12/1/2021

Probability distributions:

***** v.Imp.

v.Imp

→ machine Learning

→ Data Science

reqs → Research

Probability Distributions.

Discrete distribution

→ Binomial
→ poisson

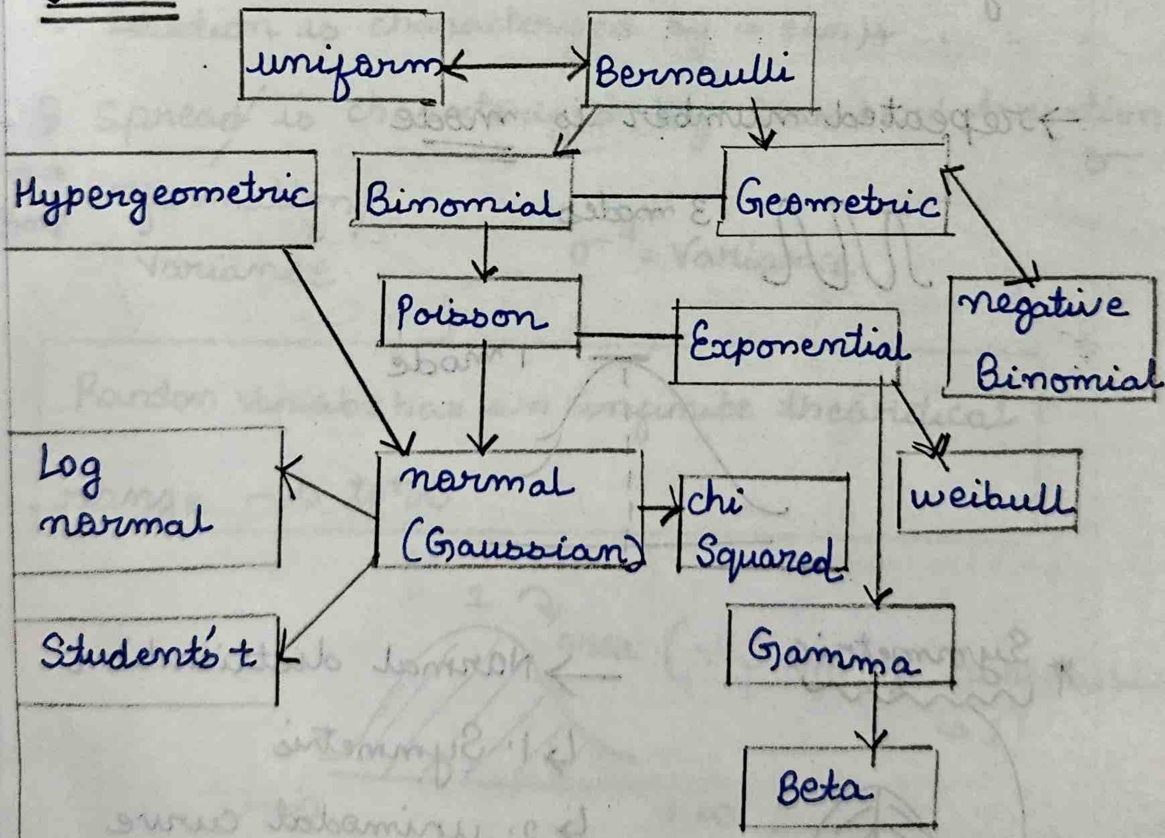
} distributions.

Continuous distribution

→ Normal (Guassion)

→ Expontential distribution

General:



******* ✓Imp. NORMAL DISTRIBUTIONS:**

✓Mother of all distributions is normal distribution.

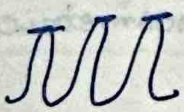
✓Guassion distribution [Normal]

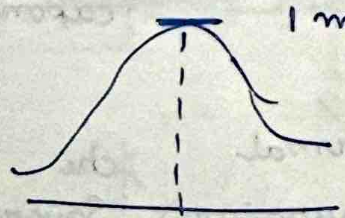
✓Biologists use his information of the normal Distribution to Study patterns in nature.

Imp To Study patterns in nature.
Gauss is
✓Father of first Predictive algorithm.

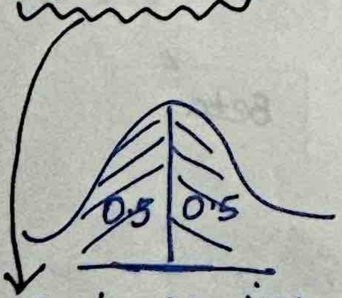
✓Linear Regression is base for most of machine learning and data Science algorithms.

→ repeated number is mode


3 modes


1 mode

* Symmetric

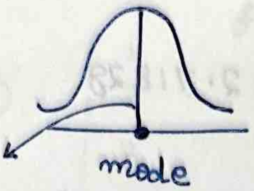


mirror image

→ Normal distribution
↳ 1. Symmetric
↳ 2. unimodal curve
↳ 3. Asymptotic

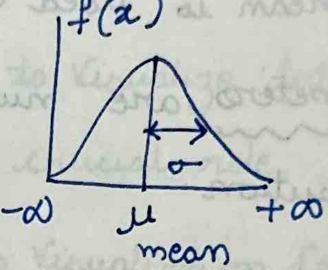
Asymptote: Tangent at ∞
↳ Touches / meets at Single Point.

x axis is tangent for normal curve



** mean = median = mode [In normal distribution]

* Properties

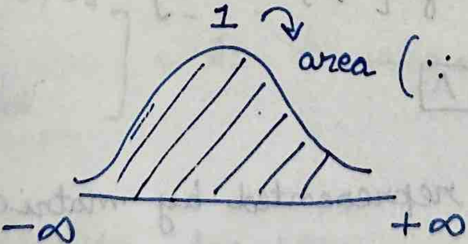


1. Bell Shaped
2. Symmetrical
3. mean, median, mode are equal
4. Location is characterized by mean μ
5. spread/ is characterized by Standard deviation σ

** Imp information/ Variance

σ^2 = Variance

Random Variable has an infinite theoritical range $-\infty$ to $+\infty$



V. Imp

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}}$$

; μ, σ are parameters

$f(x)$ = Probability density function of normal distribution

x Values $-\infty$ to $+\infty$; x is Variable

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$\downarrow$

$e = 2.71828$

Values $-\infty$ to $+\infty$ $\pi =$

x is a Variable

μ, σ are parameters

[$\because$ mean = median = mode in normal distribution
mean is used Generally]

parameters are numerical descriptors of a distribution.

* μ, σ decides Shape, Structure in normal distribution.

15/1/2021:

General Maths Lecture

Maths

- Linear Algebra
- Calculus

Statistics

- Descriptive
- Inferential

} Statistics

→ 64K images of $\{0, 1, \ldots, 9\}$ MNIST data Set

Eg: $\boxed{1}$ $\boxed{1}$ $\boxed{/}$

Images are represented by matrices

Each image to represent we require 784 Pixels

$28 \times 28 = 784$

✓ Each image is feature

$$D = \begin{array}{c|c} & f_1 \quad f_2 \quad f_3 \cdots\cdots f_{784} \\ \end{array}\begin{bmatrix} 1 & \\ 2 & \\ 3 & \\ \vdots & \\ 64K & \end{bmatrix}$$

↓

Data

64000 × 784

→ dimensions → Columns

$784 \xrightarrow{\text{Convert}}$ 2D or 3D to Visualize data

(Linear Algebra) plays crucial role

* Dimensionality Reduction
  - ✓ Data Visualization [2D or 3D] [∵ ease] ↓ even 10D
  - ✓ To apply machine learning algorithm

→ To convert to 2D

we need Eigen Values; consider top two Eigen values λ_1 λ_2

$\lambda_1 \downarrow v_1$, v_2

Each Eigen Value ↓ has Eigen Vector

with principle Component analysis (1933)

$$D = \begin{array}{c} f_1 \; f_2 \end{array}\begin{bmatrix} 1 \\ 2 \\ \vdots \\ 64K \end{bmatrix}$$ 2D Visualization.

Eg: To identify dog, cat in an image we need Linear algebra

Calculus:

- Geoffrey Hinton is father of modern deep Learning

→ Deep learning $\xrightarrow{\text{based on}}$ chain rule
↓
** Calculus

Back propagation algorithm → Calculus chain rule

Statistics:

* Decision making
  - ↳ fact / fast based
  - ↳ not by Emotions
  - → Backed up data [information]

→ Branch of Mathematics is Statistics
Collect, organize, interpret data

Descriptive Statistics
- uni variate
- Bi Variate
- multi Variate

To Solve inferential
↓ depends on
descriptive Statistics

→ Summarize data

*** v Imp
✓ Inferential Statistics:
- ↳ Research
- ↳ academics
- ↳ Industry
  Eg: Covid
- ↳ medical
- ↳ Engineering
- ⋮

Inferential Statistics:

mp ✓ Hypothesis testing

model fitting → confirms which model fits best.

↳ $\left.\begin{array}{l}\text{regression}\\ \text{Logistic}\\ \text{multi linear}\end{array}\right\}$ model

→ Model is relation between dependent Variable and independent Variable

Descriptive Statistics.

- Mean
- median
- mode
- Standard deviation
- Sample variance
- Kurtosis
- Skewness
- Range
- maximum
- minimum
- Sum
- Count
- Geometric mean
- Harmonic mean
- AAD
- MAD
- IQR

$X = [\ \ ]_{64K \times 784}$

$X^T = [\ \ ]_{784 \times 64K}$

$C = X^T \times [\ \ ]$ ← 784×784

↓

← Covariance matrix

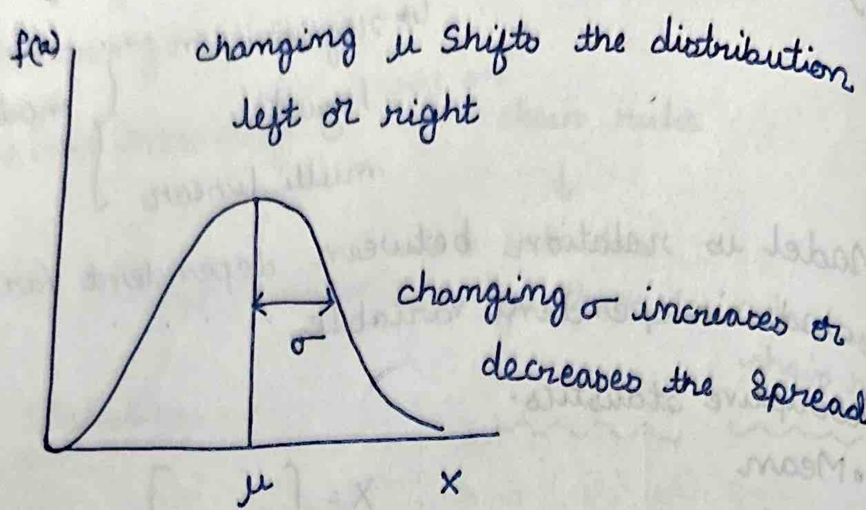
↓ under multivariate

descriptive matrix

16/1/2021:

Normal distribution .... [Contd]

Shape



→ μ → we can shift mean

→ σ → spread of location

★ Converting $X \to Z$

↳ Normal distribution (X)

Imp ★ Standardized Normal distribution (Z)

① Needs to be transform x units into Z units

② mean is 0 and standard deviation is 1

③ ✳ $\mu = 0$ $\sigma = 1$ (in Z)

✳ Standardized Normal distribution:

$$\boxed{Z = \frac{X - \mu}{\sigma}}$$

Z → Standard normal variate / variable

X → normal distribution

μ → mean

σ → Standard deviation

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} \rightarrow$$ [Pdf of Standard normal deviation distribution]

$$Z = \frac{x-\mu}{\sigma}$$ [$\because \underbrace{\mu = 0, \sigma = 1}$] Conditions.

$$Z = \frac{x-0}{1}$$

Z is increasing; f(z) is decreasing

$$\left[\because f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty \le x \le \infty\right.$$

$$= \frac{1}{1 \times \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-0}{1}\right)^0} \quad -\infty \le x \le \infty$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2} \quad -\infty \le x \le \infty$$

$$\left.\left(\frac{x-0}{1} = Z\right)\right]$$

$\boxed{\text{mean is 0 and Standard deviation is 1}}$

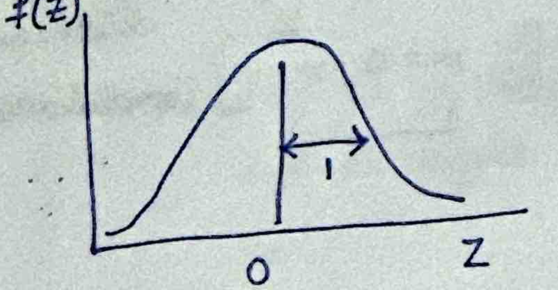
$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}}$	Probability density function of Standardized normal distribution.

where e = the mathematical constant approximated by 2.71828

π = the mathematical constant approximated by 3.14159

Z = any value of the Standardized normal distribution

Standardized Normal distribution:



68
34 34

95% $\mu+2\sigma$ $\mu-2\sigma$

68 < 95

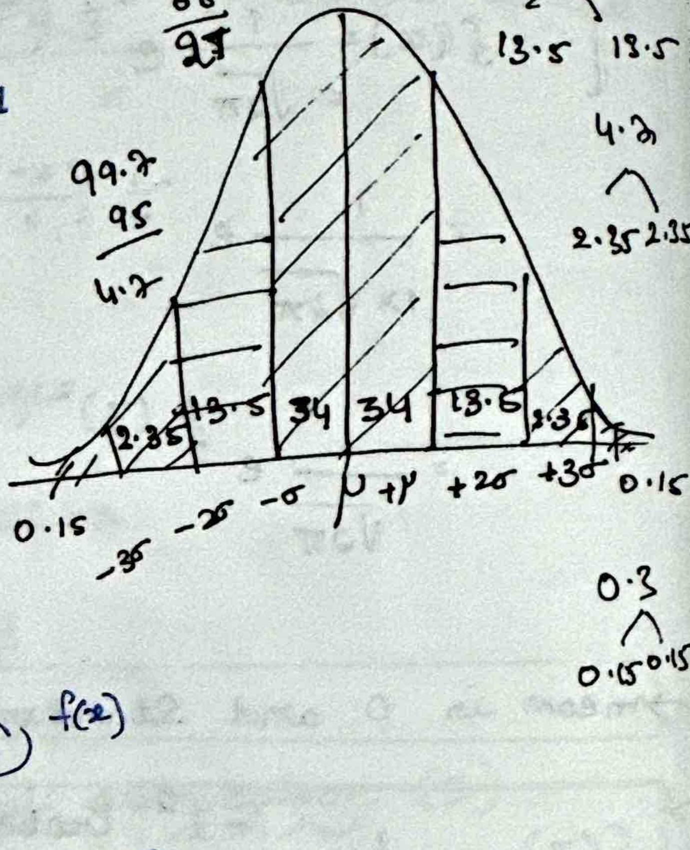
68 + 27

13.5 13.5

4.7

2.35 2.35

99.7
95
4.7



0.3
0.15 0.15

- → Z distribution
- → mean 0
- → Standard deviation 1
- → positive Z values
- → Negative Z values.

Eg:

Supposely

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

$$f(z) = e^{-z^2}$$

1SD 34

$z = 0$

$f(0) = 1$

2SD 2.35

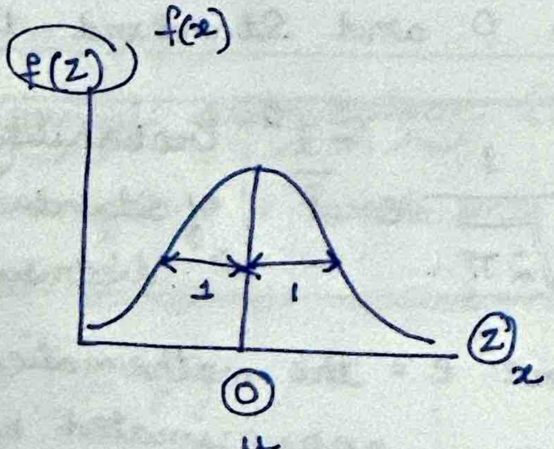
$z = 1$

$f(1) = e^{-1}$

3SD 0.15

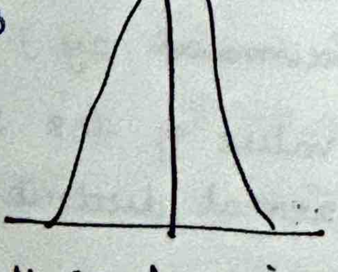
$= \frac{1}{e}$

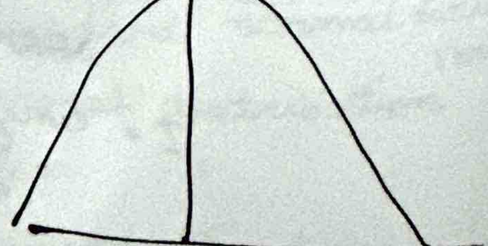
$= \frac{1}{2.713}$



34.00
02.35
00.15
96.50

Role of σ





Small SD few points far from mean

large

* Calculate the height of all individuals in world. 68% of people will be with 1 SD. avg in terms of ht.

① If x is distributed normally with mean of 100 and σ of 50, the z value for $x = 200$ is

Sol:

$$Z = \frac{x - \mu}{\sigma}$$; σ is Standard deviation.

$$\Rightarrow \frac{200-100}{50} \Rightarrow \frac{\cancel{100}\,2}{\cancel{50}}$$

$$= 2.0$$

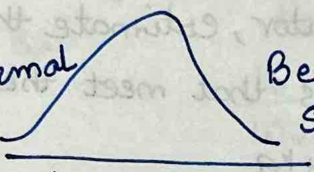
☆ applicable for forecasting when actual data pred. is ☹

** Empirical Rule for Normal Distribution:

also known (68-95-99.7 rule)

* Rule holds good only for (applicable) Normal distribution

Symmetrical curve.



Empirical rule is also called as

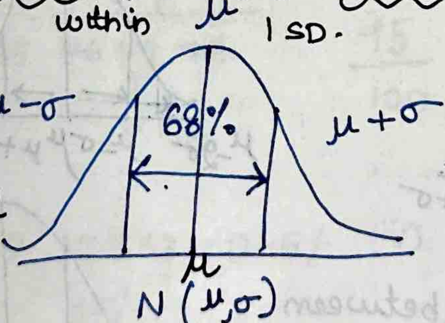
↳ 68-95-99.7 rule

↳ Three Sigma rule

↳ 3 Standard deviations of a mean

* One Standard deviation distance:

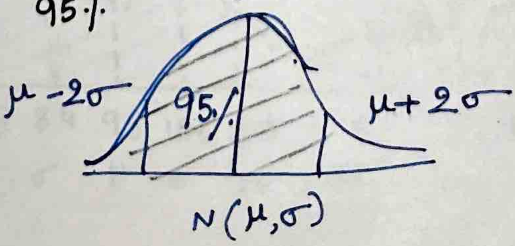
Supposely datapoints → 68% of worlds height avg will be in 1σ



many points closer to average.

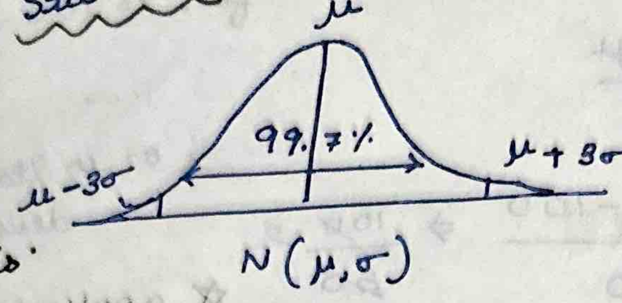
* Two Standard deviation

extreme values exists. ↑↓ } outliers



Empirical Rule

Three Standard deviation: 99.7



99.7% within 3 Standard deviations of mean

② Assume that the mean of 1 year-old girls in the US is normally distributed with mean of about 9.5 kg with standard deviation of ≃ 1.1 kg. without using calculator, estimate the % of 1 year-old girl in the US that meet the following conditions:

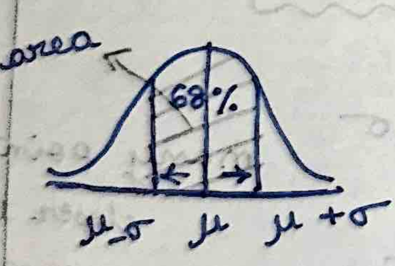
*** ** v.Imp

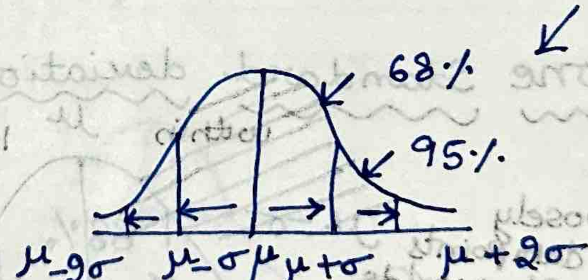
1. Less than 8.4 kg
2. Between 7.3 kg and 11.7 kg
3. more than 12.8 kg

Sol: Given $\mu = 9.5$ kg, $\sigma = 1.1$ kg

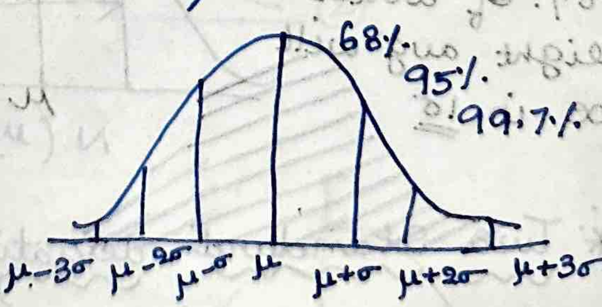
rule * 68 − 95 − 99.7

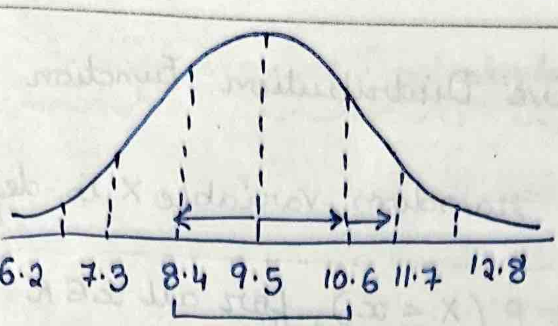
68% ⊂ 95%





68% is area between μ−σ and μ+σ
μ−σ μ+σ





$\mu = 9.5\,kg$

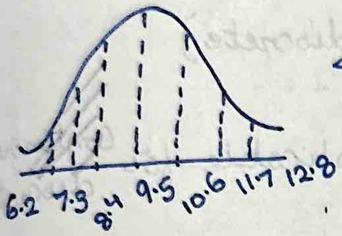
$\sigma = 1.1\,kg$

$\mu - 1\sigma = 9.5 - 1.1 = 8.4$

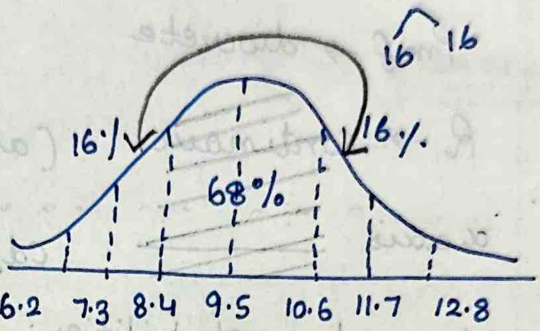
$\mu + 1\sigma = 9.5 + 1.1 = 10.6$

[∵ Symmetric distribution

(i) Less than 8.4 Kg



$< 8.4\,kg$

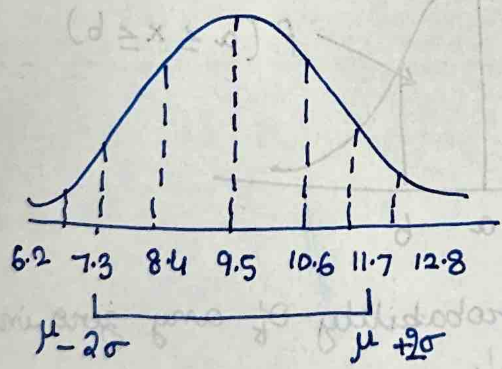


$100\% \rightarrow 100 - 68$
$= 32$

∴ Less than $8.4\,kg = 16\%$.

$\frac{16}{100}$ is Probability.

(ii) Between 7.3 Kg and 11.7 Kg $= \frac{95}{100} = 0.$



area between $\mu - 2\sigma$ and $\mu + 2\sigma$ is 95%

~~100-95=5%~~

$\frac{95}{100}$ probability

(iii) more than 12.8 Kg = 0.15%

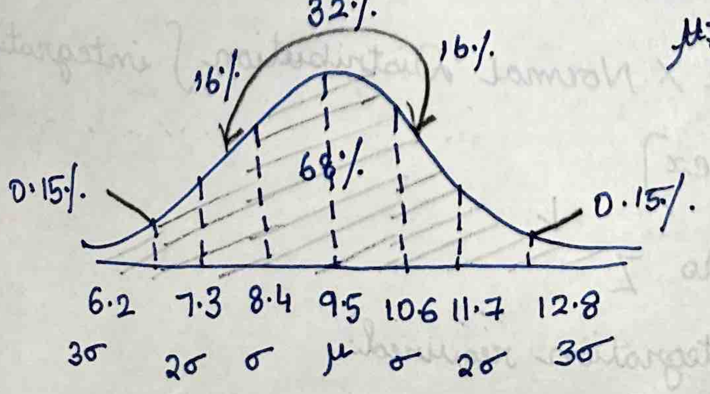
(iv) below than 6.2 is 0.15%

$\mu - 3\sigma$ and $\mu + 3\sigma$

99.7%

$100 - 99.7$
$= 0.3\%$

0.15 0.15



Cumulative Distribution Function

CDF of random variable X is defined as

$$\boxed{F_X(x) = P(X \leq x), \text{ for all } x \in R}$$

cdf → describes distribution of a random variable

Pdf → continous

Pmf → discrete

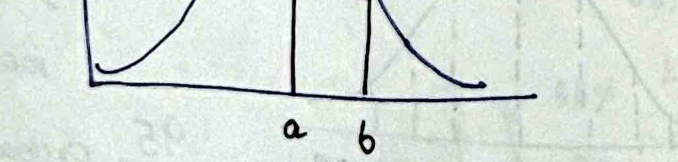
R → continous (also for discrete)

x axis ———— x axis - - - -

cdf is applicable for continous / discrete

Normal probabilities:

→ Probability is measured by the area under the curve



Note that the probability of any zero individual value is zero.

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \quad (\because \text{area})$$

[∵ In X Normal Distribution ∫ integration is complex]

↓

convert to Z

no integration required.

Applications of Cumulative Distribution Function.

* Standard Normal table
* unit Normal Table (also called)
* Z table

} is applicable in field of data Science.

[∵ we need not to find integration to find probability using above tables.

→ using PDF and CDF we can find probability of a Random Variable.

$f(x) = Pdf$

$F(x) = Cdf$

$$f(x) = \frac{d}{dx}\left[F(x)\right]$$

Statistics in Understanding Data

* To understand data we require Statistics.

"There are two kinds of Statistics, the kind you look up and the kind you make up"
Rex Stout

What is Statistics?

* A branch of mathematics that deals with collecting, analyzing and interpreting data.

✓ Descriptive
✓ Inferential

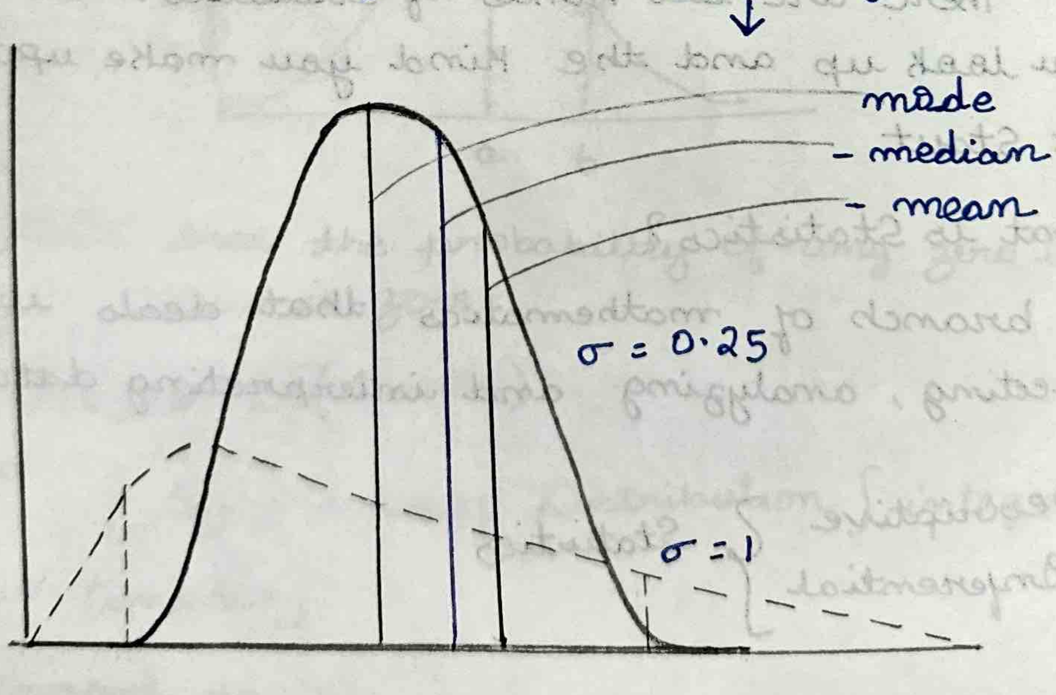
} Statistics

Descriptive Statistics:

- Summarize, describes the data
- Identifies important elements in dataset without Seeking to find explanations of those elements
- Depending on how many variables are involved descriptive statistics are of 3 types

- Summarize single Variable → univariate
- Relation/b/w 2 Variables → BiVariate
- Describe → multiple Variables multivariate

- Summarize data as it is
- Do not posit any hypothesis about data
- Detect outliers
- Plan how to prepare data
- precursor to feature engineering



✓ Feature engineering involves extracting (features) data to build mode

↓ we require features → we require feature engineering

✓ Precursor to feature engineering.

univariate (Single variable)

→ Frequency
→ Central Tendency
→ Dispersion

Measures of frequency:

↳ Frequency Tables

↳ Histograms

Summary of Statistics:

Measures of Central Tendency:

Measures of Location

- Arithmetic mean
- ✓ weighted mean
- ✓ median
- ✓ mode
- ✓ Percentile
- ✓ Geometric } means → infrequently used
- ✓ Harmonic }

Measures of Dispersion

- ✓ Skewness
- ✓ Kurtosis
- ✓ Range
- ✓ Interquartile range
- ✓ Variance
- ✓ Standard Score
- ✓ Co-efficient of Variance

* Arithmetic Mean

→ Data : {16, 17, 10, 13, 20, 18, 13, 14, 18}

$$\bar{x} = \frac{16+17+10+13+20+18+13+14+18}{9}$$

$$\bar{x} = \frac{139}{9}$$

$$\bar{x} = 15.444$$

* Commonly known as 'mean'
* Average of Group of numbers
* Applicable for interval and ratio data

Eg:

Interval data

[2000, 2020]

ratio data → Heights, weights → cuts are allowed

* Not applicable for namidal or ordinal data

Imp

Gender ⟨ male - 1, female - 0

1. Prof, AssoP, Assis P
2. Grades of Students

* Affected by each value in the dataset, including extreme values

* Computed by summing all values in the dataset and dividing the sum by the numbers of values in the dataset.

Population Mean:

Data: 60 20 10 40 50 30

$$\mu = \frac{\Sigma x_i}{n} = \frac{60+20+10+40+50+30}{6} = 35$$

* Impact of outliers:

Data: 60 20 10 40 50 30 1000

$$\mu = \frac{\Sigma x_i}{n} = \frac{60+20+10+40+50+30+1000}{7}$$

Mean = 172.85

Population mean $\rightarrow \mu$

Sample mean $\rightarrow \bar{x}$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

$$= \frac{57+86+42+38+90+66}{6}$$

$$= \frac{379}{6}$$

$$= 63.167$$

* Mean of Grouped Data

* weighted average of class midpoints
* class frequencies are the weights

$$\mu = \frac{\Sigma f M}{\Sigma f}$$ frequency, class midpoint

$$= \frac{\Sigma f m}{N}$$

$$= \frac{f_1 M + f_2 M_2 + f_3 M + \cdots + f_i M_i}{f_1 + f_2 + f_3 + \cdots f_i}$$

f_1 is frequency of first class

M_1 is class midpoint of first class

Calculation of Grouped Mean:

class Interval	Frequency (f)	class midpoint (M)	fM
20-under 30	6	25	150
30-under 40	18	35	630
40-under 50	11	45	495
50-under 60	11	55	605
60-under 70	3	65	195
70-under 80	1	75	75
	50		2150

$\frac{20+30}{2} = 50/2$

$$\mu = \frac{\Sigma f M}{\Sigma f} = \frac{2150}{50}$$

$$= 43.0$$

20/1/2021

* Finding Normal probability procedure:

→ To find $P(a < X < b)$ when X is distributed normally:

→ Draw the normal curve for the Problem in terms of X.

✓ Translate X-values to Z-values
use the Standardized Normal Table

→ Accessing Normality:

CLT
↳ Very Imp

many Analytical and Statistical tools

Data → Data cleaning → verify ND

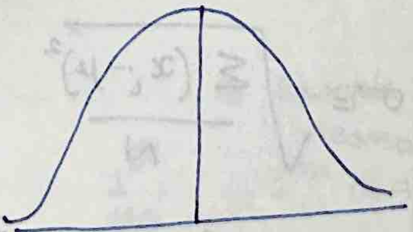
* Demo on Normal Distribution:

```
from Scipy.Stats import norm
Val, m, S = 68, 65.5, 2.5
print(norm.cdf(val, m, S))
```

Val = 68, m = 65.5, S = 2.5
↓ ↓ ↓
$x = 68$ $\mu = 65.5$ $\sigma = 2.5$

$$Z = \frac{x - \mu}{\sigma}$$

$$Z = \frac{68 - 65.5}{2.5}$$



$$= \frac{2.5}{2.5}$$

$$= 1$$

20/1/2021

**
**
**
vgmp

Central Limit Theorem:-

↳ Real time applications

↳ Super useful

↳ Statistics.

↳ Mathematics

Developing a Sampling distribution:

Assume there is a population...

Polulation Size $N=4$

Random variable x, is age of individuals

Values of x:

18, 20, 22, 24 (years)

$$\mu = \frac{\sum x_i}{N}$$

$$= \frac{18+20+22+24}{4} \Rightarrow 21$$

μ is Polulation mean

$\downarrow$

(by default)

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

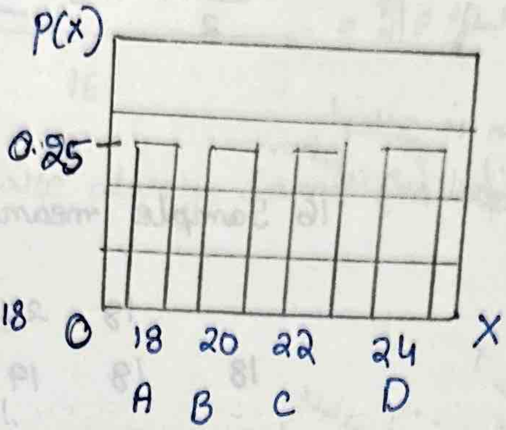
$$= (x_1-\mu)^2 + (x_2-\mu)^2 + (x_3-\mu)^2 + (x_4-\mu)^2$$

18	20	22	24
x_1	x_2	x_3	x_4

on Substitution

$\sigma = 2.236$

$\therefore \mu = 21$



$\therefore$ Probability for 18

$\frac{1}{4}$

uniform distribution

Slly 18, 20, 22, 24

$\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}$ $\left\{ \frac{\text{favourable cases}}{\text{Total}} \right\}$

Now consider all possible Samples of Size $n = 2$

1st obs	2nd observation 18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
N^n 24	24,18	24,20	24,22	24,24

16 possible Samples (Sampling with replacement)

$\rightarrow N=4, n=2$ $N^n = 4^2 = 16$

note:

$N^n \rightarrow$ with repetition

${}^N C_n$ without

${}^N C_n \rightarrow$ without repetition $N = 4$, $n = 2$

$= 16$

1st Obs	2nd observation 18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

16 Sample means

1st Obs	2nd obs			
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

above 2^{nd} table is obtained by averages

avg: $\frac{18+18}{2} = \frac{36}{2} = 18$

$\frac{18+18}{2} = 18$, $\frac{18+20}{2} = 19, \ldots$ so on

mean is called first expectation

16 Sample means

CLT / not applicable

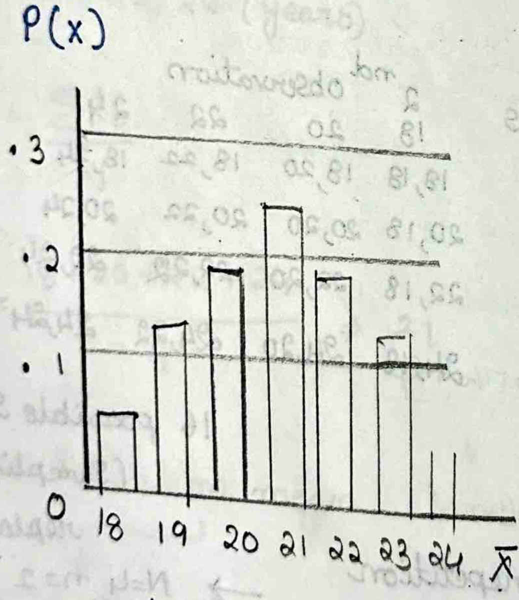
Pareto ✓

{Cauchy ✗ ↓ not use in CSE

	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

means of sample size = 2

Sample means distribution



no longer uniform

$\therefore \mu$ is always population mean

$\bar{x}$ sample mean

Summary measures of this Sampling distribution

$$E(\bar{x}) = \frac{\Sigma \bar{x}_i}{N} = \frac{18+19+21+\cdots+24}{16} = 21 = \mu$$

μ and $E(\bar{x})$ are always Same (values)

$$\sigma_{\bar{x}} = \sqrt{\frac{\Sigma(\bar{x}_i-\mu)^2}{N}}$$

$$= \sqrt{\frac{(18-21)^2+(19-21)^2+\cdots+(24-21)^2}{16}}$$

$= 1.58$

σ Value 2.236

always $\sigma_{\bar{x}} < \sigma$ values

Comparing the population with its Sampling distribution

Polulation

$N=4$

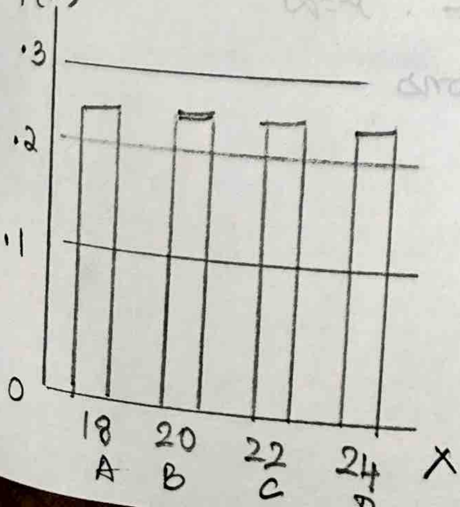
$\mu=21$ $\sigma=2.236$

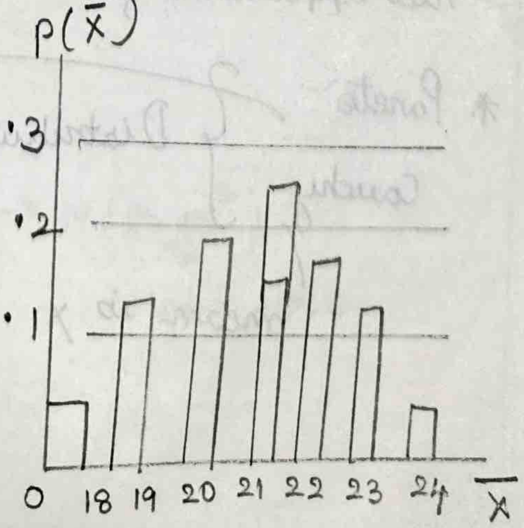
Sample means distribution

$n=2$

$\mu_{\bar{x}}=21$, $\sigma_{\bar{x}}=1.58$

fig (Graphs)





Even if you're not
normal...
... the average...
... is normal!!!

* CLT [Central Limit Theorem]

Statement:

let x be an independent identically distributed random Variable with finite Polulation mean μ and finite population Variance σ^2 then a random Variable Converges in a distribution to the Standard normal Variables as n goes to infinity with mean μ and Variance σ^2/n

Here n is the Size of Sample from given distribution

no matter what is the Shape of the population is distribution of Sampling distribution will be normal if n is Very large.

Limitations:

not applicable for

* Pareto $\quad \because \mu = \infty$

Cauchy } Distributions

$\because$ mean is $\times$

Let us assume X is a random Variable with finite Population mean and finite Population Variance

$X(\mu, \sigma^2)$

I am Collecting m Samples

$S_1, S_2, S_3 \ldots S_m$

Size of each Sample is $n = 30$, my Sample Size

Let me assume I am Collecting 1000 Samples

$n = 30$, $m = 1000$

$m \times n = 1000 \times 30$

$= 30K$ Samples

For each of the Sample, let us Calculate mean

$$\begin{array}{cccc} S_1, & S_2, & S_3, & \ldots S_m \\ \downarrow & \downarrow & \downarrow & \downarrow \\ \bar{x}_1 & \bar{x}_2 & \bar{x}_3 & \bar{x}_m \end{array}$$

Distribution of $\bar{x}_i$, $i = 1$ to m

Distribution of $\bar{x}_i$ = Sampling distribution of Sample mean

Now Central Limit theorem Says

$\bar{x}_i$ is distributed with Gaussian Normal distribution with mean μ and Variance $\frac{\sigma^2}{n}$

as $n \to \infty$

i.e $\bar{x}_i \longrightarrow N\left(\mu, \frac{\sigma^2}{n}\right)$

as $n \to \infty$

* Sampling distribution Properties.

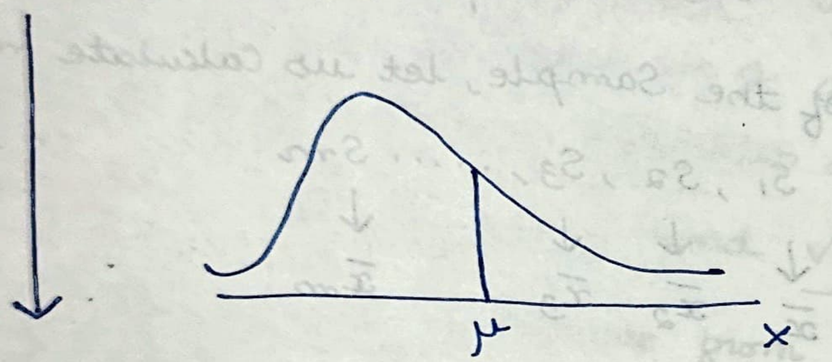
Central Tendency

$\mu_{\bar{x}} = \mu$

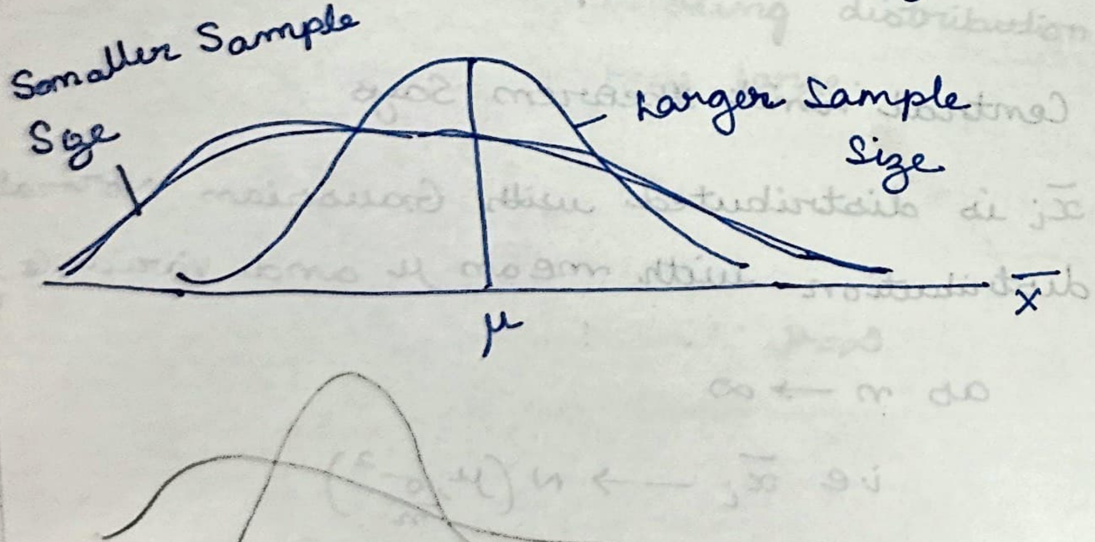
Variation

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

* Polulation distribution:



* Sampling distribution (becomes normal as n increases)



dat

21/11/2021.

① Suppose a large population has mean $\mu = 8$ and Standard deviation $\sigma = 3$. Suppose a random Sample of Size $n = 36$ is Selected.
what is the probability that the Sample mean is between 7.8 and 8.2?

$n \geq 25$

Sol: $\because n = 36$ ✓ large Size $\rightarrow$ CLT is applicable

Even if the Population is not normal distributed, the central theorem can be used
$(n > 25)$

$\mu_{\bar{x}} = \mu$

$\mu_{\bar{x}}$ ↓ Sample mean

μ ↓ population mean

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

So the sampling distribution of $\bar{x}$ is approximately normal

$-0.5 < Z < +0.5$

$\because$ 7.8   8.2

with mean $\mu_x = 8$

and Standard deviation

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0.5$$

$$P(7.8 < \mu_x < 8.2) = P\left(\frac{7.8 - 8}{3/\sqrt{36}} < \frac{\mu_x - \mu}{\sigma/\sqrt{n}} < \frac{8.2 - 8}{3/\sqrt{36}}\right)$$

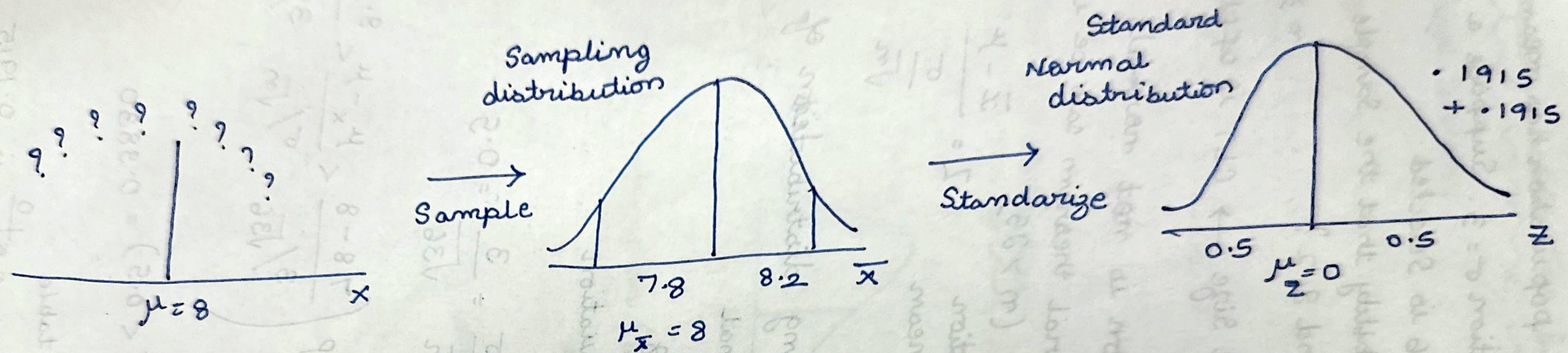
0.1915

$= P(-0.5 < Z < 0.5) = 0.3830$

$= \because$ from Z table

Value of $0.5 + 0$ is 0.1915

$[\because -0.5$ to 0 0.5 to $0]$ 2×0.1915



? ? ? ? ? ? ? ? ?
$\mu = 8$
x
Sample
Sampling distribution
7.8
8.2
$\bar{x}$
$\mu_{\bar{x}} = 8$
Standarize
Standard Normal distribution
.1915
+ .1915
0.5
0.5
$\mu_z = 0$
Z

⁂ MLE :

Maximum Likelihood Estimation :

→ In Statistics, MLE is a method of estimating the parameters of probability distribution by maximizing a likelihood function, so that under the assumed statistical model is most probable. It is Generally a function defined over the Sample Space.

→ It determines value for the parameters of a mode. Parameter values are found such that they maximize the likelyhood that the process described by the model.

① Eg: Suppose that x is a discrete Random Variable with a following Probability mass function where $0 < \theta < 1$ is a Parameter. The following 10 independent Observations

X	0	1	2	3
P(x)	$\frac{2\theta}{3}$	$\frac{\theta}{3}$	$\frac{2(1-\theta)}{3}$	$\frac{1-\theta}{3}$

PMF

θ is Parameter.

from such a distribution (3, 0, 2, 1, 3, 2, 1, 0, 2, 1) what is the maximum likelihood Estimation for θ?

Sol:

Sol: Since the Sample is (3, 0, 2, 1, 3, 2, 1, 0, 2, 1),

the likelihood is $L(\theta) = P(x=3)\, P(x=0)\, P(x=2)\cdot$

$P(x=1)\, P(x=3)\, P(x=2)$

$P(x=1)\, P(x=0)\, P(x=2)$

$P(x=1)$

Substituting from the probability distribution given above

$$L(\theta) = \prod_{i=1}^{n} P(x_i|\theta) = \left(\frac{2\theta}{3}\right)^2 \left(\frac{\theta}{3}\right)^3 \left(\frac{2(1-\theta)}{3}\right)^3 \left(\frac{1-\theta}{3}\right)^2$$

Clearly the likelyhood function $L(\theta)$ is not easy to maximize

$$\left[\begin{array}{l} f'(x) = 0 \\ f''(x) < 0 \begin{cases} \text{maxima} \\ \text{minima} \end{cases} \end{array}\right] \text{recall}$$

$d(uv) = uv' + vu'$ — differentiation is very complex

So, apply log

$l(\theta) = \log_e L(\theta)$

$$\begin{cases} \log_e x^m = m \log_e x \\ \log_e mn = \log_e m + \log_e n \end{cases}$$

$$= 2\left[\log_e\left(\frac{2}{3}\right) + \log_e \theta\right] + 3\left[\log_e\left(\frac{1}{3}\right) + \log_e \theta\right]$$

$$+ 3\left[\log_e\left(\frac{2}{3}\right) + \log_e(1-\theta)\right] + 2\left[\log_e\left(\frac{1}{3}\right) + \log_e(1-\theta)\right]$$

Constants

$l(\theta) = C + 5\log\theta + 5\log(1-\theta)$

where C is a Constant which does not depend on θ.

It can be seen that log likelihood function is easier to maximize. Compare to likelihood function.

$$l(\theta) = C + 5\log\theta + 5\log(1-\theta)$$

$$\frac{d\,l(\theta)}{d\theta} = 0 + \frac{5}{\theta} + \frac{5}{1-\theta}(0-1) \quad \left[\frac{d}{d\theta}\theta = 1\right]$$

$$\left[\frac{d}{dx}(\log x) = \frac{1}{x}\right] \quad \frac{d\,l(\theta)}{d\theta} = \frac{5}{\theta} - \frac{5}{1-\theta} \Rightarrow \frac{d}{d\theta}l(\theta) = 0$$

$$\frac{d\,l(\theta)}{d\theta} = 0 \Rightarrow \frac{5}{\theta} - \frac{5}{1-\theta} = 0$$

$$\frac{5(1-\theta) - 5\theta}{\theta(1-\theta)} = 0$$

$$5 - 5\theta - 5\theta = 0$$

$$10\theta = 5$$

$$\boxed{\theta = 0.5}$$

* Stochastic process:

Mathematically, a Stochastic process is a set of random variables $\{x_t\}$ or $\{x(t)\}$ depending on some real Parameter like time t. These are also known as random processes or random functions.

[Random process/stochastic process are same]

Eg: 1) Queuing

2) unbiased die

X_n outcome of n^{th} throw

$\{X_n \mid n \geq 1\}$

$\downarrow$

family of Random variables

$x_1 = 1$ to 6

$x_2 = 1$ to 6

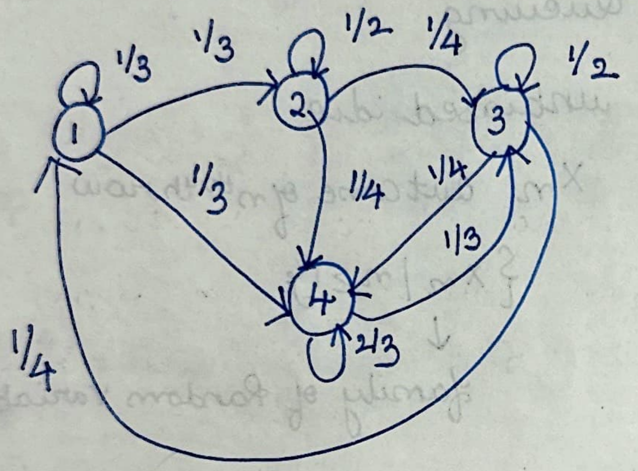
3) X_n is maximum number shown in first n rows.

2/2/2021:

(i) Determine if the following transition matrix is ergodic markov chain?

$$
\text{Present States}\quad
\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}
\begin{array}{c}
\begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\
\begin{bmatrix}
1/3 & 1/3 & 0 & 1/3 \\
0 & 1/2 & 1/4 & 1/4 \\
1/4 & 0 & 1/2 & 1/4 \\
0 & 0 & 1/3 & 2/3
\end{bmatrix}
\end{array}
\rightarrow \text{Probabilities}
$$

Sol: The transition diagram for the given transition matrix.



✓ In the above transition diagram, State 1 can go to any other State directly except to State 3.

→ To go to State 3, one must go from State 1 to State 2 and then to State 3.

→ It is Possible for State 2 to directly to all other states except to State 1.

→ now from State 2 one can reach State 1 in two ways. one way is from State 3 to State 1 and the other way is from State 4 to State 3 and then State 1.

→ From State 4 it is Possible to go to State 3 and State 4 · except to State 1 and State 2.

★ { Hence the given Matrix is an Ergodic MARKOV CHAIN }

② A training process is considered as two State markov chain. If it rains, it is considered to be in State 0 and it does not rain, the chain is in the state of 1. The transition Probability of the markov chain is defined by $P = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$. find the Probability that it will rain for 3 days from today assuming that it is raining today. Assume that the mutual probabilities of State 0 and State 1 as 0.4 and 0.6 respectively.

Sol:

$$P = \begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix} \end{array}$$ → 2 State markov chain

$2^2 = 4$

$$P^2 = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$$

$$= \begin{bmatrix} 0.44 & 0.56 \\ 0.28 & 0.72 \end{bmatrix}$$

$P^2 \cdot P \Rightarrow \begin{bmatrix} 0.44 & 0.56 \\ 0.28 & 0.72 \end{bmatrix} \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$

$P^3 = P^2 \cdot P$

[∵ $P^{(3)}$ → 3 days]

$$= \begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} 0.376 & 0.624 \\ 0.312 & 0.688 \end{bmatrix} \end{array}$$

The probability that it will rain on 3rd day given that it will rain today is 0.376

(3) The number of units of an item that are withdrawn from inventory on a day-to-day basis is a markov chain process in which requirements for tomorrow depend on today's requirements. A one-day-transition matrix is given below.

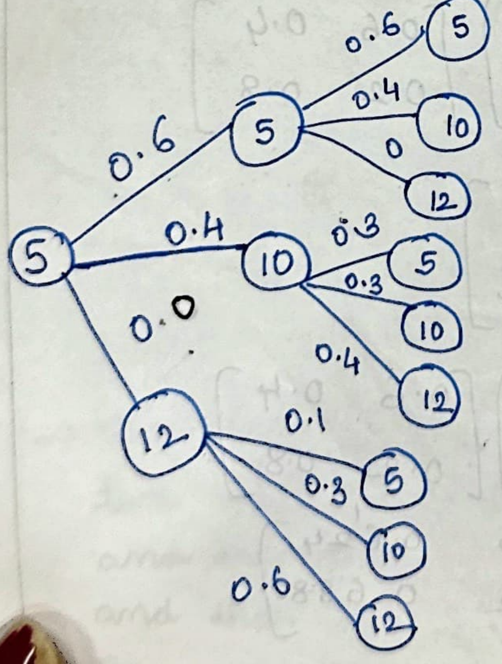
3/2/21

$$\begin{array}{cc} & \text{Tomorrow} \\ & \begin{array}{ccc} 5 & 10 & 12 \end{array} \\ \text{Today} \begin{array}{c} 5 \\ 10 \\ 12 \end{array} & \begin{bmatrix} 0.6 & 0.4 & 0.0 \\ 0.3 & 0.3 & 0.4 \\ 0.1 & 0.3 & 0.6 \end{bmatrix} \end{array}$$

(i) Construct a tree diagram showing inventory requirements on two consecutive days.

(2) Develop a two-day transition matrix.

Sol: (I) The inventory requirements on two consecutive days is represented by a tree diagram as shown below



tree diagram (must be same line)

$P_{11}^{(2)} = (0.6)(0.6) + (0.4)(0.3) + 0(0.1)$

$= 0.36 + 0.12 + 0$

$P_{11}^{(2)} = 0.48$

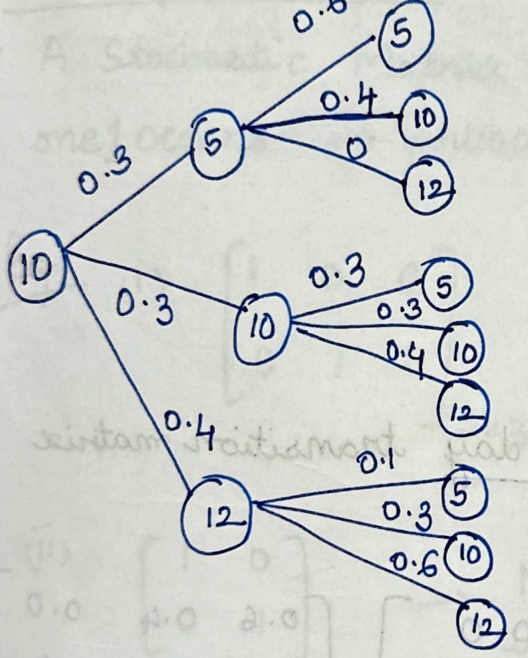
$P_{12}^{(2)} = (0.6)(0.4) + (0.4)(0.3) + 0(0.3)$

$= 0.24 + 0.12 + 0$

$P_{12}^{(2)} = 0.36$

$P_{13}^{(2)} = (0.6)(0) + (0.4)(0.4) + 0(0.6)$

$P_{13}^{(2)} = 0.16$



$P_{23}^{(2)} = (0.3)(0) + (0.3)(0.4) + (0.4)(0.6)$

$= 0 + 0.12 + 0.24$

$P_{23}^{(2)} = 0.36$

$P_{21}^{(2)} = (0.3)(0.6) + (0.3)(0.3) + (0.4)(0.1)$

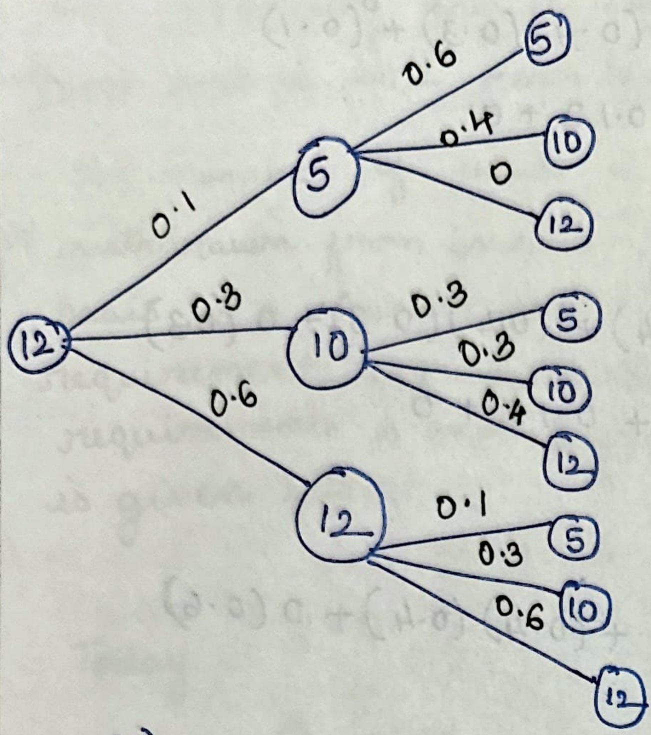
$= 0.18 + 0.9 + 0.04$

$P_{21}^{(2)} = 0.31$

$P_{22}^{(2)} = (0.3)(0.4) + (0.3)(0.3) + (0.4)(0.3)$

$= 0.12 + 0.9 + 0.12$

$P_{22}^{(2)} = 0.33$



$$P_{31}^{(2)} = (0.31)(0.6) + (0.3)(0.3) + (0.6)(0.1)$$

$$P_{31}^{(2)} = 0.21$$

$$P_{32}^{(2)} = 0.31$$

$$P_{33}^{(2)} = 0.48$$

(II) Develop a <u>two-day</u> transition matrix

$$P^2 = \begin{bmatrix} 0.6 & 0.4 & 0.0 \\ 0.3 & 0.3 & 0.4 \\ 0.1 & 0.3 & 0.6 \end{bmatrix} \begin{bmatrix} 0.6 & 0.4 & 0.0 \\ 0.3 & 0.3 & 0.4 \\ 0.1 & 0.3 & 0.6 \end{bmatrix}$$

$$= \begin{bmatrix} 0.48 & 0.36 & 0.16 \\ 0.31 & 0.33 & 0.36 \\ 0.21 & 0.31 & 0.48 \end{bmatrix}$$

* STOCHASTIC MATRIX :

→ A Stochastic Matrix (must be Square matrix) is a random matrix with non-negative elements (≥ 0) and unit row Sums.

* Regular Matrix :

→ A Stochastic Matrix P is Said to be regular if all the entries of Some Power P^m are Positive.

* Non Regular Matrix:

→ A Stochastic Matrix P is not regular if one 1 occurs in principle Mean diagonal.

Eg: (i) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}_{2\times 3}$ → Not a Square matrix ∴ It is not Stochastic

(ii) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{2\times 2}$ → Matrix is Square Matrix with non negative entries and Sum of the elements in each row is equal to (=) 1

∴ The matrix is Stochastic.

(iii) $\begin{bmatrix} 0 & 1 \\ 1/3 & 1/4 \end{bmatrix}_{2\times 2}$ → The matrix is Square matrix but Sum in each row is not equal to 1. So it is not Stochastic.

4) $\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$ Stochastic

5) $\begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}$ The matrix is not Stochastic because it contains negative elements.

6) $\begin{bmatrix} 0 & 2 \\ 1/4 & 1/4 \end{bmatrix}$ The matrix is Square but Sum in each in row is not equal to 1

$\therefore$ It is not Stochastic process

Q check whether the following Markov chain is Regular or Not.

(i) $P = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 1/2 \end{bmatrix}_{4\times 4}$

(ii)

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}$$

$$P^3 = P^2 \cdot P \Rightarrow \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix} = P$$

$P^3 = P$

$P^4 = P^3 \cdot P = P^2$

$P^6 = P^2$

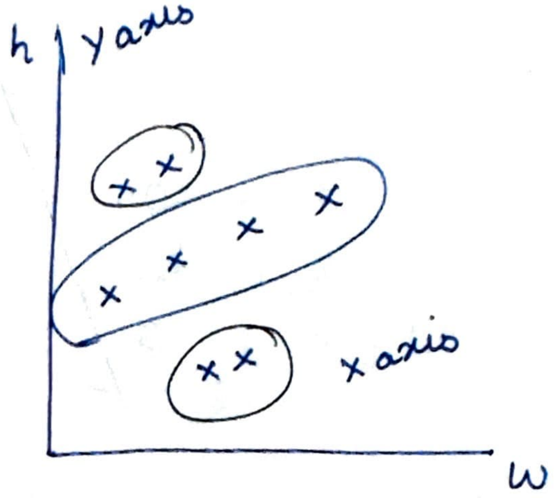
$P^5 = P$

$$\left\{ \begin{aligned} P^{2n} &= P^2 \\ P^{2n+1} &= P \end{aligned} \right\}$$

$\therefore$ Not a Regular Matrix

4/1/2021

★ Linear Regression: Geometrical Interpretation



usually height ↑ weight ↑

Exceptions : Outliers

↙ ↘

Obesity very thin

Plot the Graph for majority (max) Points.

minimize the errors:

Linear Regression:

* mostly applicable on Data Analysis, Building models

Motivation:

* purpose is to build a functional relationship (model) between dependent Variable(s) and independent Variable (S).

Example:

Business: what is the effect of price on Sales? (can be ~~fix~~ used to fix the Selling price of an item).

↓ Result:

To Increase market Sale.

Engineering: Can we infer difficult to measure properties of a product from other easily measured Variables? (mechanical Strength of a polymer from temperature, Viscosity or other process Variables) - also known as Soft Sensor.

↓

→ Develop a model

→ use the model to predict the mechanical Strength, temp., Viscosity → also Known as Soft Sensor

Software Sensor in literature.

~~This model is used in practise to~~ ~~measure~~ infer the Variables

* This model is used in practise to infer the values of Variables which are difficult to measure using an instrument.

Purpose:

(i) we are building the model for given Purpose

(ii) purpose is defined depending the area that you are working

* Regression

→ ~~one~~ one of the widely used Statistical technique.

→ Dependent Variables also Known as Response Variable, Regress and, predicted Variable, Output Variable - denoted as Variable/S y.

→ Variable whose Output is designed to predict based on model. Symbolic way is output y.

→ Independent Variable(s) also known as predictor/ Regressor/ Exploradatory Variables or input Variable indicated by x.

Eg: For a plant growth, if we apply fertilizers, fertilizer is independent Variable.

plant Growth is Dependent Variable.

Methods & Types of Regression:

classification of Regression Analysis

* univariate: one dependent and one independent Variable

* multivariate: many dep. and many independent Variables.

* Linear vs Non Linear

→ Linear: Relationship is linear between dependent and independent Variables.

→ Non Linear: Relationship is nonlinear between dependent and independent Variables.

* Simple vs Multiple:

→ Simple: one dependent and one independent Variable (SISO)

→ multiple: one dependent and many independent Variables (MISO)

Linear Regression methods:

→ Simple linear Regression

→ Multiple
→ Ridge
→ Principal component
} Regression.

→ Lasso

→ partial least Squares

Non linear Regression methods:

★ Polynomial
★ Spline
} Regression

★ Neural networks.

Simple linear Regression Model:

What is Regression?

→ we can estimate value of one variable with value of another variable which is known.

The statistical method which helps us to estimate the unknown value of one variable from known value of related variable is called Regression.

Line of Regression:

→ The line described in average relationship b/w 2 Variables.

→ now-a-days we call this as Estimating line.

Uses of Regression:

1. Estimate relation b/w 2 Variables.

Eg: like 2 Economic variables like Income and Expectation.

2. Highly valuable tool in Business and economics.

3. Highly used in prediction.

4. we can find relation b/w Coefficient of corelation and Coefficient of determination.

5. useful in Statistical Estimator in demand curves, Supply curves, function, Cost function, Consumption function. etc

Comparision between Corelation and Regression.

176

1. Degree of Covariability

→ measures of degree of Covariability b/w 2 Variables by regression established by functional relationship is regression (functional established functional relation b/w dependent and independent Variables So that (farmer) can be ~~dependent~~ predicted for a given value of the (Letter?) In Corelation (analysis) x and y are random Variables whereas in Regression x is R.V and y is fixed measure.

corelation ⟶ Relative Measure.

Absolute Measure ⟶ Regression.

Regression Model:

Simple Linear Regression Model:

The equation that describes how y is related to x and Error term

Eg:
$$\mu = \beta_0 + \beta_1 x + \epsilon$$

β_0, β_1 are parameters

ϵ is Random variable called error term.

$E(y) = \beta_0 + \beta_1 x$ [no error term]

$y = \beta_0 + \beta_1 x + \epsilon$

$E(y)$ is Expectation of y

In calculating β_1 we minimized the errors prediciting y $E(X)$ i.e mean value of y or $E(Y)$

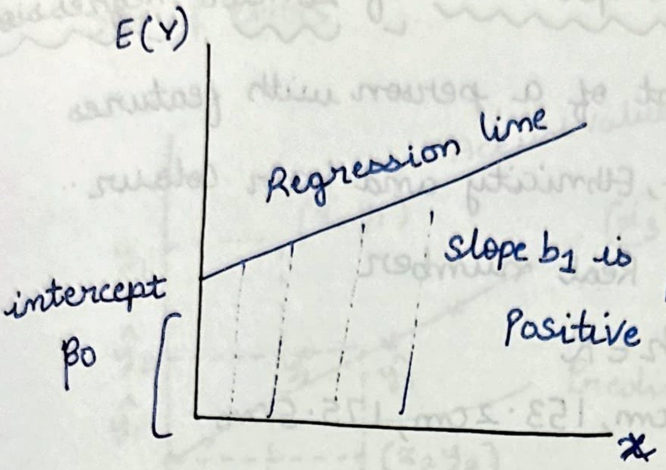
Graph is Straight line

β_0 is y intersect cept of Regression Line

β_1 is Slope of the Regression line.

$E(Y)$ is Expected value of y for an given x value.

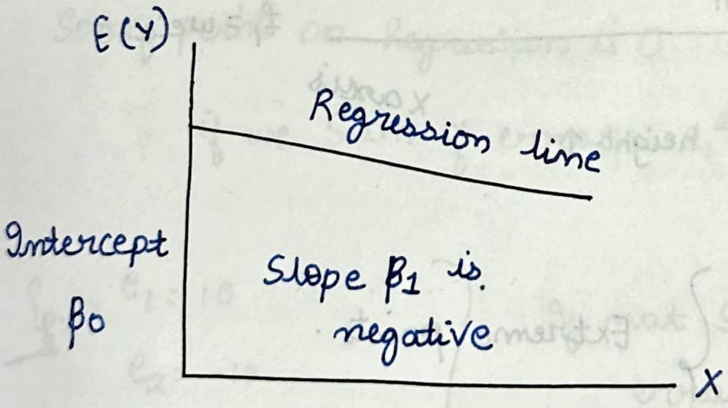
Positive Linear Relationship



$x \uparrow$, y is also $\uparrow$
~~or~~ $E(y)$

$E(Y) = \beta_0 + \beta_1 x$

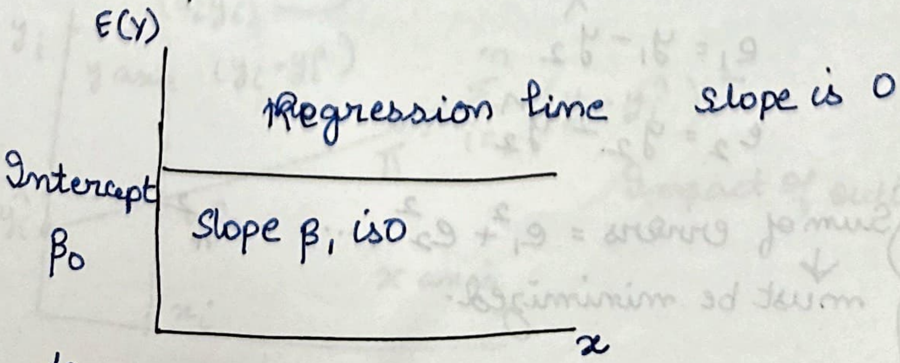
Negative Linear Relationship



$x \uparrow$ y is $\downarrow$

$\because$ Slope is negative

No Relationship.



for any value of x, $E(Y)$ are Same

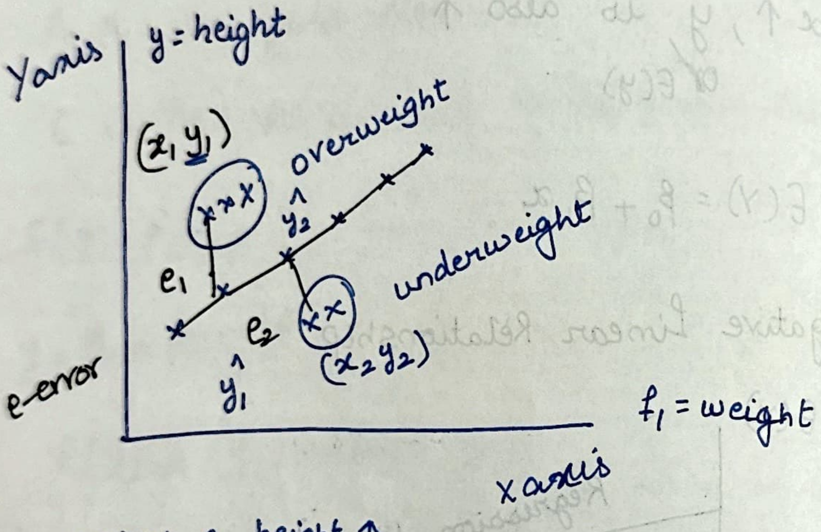
Geometrical Intepretation of Linear Regression

predict height of a person with features weight, gender, Ethnicity and hair Colour.

Height is a Real number

$h \in R$

Eg:- 165.6 cm, 153.2 cm, 175.5 cm



weight ↑ height ↑

Exceptions:

overweight
under
} Extreme points.

Line has to be drawn on all the given points.

So error is minimized.

Error = Actual Value − Predicted Value

$e_1 = y_1 - \hat{y}_2$

$e_2 = y_2 - \hat{y}_2$

Sum of errors $= e_1^2 + e_2^2 + \dots + e_n^2$

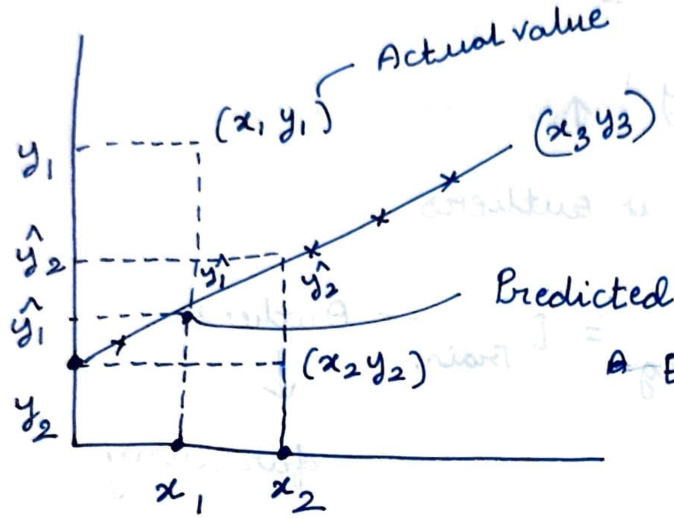
↓

must be minimized.

which ever line is minimized is least Square method [is Principle].

↓

Sum of errors



Predicted value

Error = A.V − P.V

$= y_1 - \hat{y}_1$

$e_1 = y_1 - \hat{y}_1 = +ve$

$e_2 = y_2 - \hat{y}_2 = -ve$

error for x_2

above } regression line → +ve

below } → −ve

Some point on Regression is 0

if we Sum of errors $= e_1^2 + e_2^2 + e_3^2 + \dots + e_n^2$

→ +ve

Eg:

$e_1 = 10$

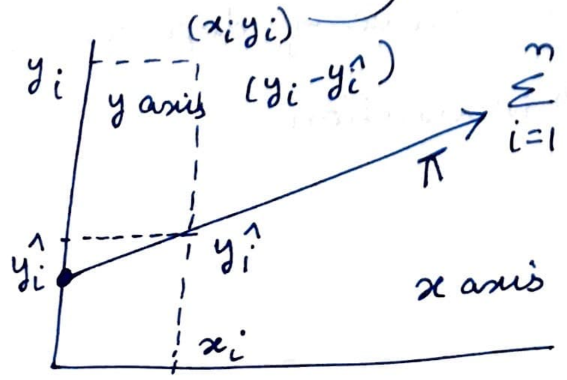
$e_2 = -10$

$e_1 + e_2 = 10 + (-10)$

$= 0$

but error exists above line

If not Sum of errors errors may be nullified.



$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Impact of outliers.

For trained data (we find)

Eg: $y = mx + c$

$y_i - \hat{y}_i \uparrow$

remove as outliers

$D_{\text{Training Data set}} = D_{\text{Train}} - \text{Outliers}$

$\downarrow$

far away

Repeat this procedure.

This is Random Sampling process.

Least Squares Method

· Least Square Criterion

$$\min \sum (y_i - \hat{y}_i)^2$$

$\hat{y}_i$ → predicted value

y_i → Actual Value

where:

y_i = Observed value of the dependent variable for the ith observation.

$\hat{y}_i$ = estimated value of dependent variable for ith observation.

In Least Squares method,

(i) we first find errors

(ii) Square of the errors.

(iii) Sum of the Squared errors.

$$\min \sum (y_i - \hat{y}_i)^2$$

$$i=1 \quad (y_1-\hat{y}_1)^2 + (y_2-\hat{y}_2)^2 + \cdots (y_n-\hat{y}_n)^2$$

with $i=2$ over the second term and $i=n$ over the last term.

Squared Error (SE) = Σy

$$y = mx + b$$

$$\Rightarrow (y_1 - (mx_1+b))^2 + (y_2 - (mx_2+b))^2 + \cdots\cdots$$

$$(a-b)^2 \qquad (y_n - (mx_n + b))^2$$

$$\hat{y}_1 = mx + b$$

$$\hat{y}_1 = mx_1 + b$$

error = AV − PV

m, b are parameters such that Squared error is minimum.

$$(y_1-(mx_1+b))^2 \Rightarrow y_1^2 + \underbrace{(mx_1+b)^2}_{(a+b)^2} - 2y_1(mx_1+b)$$

$$y_1^2 + (m^2x_1^2 + b^2 + 2mx_1b) - 2y_1x_1m - 2by_1$$

Slly for $(y_2 - (mx_2+b))^2$

$$SE = (y_1 - (mx_1 + b))^2 + (y_2 - (mx_2 + b))^2 + \ldots\ldots$$

$$(y_n - (mx_n + b))^2$$

$$= y_1^2 - 2y_1(mx_1 + b) + (mx_1 + b)^2 +$$

$$y_2^2 - 2y_2(mx_2 + b) + (mx_2 + b)^2 + \ldots\ldots$$

$$+ y_n^2 - 2y_n(mx_n + b) + (mx_n + b)^2$$

$$= y_1^2 - 2x_1y_1m - 2y_1b + m^2x_1^2 + 2mx_1b + b^2 +$$

$$y_2^2 - 2x_2y_2m - 2y_2b + m^2x_2^2 + 2mx_2b + b^2 +$$

$$\ldots\ldots + y_n^2 - 2x_ny_nm - 2y_nb + m^2x_n^2 +$$

$$2mx_nb + b^2$$

Grouping terms

$$= (y_1^2 + y_2^2 + \ldots + y_n^2) - 2m(x_1y_1 + x_2y_2 + \ldots + x_ny_n)$$

$$- 2b(y_1 + y_2 + \ldots + y_n) + m^2(x_1^2 + x_2^2 + \ldots + x_n^2)$$

$$+ 2mb(x_1 + x_2 + \ldots + x_n) + (b^2 + b^2 + \ldots + b^2)$$

$$= n\overline{y^2} - 2mn\overline{xy} - 2bn\bar{y} + m^2 n\overline{x^2} +$$

$$2mbn\bar{x} + nb^2$$

Proof:

$$\bar{y}_2 = \frac{y_1^2 + y_2^2 + \ldots\ldots + y_n^2}{n}$$

$$ny_2 = \underbrace{y_1^2 + y_2^2 + \ldots + y_n^2}$$ Sub

$$\frac{\overline{xy}}{1} = \frac{x_1 y_1 + x_2 y_2 + \cdots + x_n y_n}{n}$$

$$n\overline{xy} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

$$\frac{\bar{y}}{1} = \frac{y_1 + y_2 + \cdots + y_n}{n}$$

$$n\bar{y} = y_1 + y_2 + \cdots + y_n$$

$$\frac{\overline{x^2}}{1} = \frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}$$

$$x_1^2 + x_2^2 + \cdots + x_n^2 = nx^2$$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$n\bar{x} = x_1 + x_2 + \cdots + x_n$$

Standard error $= n\overline{y^2} - 2mn\overline{xy} - 2bn\bar{y} +$

[SE] $m^2 n \overline{x^2} + \cancel{2b} \, 2mbn\bar{x} + nb^2$

$y = mx + b$

m, b are parameters

find m, b

differentiate partially

$\frac{d}{dx}(x^n) = nx^{n-1}$

$\frac{d}{dx}(x) = 1$

$\frac{\partial(SE)}{\partial m} = -2n\bar{x}\bar{y} + 2mn\overline{x^2} + 2bn\bar{x} = 0$

$\frac{\partial(SE)}{\partial m} = -2n\overline{xy} + 2mn\overline{x^2} + 2bn\bar{x} = 0$

[2n common]

$= -\overline{xy} + m\overline{x^2} + b\bar{x} = 0$

$m\overline{x^2} + b\bar{x} = \overline{xy}$ [$\div$ with $\bar{x}$]

$= m\frac{\overline{x^2}}{\bar{x}} + b = \frac{\overline{xy}}{\bar{x}}$

$mx + b = y$

one Point $\left(\frac{\overline{x^2}}{\bar{x}}, \frac{\overline{xy}}{\bar{x}}\right)$

$SE = n\overline{y^2} - 2mn\,\overline{xy} - 2bn\bar{y} + m^2 n\overline{x^2} +$
$\times \qquad \times \qquad \times$
$2mbn\bar{x} + nb^2$

differentiating partially w.r.t b [constant 0]

$$\frac{\partial(SE)}{\partial b} = -2n\bar{y} + 2mn\bar{x} + 2nb = 0$$

$= -\bar{y} + m\bar{x} + b = 0$

$y = mx + b$ [passing through $\bar{x}, \bar{y}$]

$\bar{y} = m\bar{x} + b$

another point $(\bar{x}, \bar{y})$ $\qquad \frac{dy}{dx} = 0$

$\frac{d^2y}{dx^2} < 0$

$\therefore$ line must pass through $\left(\frac{\overline{x^2}}{\bar{x}}, \frac{\overline{xy}}{\bar{x}}\right)$ and $(\bar{x}, \bar{y})$

we get max. value

SLOPE: $\left(\frac{\overline{x^2}}{\bar{x}}, \frac{\overline{xy}}{\bar{x}}\right), (\bar{x}, \bar{y})$
$\qquad x_1 \quad y_1 \qquad x_2 \quad y_2$

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{\bar{y} - \frac{\overline{xy}}{\bar{x}}}{\bar{x} - \frac{\overline{x^2}}{\bar{x}}} = \frac{\bar{x}\bar{y} - \overline{xy}}{(\bar{x})^2 - \overline{x^2}}$$

$$m = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\overline{xy})^2} = \frac{Cov(x, y)}{Var(x)}$$

$Cov(x, y) = E(x - \bar{x})(y - \bar{y})$

$= E(xy - x\bar{y} - \bar{x}y + \bar{x}\bar{y})$

$= E(xy) - \bar{y}\bar{x} - \cancel{\bar{x}\bar{y}} + \cancel{\bar{x}\bar{y}}$

$= E(xy) = \overline{xy} - \bar{x}\bar{y}$

$$Var(x) = E(x-\bar{x})^2$$

$$= E(x^2 - 2x\bar{x} + \bar{x}^2)$$

$$= E(x^2) - 2(\bar{x})^2 + (\bar{x})^2$$

$$= \overline{x^2} - (\bar{x})^2$$

$$\therefore m = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2}$$

$\therefore$ x is Covariance of (x, y) and Variance of x

$$y = mx + b$$

$$y = b_0 + b_1 x$$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$$

y intercept for estimated Linear Regression

$$\bar{y} = b_0 + b_1 \bar{x}$$ Parameters.

$(\bar{x}, \bar{y})$ $\quad b_0 = \bar{y} - b_1 \bar{x}$

188-191

Multiple Regression Model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots\cdots + \beta_x x_p + \epsilon$$

If more than 1 independent variable

Multiple Regression Equation

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots\cdots + \beta_x x_p$$

$\left[E(\epsilon) = 0 \right]$ $\beta_0, \beta_1, \beta_2, \ldots\ldots, \beta_p$ are unknown parameters

The Estimation process for Multiple Regression

To find unknown parameters from the population we collect Sample data

y is dependent variable

Sample data

x_1	x_2	......	x_p	y
⋮	⋮		⋮	⋮

↓

Compute the estimated multiple Regression model Equation

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots\cdots + b_p x_p$$

$b_0, b_1, b_2, \ldots\ldots b_p$ are Sample Statistics.

↳ $b_0, b_1, b_2 \ldots\ldots b_p$ provides the estimates of $\beta_0, \beta_1, \beta_2, \ldots\ldots \beta_p$

~~find~~ How can we decide $\beta_0, \beta_1, \beta_2, \ldots \beta_n$ are equal to 0 or not

→ By using Significant test

Similar to Linear Regression model.

Simple vs multiple Regression.

In Simple linear regression, b_0 and b_1 were the Sample Statistics used to estimate the parameters β_0 and β_1

Multiple regression parallels this Statistical inference process, with $b_0, b_1, b_2, \ldots, b_p$ denoting the Sample Statistics used to estimate the parameters $\beta_0, \beta_1, \beta_2, \ldots \beta_p$

$b_0 \rightarrow \beta_0$

Sample Statistics

$b_1 \rightarrow \beta_1$

~~$b_2 \rightarrow \beta_2$~~

$b_2 \rightarrow \beta_2$

$\vdots$

$b_p \rightarrow \beta_p$

Going to predict the population parameters with help of Sample Statistics

In Simple regression we have only 1 independent Variable

multiple → more than 1 independent Variable

~~Explo~~ Example: Trucking Company:

[Source: Statistics for Business and Economics, 2012, Anderson.]

As an illustration of multiple regression analysis, we will consider a problem faced by the Trucking Company.

A major portion of business involves deliveries throughout its local area.

To develop better work schedules, the managers want to estimate the total daily travel time for their drivers.

preliminary data for butler trucking.

	Driving assignment	x_1 = miles travelled	no. of deliveries	y = Travel Time (hours)
0	1	100	4	9.3
1	2	50	3	4.8
2	3	100	4	8.9
3	4	100	2	6.5
4	5	50	2	4.2
5	6	80	2	6.2
6	7	75	3	7.4
7	8	65	4	6.0
8	9	90	3	7.6
9	10	90	2	6.1

Independent Variable

dependent

x_1 miles travelled

↳ deliveries → Independent.

$x1$ miles travelled

Dependent Variable : travel_time

Independent Variables:

$x1$ = miles travelled

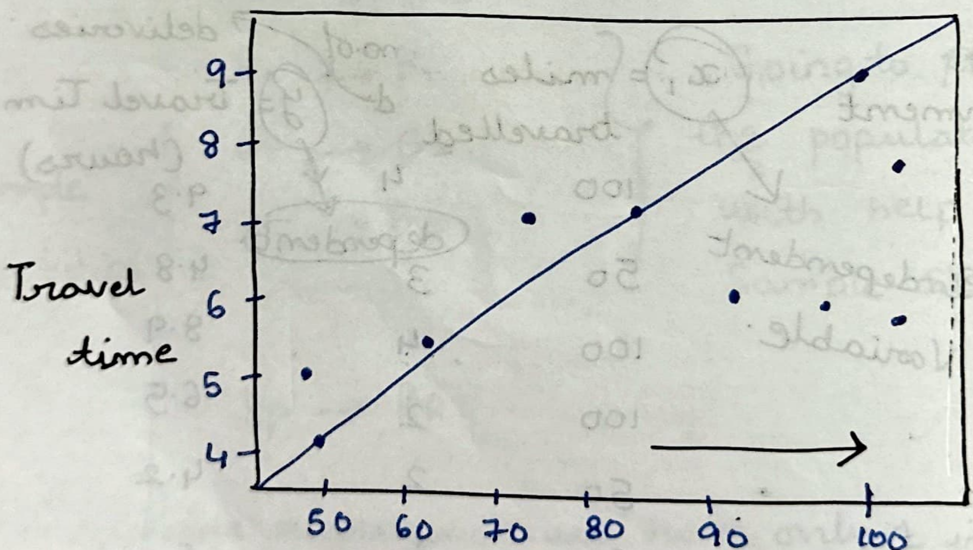
n_of_deliveries.

Let's understand the relation b/w $x1$ miles travelled and travel time.

we need to draw Scatter plot.

Scatter diagram of preliminary Data for Trucking x_1

Simple linear regression with miles travelled

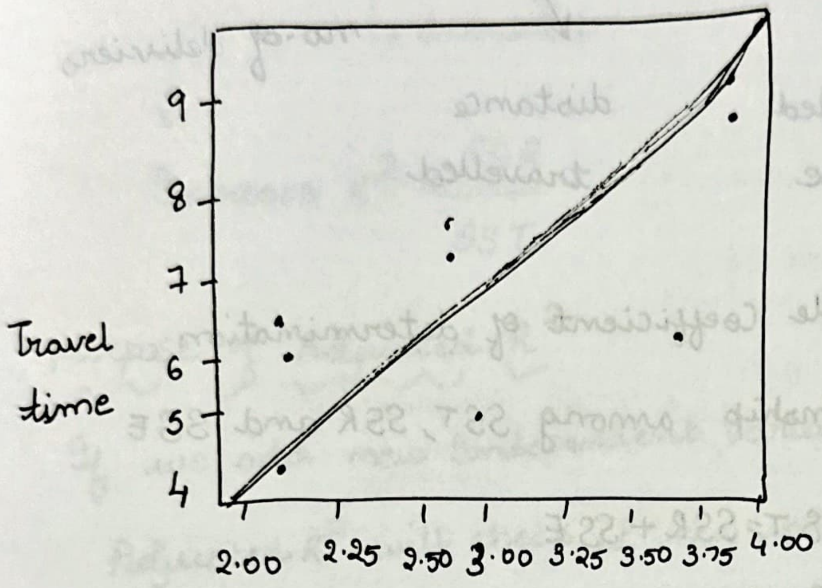


✓ Relation between one independent and one dependent Variable.

Travel time depends on miles travel

now consider no. of deliviers

Simple linear regression with number of deliveries.

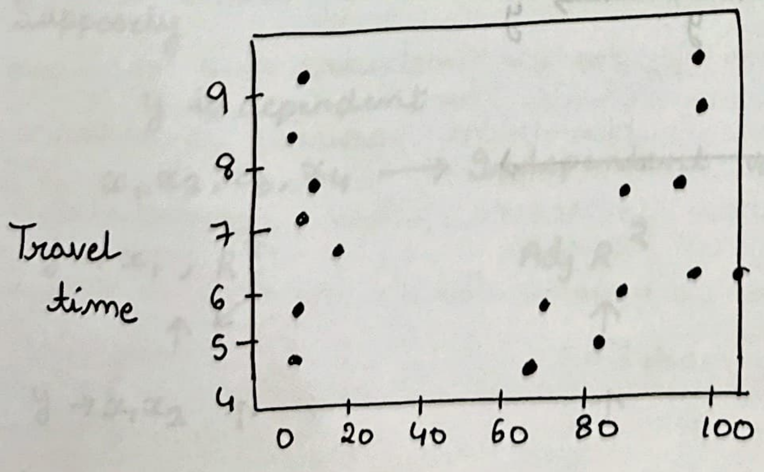


linear regression

There is a +ve corelation

? why corelation → If there is no corelation there is no relation between dependent and independent variable so regression model is not used in such cases.

Multiple regression.



Multiple regression:

Eg: $\hat{y} = -.869 + .0611 x_1 + .923 x_2$

Travelled time

distance travelled

no. of deliviers

multiple Coefficient of determination

Relationship among SST, SSR and SSE

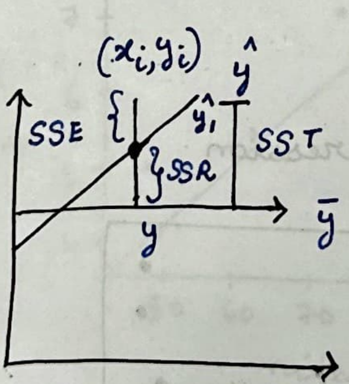
$$SST = SSR + SSE$$

where

SST = tatal Sum of Squares $= \Sigma (y_i - \bar{y})^2$

SSR = Sum of Squares due to regression $= \Sigma (\hat{y}_i - \bar{y})^2$

SSE = Sum of Squares due to error $= \Sigma (y_i - \hat{y}_i)^2$



Adjusted Multiple Coefficient of Determination:

$SST = SSR + SSE$

$SSR = SST - SSE$

↳ if decreased

Increase $R^2 = \frac{SSR}{SST}$

purpose of Adjusted R^2

If we add new Independent Variable

Adjusted R^2 will check if new independent Variable is explaining variable or noise Variable

n = number of Observations

P = denoting the no. of independent Variables.

Adjusted R^2

$$R^2_a = 1 - (1 - R^2)\frac{n-1}{n-P-1}$$

$$R^2_a = 1 - (1 - 0.904)\frac{10-1}{10-2-1} = 0.88$$

Supposely

y is dependent

$x_1, x_2, x_3, x_4 \rightarrow$ Independent variables

$y \rightarrow x_1$, R^2 ↑   Adj R^2 ↑

$y \rightarrow x_1 x_2$ ↑   ↑

$y \rightarrow x_1 x_2 x_3$ (x_3 is noise data) noisy value implies it will not explain y, and disturbs existing relation

R^2 $\uparrow$ Adj R^2 $\downarrow$

$y \to x_1\, x_2\, x_3$

If R^2 and Adj R^2 are Same we need not add Variables.

If R^2 is 0.9, Adj $R^2 = 0.3$ $\rightarrow$ we can add more values to fill the Gaps

Adj multiple Coefficient Vs multiple Coefficient:

Every time you add a independent Variable to a model, the R-Squared increases, even if the independent Variable is insignificant. It never declines. whereas Adjusted R-Squared increases only when independent Variable is Significant and affects dependent Variable.

In the table ~~above~~ below, adjusted r-Squared is maximum when we included two Variables. It declines when third Variable is added whereas r-Squared increases when we included third Variable. It means third Variable is insignificant to the model.

Variables	R-Squared	Adjusted R-Squared
1	67.5	67.1
2	85.9	84.2
3	88.9	81.7
	incr.	decreasing

Adj

Adjusted r-Squared can be negative when r-Squared is close to Zero.

Adjusted r-Squared Value always be less than or equal to r-Squared value.

Testing for Significance:

F-test

t-test

The F test is used to determine whether a Significant relationship exists between the dependent Variable and the Set of all the independent Variables; we will refer to the F test as the test for overall Significance.

If the f test Shows an overall Significance, the t test is used to determine whether each of the individual independent Variable is Significant.

A Seperate t test is Conducted for each of the independent Variables in the model; we refer to each of these t tests as a test for individual Significance.

F test:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon$$

The hypotheses for the F test involve the parameters of the multiple regression model.

$$H_0 = \beta_1 = \beta_2 = \dots = \beta_p = 0$$

H_a: one or more of the parameters is not equal to zero.

Test Statistic:

$$F = \frac{MSR}{MSE}$$

~~Relation~~ Rejection rule:

p Value approach   Reject H_0 if $P\text{-value} \leq \alpha$

Critical Value approach   Reject H_0 if $F \geq F_\alpha$

where

F_α is based on an F distribution with p degrees of freedom in the numerator and $n-p-1$ degrees of freedom in the denominator

$\beta_1 = 0$

There is no relation between x_1 and y

Coefficient of β_1 dependent Variable

$\beta_2 = 0$

no relation between x_2 and y

Eg:

$$MSR = \frac{SSR}{\text{degree of freedom}} = \frac{21.608}{2} = 10.804$$

$$MSE = \frac{SSE}{n-P-1} = \frac{2.299448}{10-2-1} = 0.328491$$

$$F = \frac{MSR}{MSE} = \frac{10.804}{0.328491} = 32.88$$

t test for individual Significance.

for any parameter β_i $H_0: \beta_i = 0$ — β_i

$H_a: \beta_i \neq 0$

Test Statistic $t = \frac{b_i}{S_{b_i}}$

$$t = \frac{b_i - \beta_i}{S_{b_i}} \quad [\text{if } \beta_i = 0]$$

Rejection Rule:

P: Value approach Reject H_0 if P-Value $\leq \alpha$

Critical Value approach Reject H_0 if $t \leq t_{\alpha/2}$ or $t \geq t_{\alpha/2}$

where $t_{\alpha/2}$ is based on a t distribution with $n-p-1$ degrees of freedom

t Test for individual Significance

$b_1 = 0.061135$ $S_{b_1} = 0.009888$

$b_2 = 0.9234$ $S_{b_2} = 0.2211$

$t = 0.661135/0.009888 = 6.18$

$t = 0.9234/0.2211 = 4.18$

Reject Ho if $t \leq -t_{\alpha/2}$ or $t \geq t_{\alpha/2}$

$t_{\alpha/2} = 2.262$

$\alpha = 5\% = 0.05$ $6.18 > 2.262$

$\alpha/2 = 0.025$ reject Ho in Case of 1st variable

$4.176 > 2.262$

In Case of 2nd variable Reject Ho

Logistic Regression:

Linear Regression → Independent Variable } Continous
→ Dependent Variable

→ categorical Data → we require Dummy Variable Regression

There may be a possibility that dependent variable is categorical data. In such cases we need Logistic Regression.

If Dependent Variable → Categorical

The essential difference between these two is that Logistic Regression is used when the dependent variable is binary in nature. In Contrast, Linear regression is used when the dependent variable is continous and nature of the regression line is (nature) linear

Dependent Variable in Regression applications; need of Logistic Regression

Dependent Variable
Gender: M or F
Success or Failure

buy or not buy

categorical

0, 1

Gender: male or female dependent variable

Quality of the product good or bad.

In linear regression

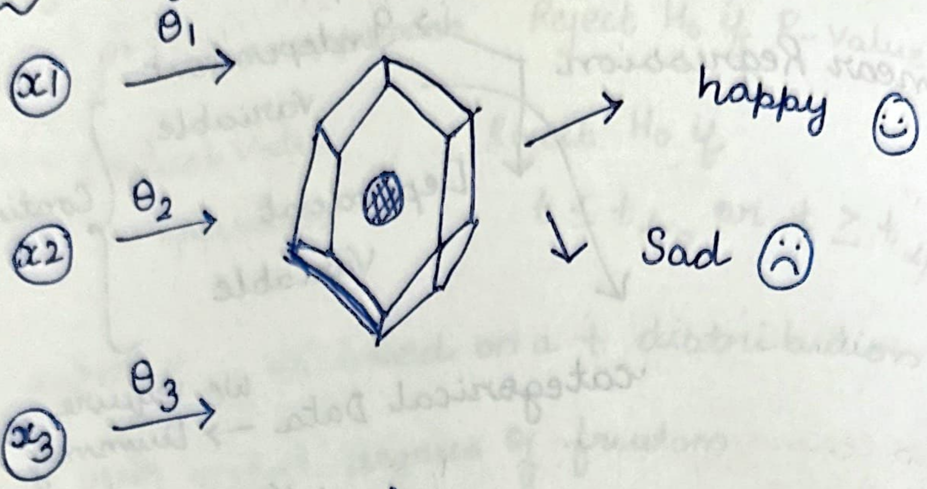
$y = a + b_1 x_1 + b_2 x_2$

independent

categorical data

↓ Dummy Variable Regression

Need of Logistic Regression.



based on product;

dependent Variable is categarical data

Approval of Credit Card:

approves Credit Card or not

y=1 bank approved

y=0 bank rejects

we can estime the probability of bank to approval of credit card

- Payoff Personal Loan:

will pay Loan or not.

Applying for job

dependent Variable: Getting a job / not getting a job

Example:

Let us Consider an application of Logistic regression involving a direct mail promotion being used by Simmons Stores.

Simmons owns and operates a national chain of women's apparel Stores.

5000 Copies of an expensive four-Color Sales catalog have been printed, and each catalog includes a Coupon that provides a \$50 discount on purchases of \$200 or more

The Catalogs are expensive and Simmons would like to send them to only those customers who have the highest probability of using the coupon.

Variance:

Management thinks that annual Spending at Simmons Stores (independent) and whether a Customer has a Simmons Credit Card are two Variables that might be helpful in predicting wheather a Customer who receives the Catalog will use the Coupon.

Simmons Conducted a pilot Study using a Random Sample of 50 Simmons Credit Card customers and 50 other Customers who do not have a Simmons Credit Card.

Simmons Sent the catalog to each of the 100 customers Selected.

At the end of a test period, Simmons noted whether the customer used the Couple or not?

Data (10 customer out of 100)

categorical Variable

customer	Spending (x_1)	Card (x_2)	Coupon (y)
1	2.291	1	0
2	3.215	1	0
3	2.135	1	0
4	3.924	0	0
5	2.528	1	0
6	2.473	0	1
7	2.384	0	0
8	7.076	0	0
9	1.182	1	1
10	3.345	0	0

Independent variables

dependent Variable

0 → not using coupon

Explanation of Variables:

Amount of each customer Spent last year at Simmons is Shown in thousands of dollars and credit card info coded as 1 ✓ 0 no

In the coupon Column, a 1 is recorded if the sampled Customer used the coupon and 0 if not.

Logistic Regression Equation:

* If the two values of the dependent variable y are Coded as 0 or 1, the value of $E(Y)$ in equation given below provides the probability that $y=1$ given a particular Set of values for the independent variables $x_1, x_2, \ldots, x_p$.

$$E(y) = \frac{e^{\overbrace{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}^{\infty}}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}}$$

y is Coupon

$P(y=1/x_1, x_2, \ldots x_p)$

$\hookrightarrow$ independent Variables.

$$\frac{e^{\infty}}{1+e^{\infty}} = \frac{\infty}{\infty} \quad (\because \text{ indeterminant value})$$

$$\downarrow \frac{\cancel{e^{\infty}}}{\cancel{e^{\infty}}(1+e^{-\infty})} = \frac{1}{1+0} = 1$$

$\downarrow$ max value

$$\frac{e^{-\infty}}{1+e^{-\infty}} = \frac{e^{-\infty}}{e^{-\infty}(1+e^{\infty})} = \frac{1}{1+e^{\infty}} = 0$$

$E(y)$ maximum value of $E(y)$ is 1

minimum value of $E(y)$ is 0

Because of the interpolation
linear
Error

↳ normal distribuction

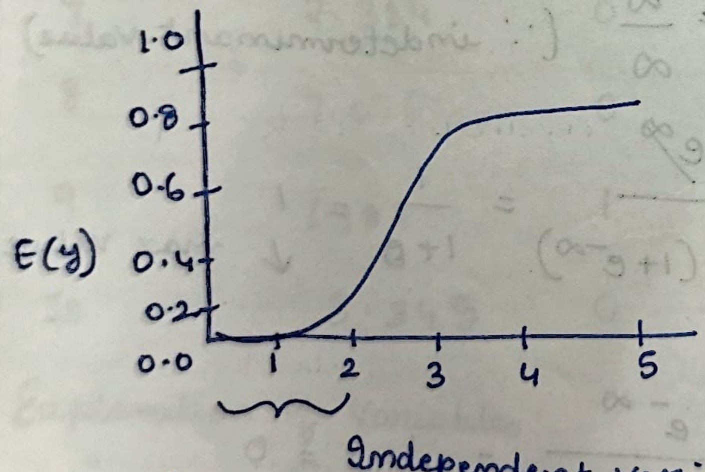
y caupon $\left[\begin{matrix}1\\0\end{matrix}\right\}$ dependent Variable. Binomial distribution

If dependent Variable is of binomial distribution we must apply logistic regression and the error term is follows Binomial ; linear Cant be used

Because of the interpretation of $E(y)$ as a probability, the logistic regression is

$$E(y) = P(y=1|x_1, x_2, x_3, \dots, x_p)$$

Logistic regression equation for β_0 and β_1



3
x is ↑ , E(y) is almost 1

x is below 1 $\xrightarrow{E(y)}$ 0

Estimating the Logistic Regression Equation.

In Simple linear and multiple regression the least Squares method is used to Compute $b_0, b_1, b_2, \ldots, b_p$ as estimates of the model parameters $(0, 1, \cdots, p)$.

The nonlinear form of the logistic regression equation makes the method of Computing estimates more Complex.

we use maximum Likelihood Estimation in case of Logistic Regression.

$$\hat{y} = \text{estimate of } P(y=1 \mid x_1, x_2, \cdots x_p) = \frac{e^{b_0 + b_1 x_1 + \cdots + b_p x_p}}{1 + e^{b_0 + b_1 x_1 + \cdots + b_p x_p}}$$

Here y ~~Has~~ is dependent Variable (Caupon)

y has provides an estimate of the probability that $y = 1$, given a particular Set of Values for the independent Variables.

$$\left.\begin{array}{l} \beta_0, \beta_1, \ldots \beta_p \rightarrow \text{Population} \\ b_0, b_1, \ldots, b_p \rightarrow \text{Sample} \end{array}\right\} \text{Parameters}$$

estimate population parameters with help of Sample parameters.

Logistic Regression Analysis:

Eg: Vote for a DemoCrat or Republican.

Social Sciences

Marketing

Health Care

Managerial use

dependent variable

$P(Y=1 \mid x_1=2, x_2=0) = 0.1880$

coupon — independent

prob. that customer using coupen

$x_1 = 2 \to 2000\$$

$x_2 = 0 \to$ not using Simmon's Credit Card

x_1 = annual Spending

x_2 = Simmon's Credit Card

$$\hat{y} = \frac{e^{-2.14637 + 0.341643(2) + 1.09873(0)}}{1 + e^{-2.14637 + 0.34164(2) + 1.09873(0)}}$$

$$= \frac{e^{-1.4631}}{1+e^{-1.4631}}$$

$$= \frac{0.2315}{1.2315}$$

$$= 0.1880$$

$P(y=1 \mid x_1 = 2 \mid x_2 = 1) = 0.4099$

$$= \frac{e^{-0.3644}}{1+e^{-0.3644}} = \frac{0.6946}{1.6946} = 0.4099$$

person have more Credit card is having more probability (using Coupon).

It appears that the probability of using the coupon is much higher for customers with a Simmons Credit card.

		Annual Spending						
		\$ 1000	\$ 2000	\$ 3000	\$ 4000	\$ 5000	\$ 6000	\$ 7000
credit card	yes	0.3305	0.4099	0.4943	0.5791	0.6594	0.7315	0.7931
	no	0.1413	0.1880	0.2457	0.3144	0.3922	0.4759	0.5610

Testing for Significance:

$H_0 : \beta_1 = \beta_2 = 0$

H_a: one or both of the parameters is not equal to 0.

G Statistics: The test for overall Significance.

Degrees of freedom df : No. of Independent Variables.

If the null hypothesis is true, the Sampling distribution of G follows a chiSquare dist. with df to equal no. of independent Variables in the model.

Interpreting odds ratios in logistic Regression:

Odds Ratio:

probability to odds to log of odds

p^* of Success of Some event

$p = 0.8$

prob. of failure $= 1 - 0.8 = 0.2$

The odds of Success $= \frac{p}{1-p} = \frac{0.8}{0.2} = \frac{4}{1}$

$p = 0.5 \rightarrow 50-50$

$q = 0.5$

$\frac{p}{1-p} = \frac{0.5}{0.5} = \frac{1}{1}$

Transformation is montonic

↓

from prob. to odds

→ Prob   odds

↑   ↑

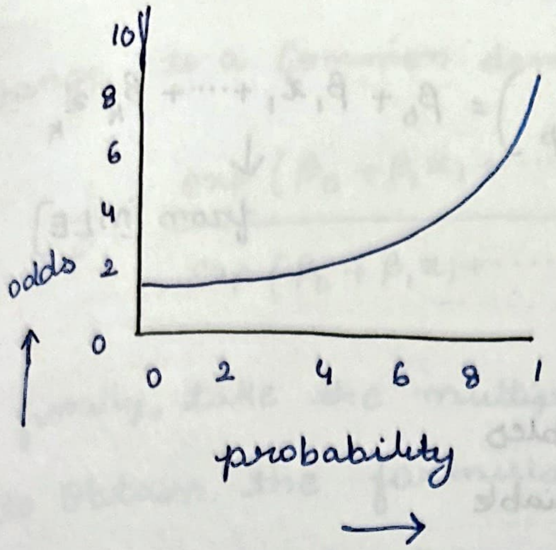
Eg. .001   .001001

{prob ranges from 0 to 1

$p = 0$ odds $= \frac{p}{1-p} = \frac{0}{1-0} = 0$

odds ranges 0 to ∞}

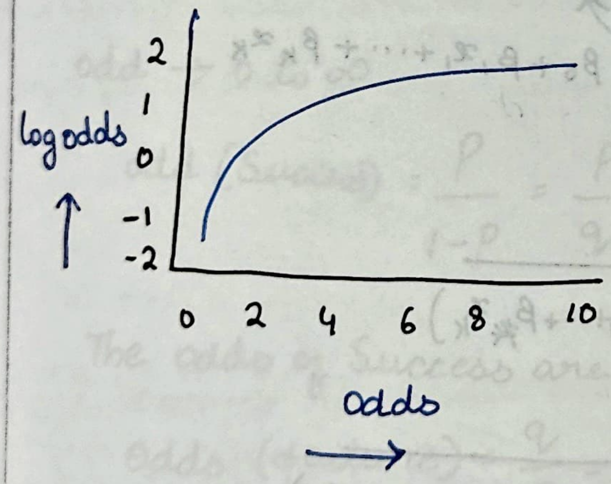
$p = 1$ odds $= \frac{1}{1-1} = \frac{1}{0} = \infty$



Log Transformation is monotonic

from Odd to log odd

P	odds ⟷	log odd
.001	.001001	-6.906755



Logistic regression model allows us to establish the relationship b/w a binary outcome Variable (dependent) and Group of predicted Variables (Independent)

Logistic)

$$\text{logit}(P) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

from [MLE]

$p = p(y=1)$

$x_1, x_2, \ldots x_k$

predicter Variables

Independent Variable

Exponentiate and take the multiplicative inverse of both Sides

$$\frac{1-P}{P} = \frac{1}{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K)}$$

$$\left[\because \log_e\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K\right.$$

$$e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K}$$

$$\frac{1}{P} - 1 = \frac{1}{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}$$

$$\frac{1}{P} = 1 + \frac{1}{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}$$

we above eq. we have

partial out the fraction on LHS of eq and add 1 to both Sides.

change 1 to a Common denominator

$$\frac{1}{p} = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) + 1}{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K)}$$

finally, take the multiplicative inverse again to Obtain the formula for the prob. $P(Y=1)$

$$P = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K)}.$$

$p = 0$ to 1.

Interpreting the logistic Regression Equation.

Ratio 0 to 1 Prob(Success)

$P = 0.8 \quad q = 1 - P = 1 - 0.8 = 0.2$

odd → 0 to ∞

$$\text{odd (Success)} = \frac{P}{1-P} = \frac{P}{q} = \frac{0.8}{0.2} = 4 = \frac{4}{1}$$

The odds of Success are 4 to 1

$$\text{Odds (failure)} = \frac{q}{P} = \frac{0.2}{0.8} = 0.25 = \frac{1}{4}$$

i.e $\frac{1}{4} = 0.25$ and $\frac{1}{0.25} = 4$

dependent Variable (Coupon)

$$\text{Odds} = \frac{P(y=1 \mid x_1, x_2, \cdots, x_p)}{P(y=0 \mid x_1, x_2, \cdots, x_p)}$$

$$= \frac{P(y=1 \mid x_1, x_2, \cdots, x_p)}{1-P(y=1 \mid x_1, x_2, \cdots, x_p)} = \frac{P}{1-P} = \frac{\text{Success}}{\text{not Success}}$$

Odd ratio:

$$\text{Odds Ratio} = \frac{\text{odds}_1}{\text{odds}_0}$$

1^{st} Level — having the credit card.

0^{th} Level — not having Credit card

measures the impact on the odds of a one-unit increase in only one of the independent Variables

Interpretation:

* for eg: Suppose we want to compare the odds of using the coupon for customers who spend \$2000 annually and have a Simmons Credit card ($x_1=2$ and $x_2=1$) to the odds of using the coupon for customers who spend \$2000 annually and do not have a Simmons Credit card ($x_1=2$ and $x_2=0$)

Compare odds $\left[\begin{array}{ll} x_1 = 2 & x_2 = 1 \\ x_1 = 2 & x_2 = 0 \end{array}\right.$ of using the coupon

we are interested in interpreting the effect of a one-unit increase in the independent Variable x_2

$$odds_1 = \frac{P(y=1 \mid x_1=2, x_2=1)}{1-P(y=1 \mid x_1=2, x_2=1)}$$

$$\text{estimate of } odds_1 = \frac{0.4099}{1-0.4099} = 0.6946$$

$$odds_0 = \frac{P(y=1 \mid x_1=2, x_2=0)}{1-P(y=1 \mid x_1=2, x_2=0)}$$

$$\text{estimate of } odds_0 = \frac{0.1880}{1-0.1880} = 0.2315$$

$$\text{Estimated odd ratio} = \frac{0.6946}{0.2315} = 3.00$$

The estimated odds in favour of using the coupon for customers who spent \$2000 last year and have a Simmons Credit Card are 3 times greater than the estimated odds in favour of using the coupon for customers who spent \$2000 last year and do not have a Simmons Credit card.

The odds ratio for each independent Variable is computed while holding all the other independent Variables constant.

x_1 = Spending

x_2 = Simmon's Credit Card.

But it does not matter what Constant values are used for other independent Variables.

For instance, if we (considered) computed the odd's ratio for Simmons credit card Variable (x2) using $3000, instead of $2000, as the value for annual Spending variable (x1), we would still obtain the same value for the estimated odds ratio (3.00)

$$\text{odd}\,s_1 = \frac{0.4943}{1-0.4943} = \frac{0.4943}{0.5057} = 0.97745699$$

$$\text{odd}\,s_0 = \frac{0.2457}{1-0.2457} = \frac{0.2457}{0.7543} = 0.325732$$

$$\text{Estimated odd ratio} = \frac{0.97745699}{0.325732} = \underline{\underline{3}}$$

Relationship b/w odd ratios and Coefficient of independent Variables

		Coef.	Std Err	Z	P>|z|
	Const	-2.1464	0.5772	-3.7183	0.0002
x_2	card	1.0987	0.4447	2.4707	0.0135
x_1	Spending	0.3416	0.1287	2.6551	0.0079

Estimated odds ratio $= e^{b_1} = e^{0.341643} = 1.41$

$= e^{b_2} = e^{1.09873} = 3.00$

Odd ratio $= e^{\beta_i}$

Effect of a change of more than one unit in odd ratio.

The odd ratios for an independent Variable represents the change in the odds for a one unit change in the independent Variable holding all the other independent Variables constant.

Suppose that we want to consider the effect of a change of more than one unit, say c units.

For Instance, suppose in the Simmons ex. that we want to compare the odds of using the coupon for customers who spend \$ 5000 annually ($x_1 = 5$) to the odds of using the coupon for customers who spend \$2000 annually ($x_1 = 2$)

In this case $c = 5-2 = 3$ and corresponding estimated odds ratio is very useful.

$$e^{cb_1} = e^{3(0.341643)} = e^{1.0249} = 2.79$$

This result indicates that the estimated odds of using the coupon for customers who spend $5000 annually is 2.79 times greater than the estimated odds of using the coupon for customers who spend $2000 annually.

In other words, the estimated odds ratio for an increase of $3000 in annual spending is 2.79

G vs Z

Because of the unique relationship b/w the estimated coefficients in the model and the corresponding odd ratios, the overall test for significance based upon the G statistic is also a test of overall significance for the odds ratio.

In addition, the Z test for the individual significance of a model parameter also provides a statistical test of significance for the corresponding odds ratio.

$$\left\{\begin{array}{l} G \\ Z \end{array}\right. \qquad \left\{\begin{array}{l} F \\ T \end{array}\right.$$

Logistic Regression. Linear Regression.

Linear Regression Model VS Logistic Regression Model:

Linear Regression model:

Dependent Variable

$Y_1 = x_1 + x_2 + \cdots + x_n$ independent Variable

where,

Y_1 is Continous data

Independent Variables = non metric and metric

non metric → discrete

metric → continous

Logistic Regression model

Independent

$Y_1 = x_1 + x_2 + \cdots + x_n$

dependent

where,

Y_1 = Binary nonmetric

Discrete

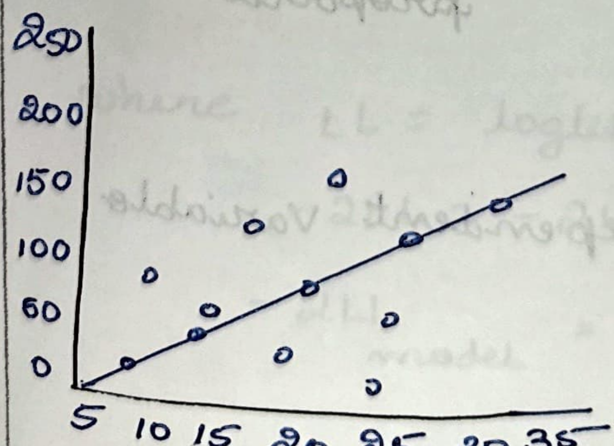
Independent Variables = non metric and metric

non metric → Discrete

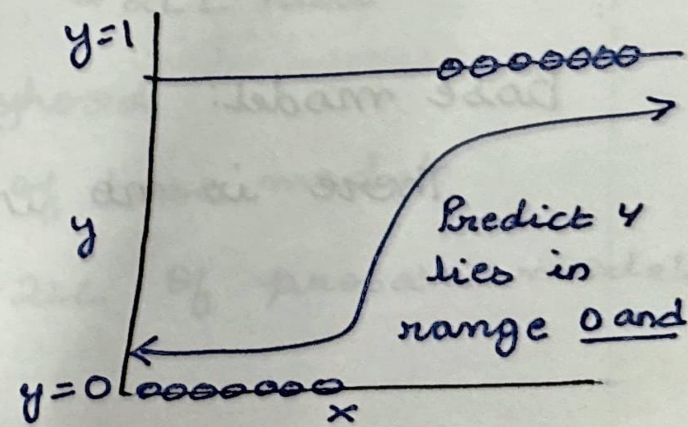
metric → Continous

Graphical Representation:

Linear Regression



Logistic Regression



correspondence of primary Elements of model fit

Linear Regression.	logistic
Tatal Sum of Squares SST	-2LL of base model ↓ loglikelihood
Error Sum of Squares SSE	-2LL of proposed model
F test of model fit $F = \frac{MSR}{MSE} = \frac{SSR/K}{SSE/(n-K-1)}$	G Test chi - Square test -2LL difference
Coefficient of determination (R^2)	Pseudo R^2 measures
Regression Sum of Squares.	Difference of -2LL for base and proposed models

Base model:

There is no independent Variable

Determination of Coefficients

Linear Regression

R^2

$r^2 = SSR / SST$

where

SSR = Sum of Squares due to Regression

SST = total Sum of Squares

Log Likelihood:

* Measures used in Logistic Regression.
* Lack of predictive fact
* SSE in linear Regression
* If you have less loglikelihood then it is a good model
* To Compare models

Logistic Regression.

$$R^2_{Logit} = \frac{-2LL_{null} - (-2LL_{model})}{-2LL_{null}}$$

where LL = loglikelyhood

$-2LL$ = $-2LL$ of base model

$-2LL_{model}$ = $-2LL$ of proposed model

Testing for Overall Significance:

Linear: | Logistic

F-test of model fit | G-test

$F = MSR / MSE$ | $G = -2\ln\left(\text{likelihood without a Variable} / \text{with Variable}\right)$

Significance of Each independent Variable

t-test | Wald test $\dfrac{\hat{\beta}_1}{\text{Std Error}(\hat{\beta}_1)}$

Linear Regression | Logistic

Error Sum of Squares | -2LL of proposed model

$SSR = \Sigma (y_i - \bar{y}_i)^2$ | Diff. b/w log likelihood

SST - SSE | $2LL_{null} - (2LL_{model})$

normally distributed | Binomially

linear Reg. assumes that residuals are approxily equal for all predicted Variable Values.	logistic Reg. does not need residuals to be equal for each level of the predicted dep. Variable Values
Based on Least Square estimation	maximum likelihood estimation
Reg. coefficients should be chosen in such a way that it minimizes the Sum of the Squared distances of each Observed response to its fitted Value.	maximizes the probability of y given x (likelihood) with MLE, the Computer uses diff. "iterations" in which it tries different Solutions until it gets the max. likelihood estimates

parametric family of distributions:

*** *** *

① Binomial distribution:

*** **Martin clothing Store problem.**

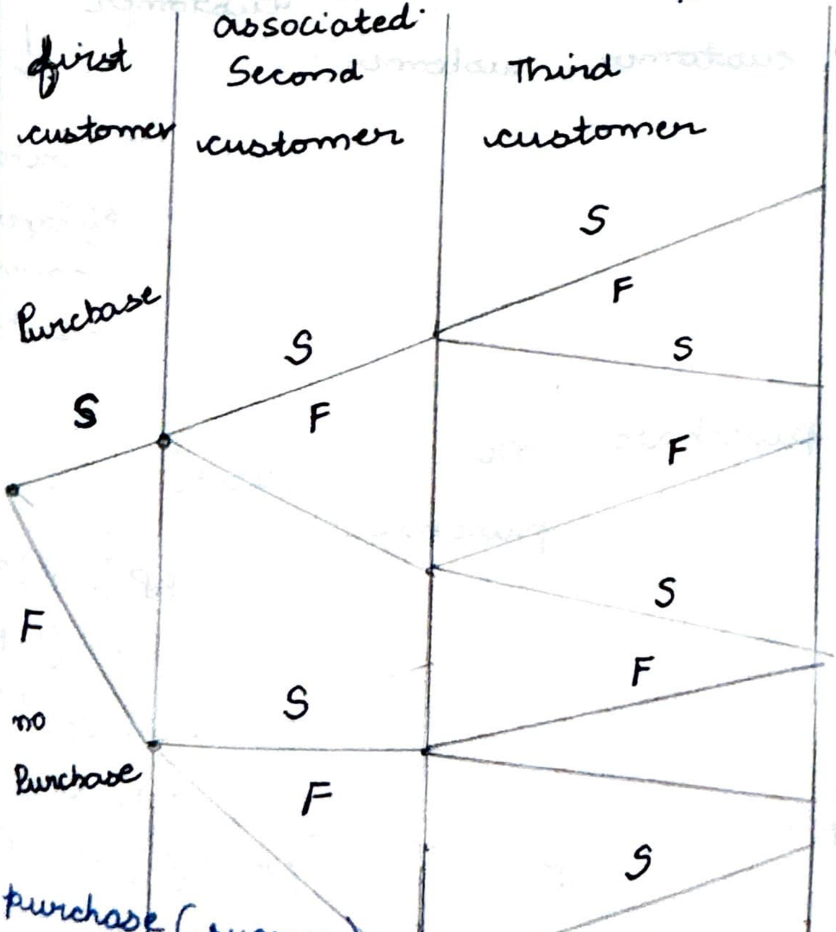
→ Let us Consider the purchase decisions of the next 3 customers who enter a Store.

→ on the basis of past experience, the store manager estimates the probability that any

→ one customer will purchase is 0·30

$$\star \begin{cases} p = 0{\cdot}30 & \text{purchasing } p^* \\ 1-p = 0{\cdot}70 & \text{not purchasing } p^* \end{cases}$$

→ what is the probability that two of the next 3 customers will make a purchase



Experimental Outcome	Values of x.
(S, S, S)	3
(S, S, F)	2
(S, F, S)	2
(S, F, F)	1
(F, S, S)	2
(F, S, F)	1
(F, F, S)	1
(F, F, F)	0

$^3C_2$

S = purchase (success)
F = no purchase
x = no. of customers making a purchase.

$^3C_2$ is no. of Combinations to purchase 2 customers out of 3 customers

2 2 2

$^3C_2 = 3$

$^nC_r = {}^nC_{n-r}$

$^3C_2 = {}^3C_1 = 3$

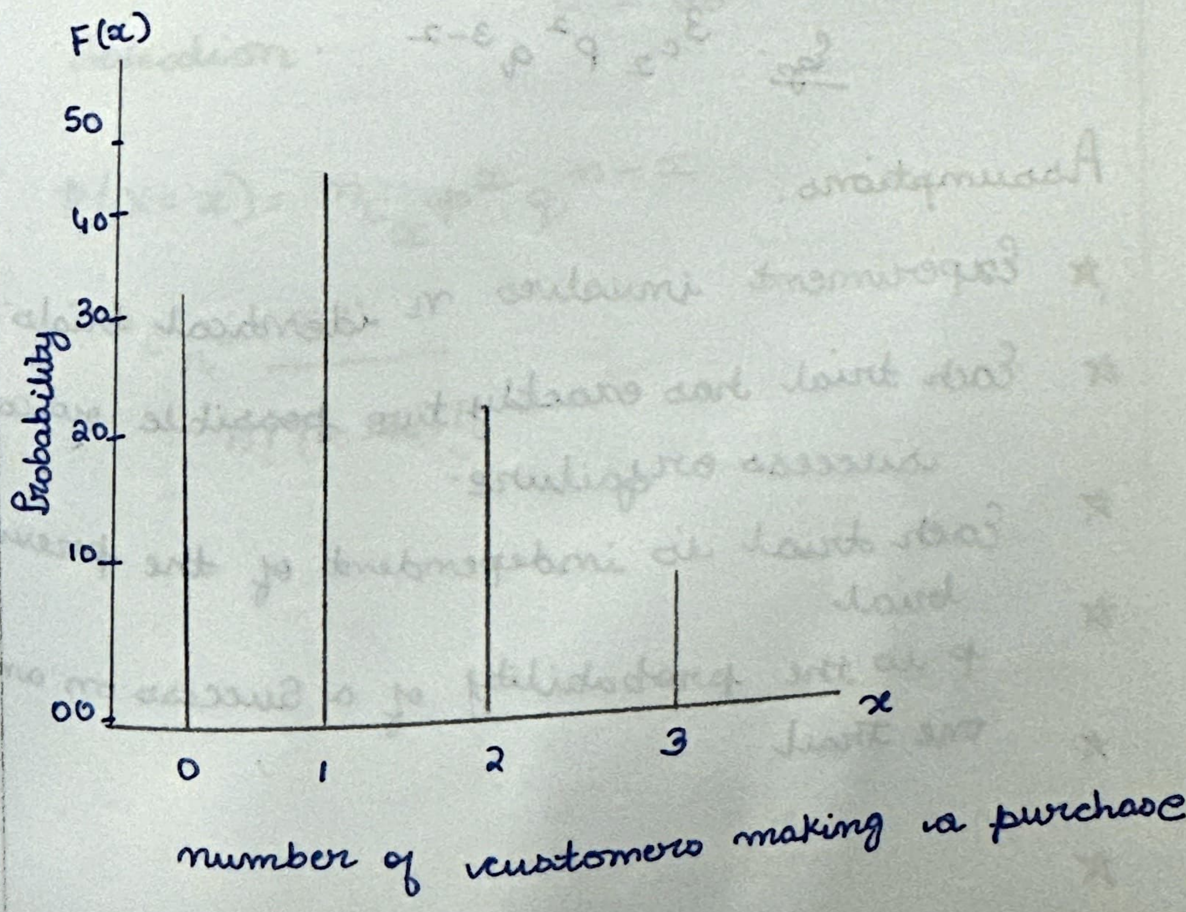
Trial outcomes:-

1st customer	2nd customer	3rd customer	Experimental outcome	P ↓ Probability of Experimental outcome
purchase	purchase	no purchase	(S,S,F)	$PP(1-P) = P^2(1-P) = (0.30)^2(0.70) = 0.063$
P	NP	P	(S,F,S)	$= P(1-P)P = 0.063$
NP	P	P	(F,S,S)	$= (1-P)(PP) = 0.063$

Success P
failure (1-P)

* Graphical representation of probability dist. for number of customers making a purchase

x	$P(x)$
0	$0.7 \times 0.7 \times 0.7 = 0.343$
1	SFF $0.3 \times 0.7 \times 0.7 +$ FSF $0.7 \times 0.3 \times 0.7 +$ $0.7 \times 0.7 \times 0.3 = 0.441$ FFS
2	0.189
3	0.027



from trial outcomes:

$$^3c_2 \, p^2 (1-q)$$

$$p^2(1-p) + p^2(1-p) + p^2(1-p)$$

$$= 3p^2 (1-p)$$

$$= {}^3c_2 \, p^2 (1-p)^{3-2}$$

$$^3c_2 = 3$$

$$^nc_x \, p^x q^{n-x}$$

Binomial distribution

$$P(X=x) = {}^nc_x \, p^x q^{n-x}$$

Eg: $^3c_2 \, p^2 q^{3-2}$

Assumptions:

☆ Experiment involves n identical trials

☆ Each trial has exactly two possible outcomes success or failure.

☆ Each trial is independent of the previous trial

☆ p is the probability of a success on any one trail

☆

☆

* $q = (1-p)$ is the probability of a failure on any one-trial.

* p and q are Constant throughout the experiment.

probability function. $P(x) = \frac{n!}{x!(n-x)!} p^x \cdot q^{n-x}$ for $0 \leq x \leq n$

mean Value $\mu = n \cdot p$

Variance and Standard deviation.

$\sigma^2 = n \cdot p \cdot q$

$\sigma = \sqrt{\sigma^2} = \sqrt{n \cdot p \cdot q}$

$p(X = x) = {}^nC_x \, p^x \, q^{n-x}$

$${}^nC_r = \frac{n!}{r!(n-r)!}$$

Mean and Variance.

☆ Suppose that for the next month the clothing store forecasts 1000 customers will enter the store.

☆ what is the expected number of customers who will make a purchase?

☆ The answer is $\mu = np$ $\quad n = 1000$

$$= (1000)(0.3) = 300$$

✓ for the next 1000 customer entering the store, the Variance and Standard deviation for the number of customers who will make a purchase are

$$\sigma^2 = np(1-p)$$

$$= 1000\,(0.3)(0.7)$$

$$= 210$$

$$\sigma = \sqrt{210}$$

$$= 14.49$$

★ $\mu = np$

$\sigma^2 = npq$

```
import Scipy
import numpy as np
from Scipy.stats import binom
```

Poisson Distribution: Assumptions:

* Statistical Distribution → how many times event is occured within specified period of time.
* Discrete event ⟨ occuring / not occuring
* Describes rare events in large population
  mutation of cells in body: mutuation acquisition.
* Each Occurence is independent any other occurences: [discrete occurences over interval)

* The expected number of occurences must hold constant throughout the experiment.

Example:

① Measurements of number of occurences within a time period. arrival patterns

Eg1: Arrivals at queuing systems

* Airports - people, airplanes and baggage
* Banks - people, automatic and loan applications
* customers at a store.
* Computer file servers - read and write operations.

number of calls to a switch board.

The number of arrivals in any service facility like at an ATM, railwaystation, petrol pumps.

Eg 2: Defects in manufactured goods.

* number of defects per 1,000 feet of extruded copper wire

* number of blemishes per Square foot of painted Surface.

* number of errors per typed page.

Eg3: Number of Extreme weather events

Eg4: The number of aircraft/road accidents in any time interval.

* poisson distributions model the number of Occurences within a fixed time period.

* can be used to approximate binomial distributions for very Small Success probabilities.

Binomial:

$$P(X=x) = {}^{n}C_{x}\, p^{x} q^{n-x}$$

$$\downarrow \; n \to \infty,\ p \to 0$$

poisson distribution

$$\frac{e^{-\lambda}\lambda^{x}}{x!} = P(X=x)$$

poisson distribution formula:

$$\frac{e^{-\lambda} \lambda^{x}}{x!}$$

e = Euler's constant ≈ 2.718

λ = mean or expected value of the variable

x = number of successes for the event

! = factorial.

mean	Variance	Standard deviation.
λ	λ	λ

$\lambda = 3.2$ customers/ 4 minutes

$X = 10$ customers/ 8 minutes

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

$$P(x=10) = \frac{6.4^{10} e^{-6.4}}{10!}$$

$= 0.0528$

$P(X = x)$

$$= \frac{\lambda x e^{-\lambda}}{x!}$$

=

$\lambda = 3.2$ customers/ 4 minutes

$X = 6$ customers/ 8 minutes

Adjusted λ

6.4 customers/ 8 minutes

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

$P(x = 6)$

$$= \frac{6.4^{6} e^{64}}{64} = 0.1856$$