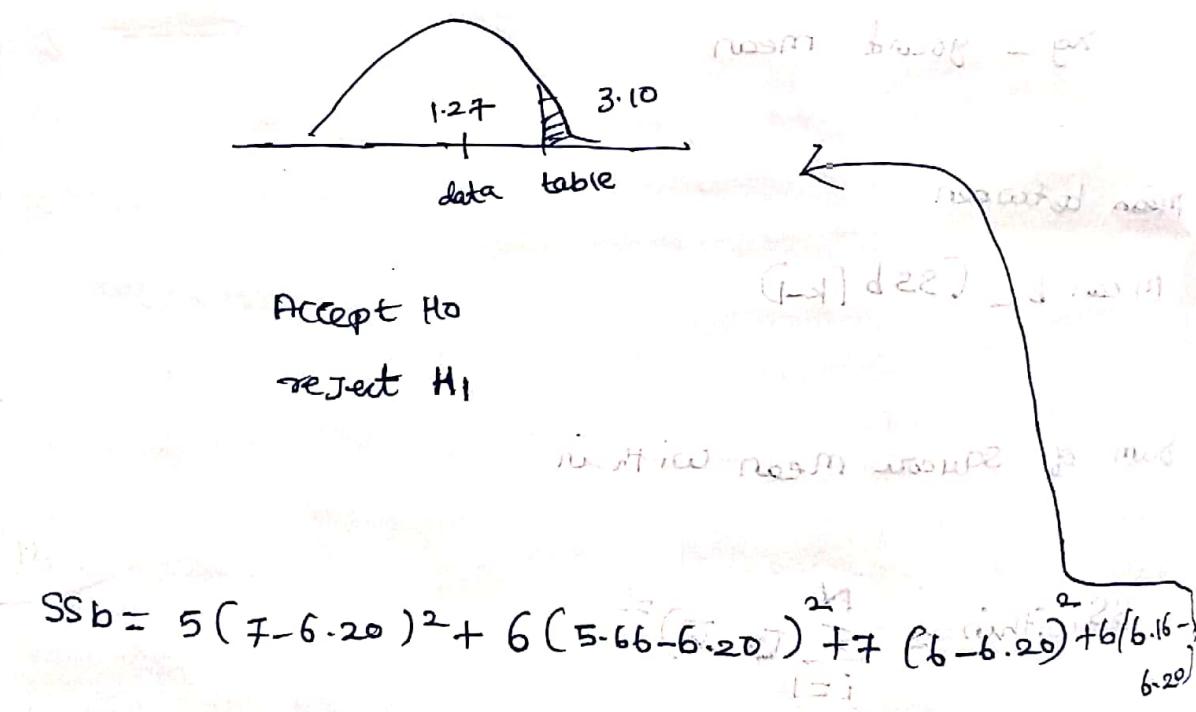
F-table



$$F = 3.10 \text{ (From table)}$$

$$F = 1.27 \text{ (From data)}$$

Diagram



$$SS_b = 5(7-6.20)^2 + 6(5.66-6.20)^2 + 7(6-6.20)^2 + 6(6.16-6.20)^2$$

$$= 4.99$$

$$SS_{\text{within}} = (0 + 5.33 + 8 + 2.83) \\ = 26.16$$

$$\text{mean}_{\text{within}} = \frac{SS_{\text{within}}}{N-k} \\ = \frac{26.16}{20} \\ = 1.30$$

27.9	x
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

$$\text{mean}_b = \frac{SS_b}{k-1} \\ = \frac{4.99}{4-1}$$

$$\frac{1}{3}x_3 + \frac{1}{3}x_4 + \frac{1}{3}x_5 + \frac{1}{3}x_6 + \frac{1}{3}x_7 = 2.5$$

$$F_{\text{test}}(\text{data}) = \frac{MSB}{MSW}$$

$$= \frac{1.66}{1.30}$$

$$(x_i - \bar{x})^2 / (n-1) = 1.66$$

$$\frac{1}{3}(2.5-2)^2 + \dots + \frac{1}{3}(2.5-2)^2 + \frac{1}{3}(2.5-2)^2 = 1.30$$

$$\frac{1}{3}P - \frac{1}{3}E \\ \frac{1}{3}P - \frac{1}{3}E = 1.30$$

sampling distribution of mean

x	$P(x)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

mean

$$\mu = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6}$$

variance

$$\sigma^2 = \sum_{n=1}^{N=6} (x - \mu)^2 P(x)$$

$$= (1-3.5) \frac{1}{6} + (2-3.5) \frac{1}{6} + \dots + (6-3.5) \frac{1}{6}$$

$$= 2.92$$

std deviation

$$\sigma = \sqrt{2.92}$$

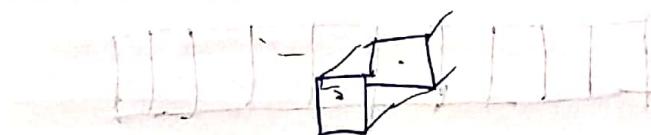
$$= 1.71$$

sample

x	μ	x	μ
1, 1	1	3, 1	2
1, 2	1.5	3, 2	2.5
1, 3	2	3, 3	3
1, 4	2.5	3, 4	3.5
1, 5	3	3, 5	4
1, 6	3.5	3, 6	4.5
2, 1	4.5	4, 1	2.5
2, 2	5.0	4, 2	3
2, 3	5.5	4, 3	3.5
2, 4	6	4, 4	4
2, 5	6.5	4, 5	4.5
2, 6	7	4, 6	5

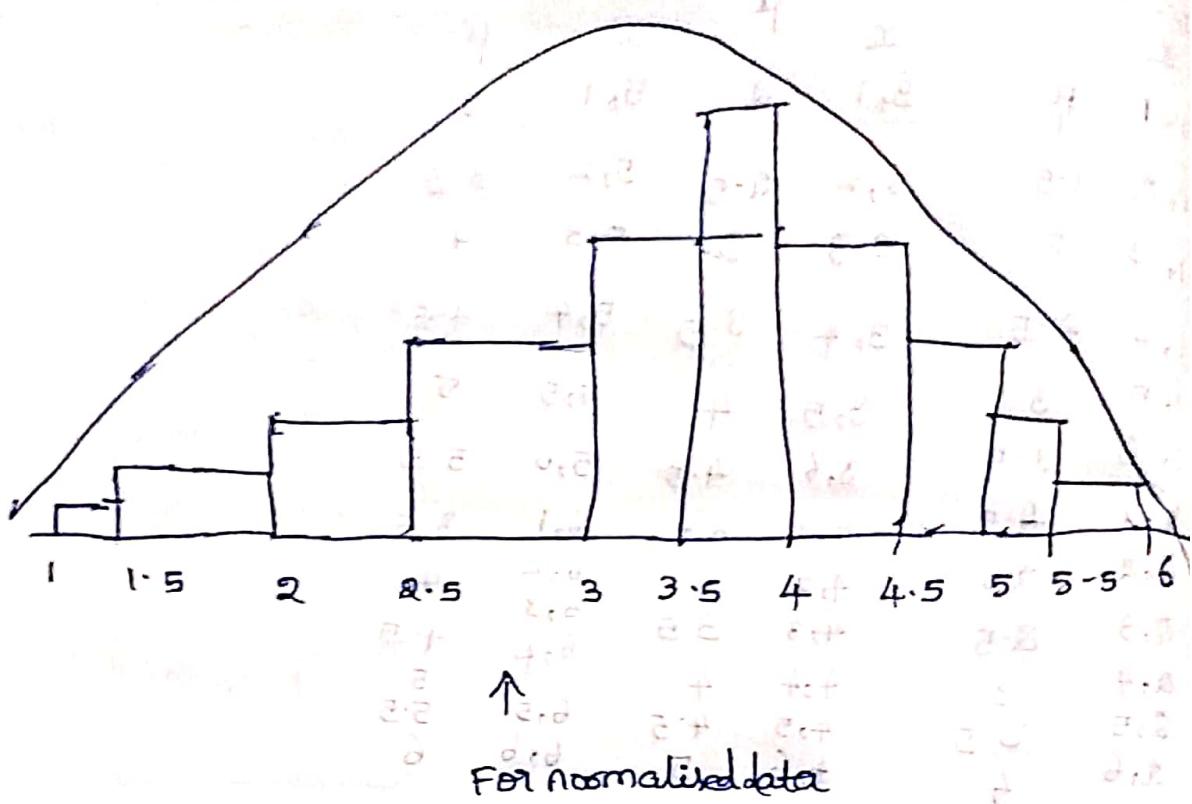


If two dice were thrown



x	$P(x)$
1	$\frac{1}{36}$
2	$\frac{2}{36}$
3	$\frac{3}{36}$
4	$\frac{4}{36}$
5	$\frac{5}{36}$
6	$\frac{6}{36}$
7	$\frac{5}{36}$
8	$\frac{4}{36}$
9	$\frac{3}{36}$
10	$\frac{2}{36}$
11	$\frac{1}{36}$
12	$\frac{1}{36}$

Plotting



for an normalized data,

Central limit theorem

IF you have taken a population, μ , s.d.,
if you have taken a large amount of sample
of population . then distribution of sample

mean will always be close to a normal distribution.

A/B testing

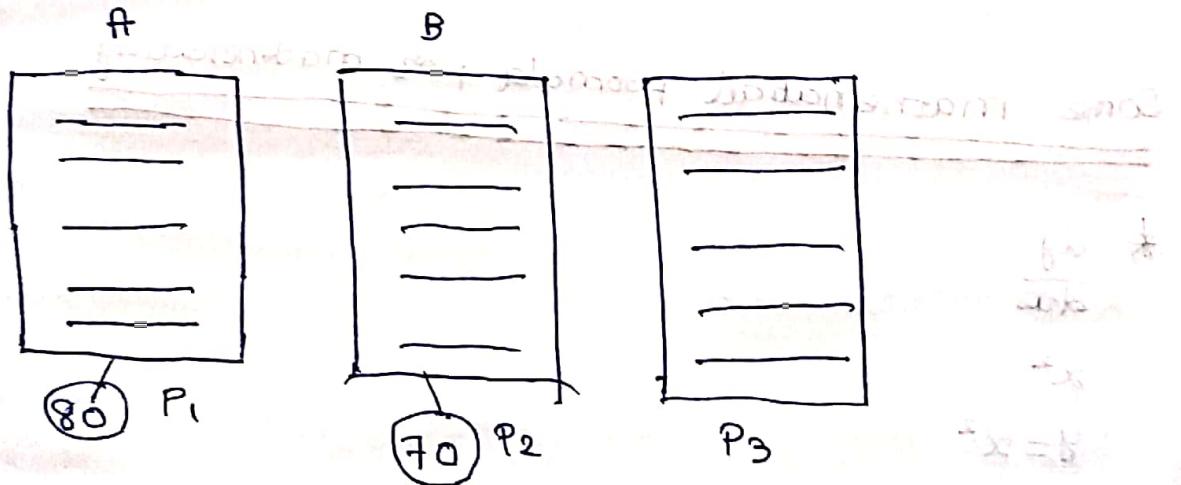
To compare diff version of experiment

Other names

- * Bucket testing
- * split testing

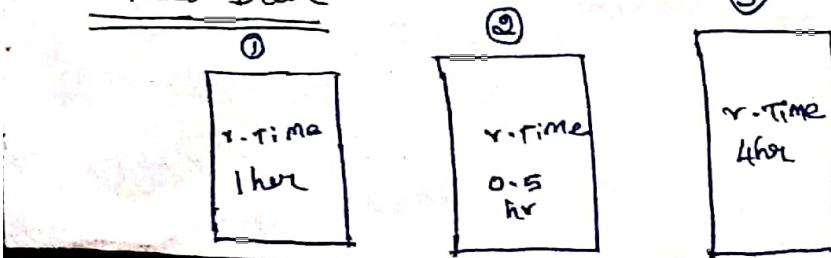
Diagram

Total 150 pages



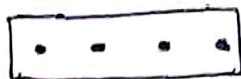
Steps for proceeding A/B testing

* collect Data



* Goals

Buttons for the ad



* Hypothesis

IF we get some profit means we will earn

or else not.

* Variations

comparing different pages

* running experiment

* analyze result

Some mathematical formula for machine learning

$$* \frac{dy}{dx}$$

$$x^2$$

$$y = x^2$$

$$\frac{dy}{dx} = 2x$$

$$* y = x^2$$

$$y = 2^2$$

$$y = 4$$

$$y = 3^2$$
$$= 9$$

$$y = 4^2$$

$$y = 16$$

$$y = 5^2$$

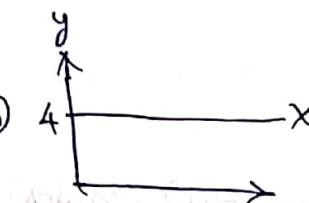
$$y = 25$$

Formula

$$* y = c$$

$$\frac{dy}{dx} = 0$$

$$\left\{ \begin{array}{l} \text{if } (y=0) \\ \text{at } x=0 \end{array} \right.$$



$$x_0 = \frac{b}{m}$$

$$(ii) \frac{dy}{dx} = x^n *$$

$$\frac{dy}{dx} = n x^{n-1}$$

$$\cos A \cos B + \cos B \cos A = 1 = \cos A$$

$$* y = x^2$$

$$\frac{dy}{dx}$$

$$= 2x$$

$$(x-1)(x+1) = 0 *$$

$$* y = \sin x$$

$$\frac{dy}{dx} = \cos x$$

sec & cosec

$$* y = \cos x$$

$$\frac{dy}{dx} = -\sin x$$

$$* y = \tan x$$

$$\frac{dy}{dx} = \sec^2 x$$

$$* \quad y = x^2$$

$$\frac{dy}{dx} = 2x$$

$$* \quad y = \ln(x)$$

$$\frac{dy}{dx} = \left(\frac{1}{x}\right)$$

$$* \quad y = e^x$$

$$\frac{dy}{dx} = e^x$$

Tools (Sine and cosine tool) (Calculator)

$$* \quad y = f(x) = h(x) \cdot g(x)$$

$$f'(x) = y' = h'(x) \cdot g(x) + g'(x) \cdot h(x)$$

Eg :-

$$* \quad y = (1-x^2)(1-x^3)$$

$$\frac{dy}{dx} = -2x(1-x^3) + -3x^2 \cdot (1-x)$$

Cosine tool

$$f(x) = y = \left(\frac{h(x)}{g(x)} \right) = \left(\frac{h'(x)g(x) - h(x)g'(x)}{(g(x))^2} \right)$$

Eg

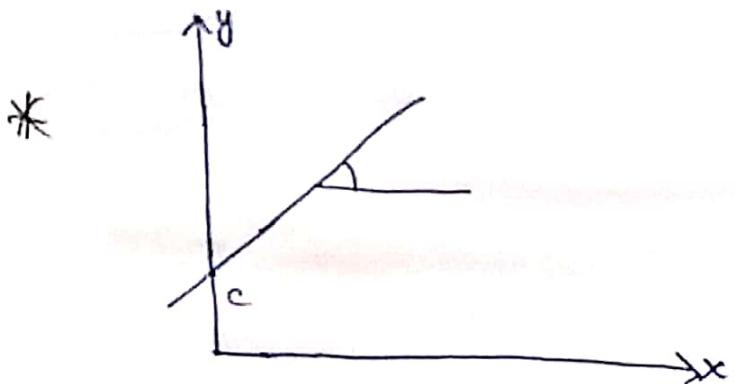
$$* y = \left(\frac{1-x^2}{1+x^3} \right)$$

$$= \left(\frac{-2x(1+x^3) + 3x(1-x^2)}{(1+x^3)^2} \right)$$

chain rule

It is used mainly in neural networks.

$$* \frac{dy}{dx} = \frac{dy}{dz} \cdot \frac{dz}{dx}$$



$$* y = ax$$

$$y = mx + c$$

when $x=0$

$$y=c$$

$$* y = mx$$

$$\frac{y}{x} = m$$

$$y = ax$$

$$\frac{y}{x} = a$$

$$* y = mx + c = 0$$

$$\frac{dy}{dx} = m$$

$$(x-a)x + (x+a) \text{ is } \\ \text{even} \quad \text{odd}$$

relation between m given below in Eq.

$$\frac{dy}{dx} = m \quad \frac{dy}{dx} = \frac{6h}{2h} = \frac{6h}{-2h} = -\frac{6h}{2h}$$



$$20h = 6 \times 2h$$

or $m = 6/2$

$m = 3$

$m = 3$

$$m = 6/2$$

$$m = -\frac{6}{2}$$