INTRODUCTION to MATHEMATICS for MACHINE LEARNING

BY- JASMEET GUJRAL

MACHINE LEARNING

PROBABILITY CALCULUS LINEAR ALGEBRA STATISTICS

OPTIMIZATION

MACHINE LEARNING ALGORITHMS

* NOTATIONS USED IN MATHEMATICS
* JARGONS USED IN MACHINE LEARNING

MOTIVATION:

DATA AVAILABLE TODAY
&
GETTING GENERATED
EVERY MINUTE

USING/ANALYZING DATA

⇓

MACHINE LEARNING

MATHEMATICS

TECHNOLOGY

LANGUAGE THAT HELPS IN
UNDERSTANDING THE FUNCTIONALITY

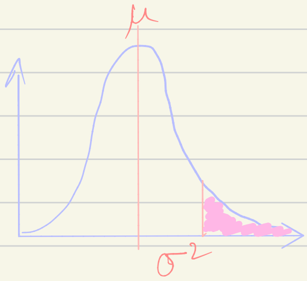
TOOL HELPS IN IMPLEMENTING

BUSINESS

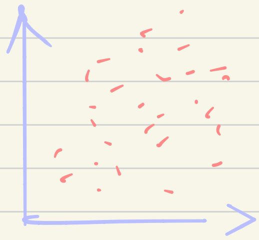
⇓

UNDERSTANDING OF PROBLEM STATEMENT

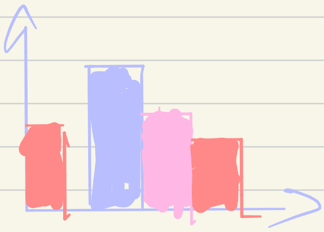
OUTLINE of SESSION 1 : ELEMENTARY



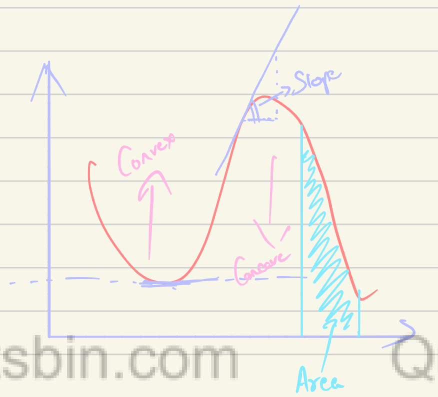
i - DESCRIPTIVE STATISTICS :



- CENTRAL TENDENCY
- DISPERSION
- VISUALIZATION

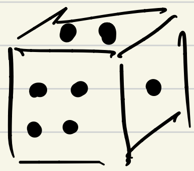


ii - CALCULUS :



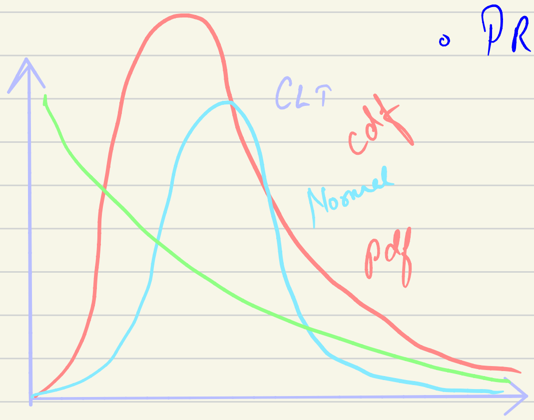
- FUNCTION / VARIABLE & CONSTANTS
- RATE / SLOPE
- SUMMATION / AREA
- MINIMA / MAXIMA
- CONVEXITY / COCAVITY

iii - DETERMINISTIC / RANDOMNESS



iv - PROBABILITY

- BASIC LAWS
- CONDITIONAL PROBABILITY
- PROBABILITY DISTRIBUTION



DESCRIPTIVE STATISTICS

SESSION 1

STATISTICAL ANALYSIS PROCESS

DESCRIPTIVE → INFERENCE → PREDICTION

↓

AIMS AT SUMMARIZING DATA

DESCRIPTIVE STATISTICS

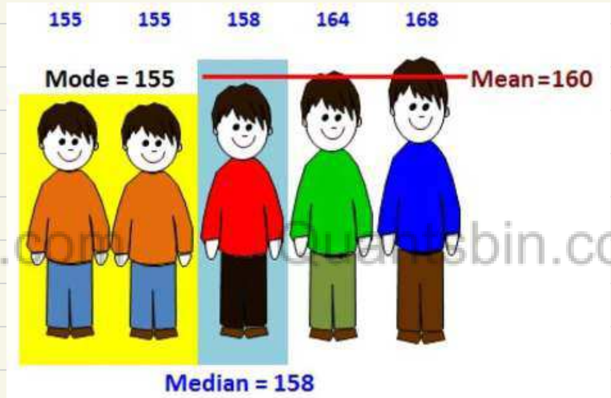
↓

- MEASURE OF CENTRAL TENDENCY
- MEASURE OF DISPERSION
- VISUALIZATION

• MEASURE OF CENTRAL TENDENCY

SUMMARIZE DATASET BY SINGLE NUMBER BY DEFINING CENTER OF DATA

→ MEAN
→ MEDIAN
→ MODE



– MEAN:

$$\frac{x_1 + x_2 + x_3 + \ldots\ldots + x_n}{n}$$

(AVERAGE / EXPECTATION)

↑

CONTEXT OF RANDOM VARIABLE

$|||$

ALSO, $$\frac{\sum_{i=1}^{n} x_i}{n}$$

MATHEMATICAL SYMBOLS

$\sum_{i=a}^{b}$ (SIGMA/SUMMATION)

SUM OF ALL THE ELEMENTS CORRESPONDING TO VALUE OF 'i' STARTING FROM 'a' till 'b'

$\equiv$ (EQUIVALENCE)

GROUPED DATA

RANGE	VALUE (x)	FREQUENCY (f)
$r_0 < x \le r_1$	x_1	f_1
$r_1 < x \le r_2$	x_2	f_2
$r_2 < x \le r_3$	x_3	f_3
$r_3 < x \le r_4$	x_4	f_4
$r_4 < x \le r_5$	x_5	f_5

$$\frac{x_1 \times f_1 + x_2 \times f_2 + \ldots\ldots + x_5 \times f_5}{f_1 + f_2 + \ldots\ldots + f_5} \equiv \frac{\sum_{i=1}^{5} x_i \times f_i}{\sum_{i=1}^{5} f_i}$$

→ $x_i = \dfrac{r_i + r_{i-1}}{2}$

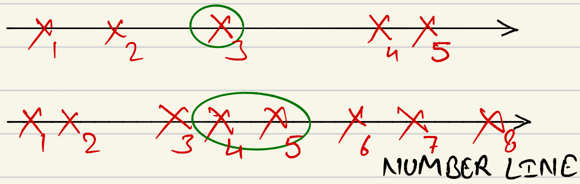
– MEDIAN

$x_1, x_2, x_3, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n$ ⬅ VALUES SORTED IN ASCENDING/DESCENDING ORDER

FOR

$n = odd \quad x_i \quad i = \frac{n+1}{2}$

$n = even \quad \frac{x_i + x_{i+1}}{2} \quad i = n/2$



GROUPED DATA

RANGE	VALUE (x)	FREQUENCY (f)	CUMMULATIVE FREQUENCY (C)
$r_0 < x \leq r_1$	x_1	f_1	$C_1 = f_1$
$r_1 < x \leq r_2$	x_2	f_2	$C_2 = f_1 + f_2$
$r_2 < x \leq r_3$	x_3	f_3	$C_3 = f_1 + f_2 + f_3$
$r_3 < x \leq r_4$	x_4	f_4	$C_4 = f_1 + f_2 + f_3 + f_4$
$r_4 < x \leq r_5$	x_5	f_5	$C_5 = f_1 + f_2 + f_3 + f_4 + f_5$

⬅ ENSURE X_i or RANGES ARE SORTED

LETS $C_2 < f_{MID} < C_3$

IF VALUES (x) ARE GIVEN

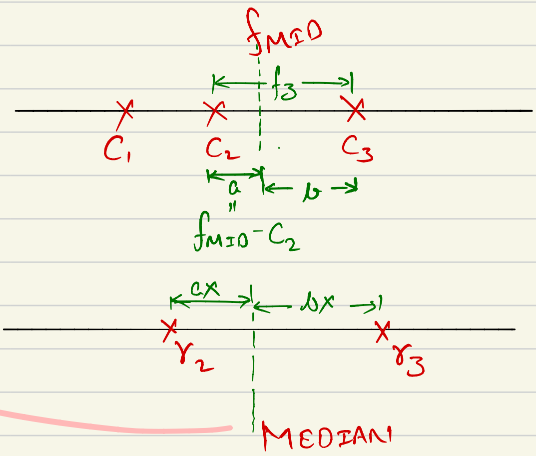
$$f_{MID} = \frac{\sum_{i=1}^{n} f_i}{2}$$

$= X_i$ CORRESPONDING TO $\sup\{C_i \in C \mid f_{MID}\}$ ⟶ X_3

IF RANGES ARE GIVEN

r_i CORRESPONDING TO $\inf\{C_i \in C \mid f_{MID}\}$

$$= r_i + (f_{MID} - C_i) \times \frac{(r_{i+1} - r_i)}{f_{i+1}}$$



– MODE

MAXIMUM OCCURING VALUE IN DATASET

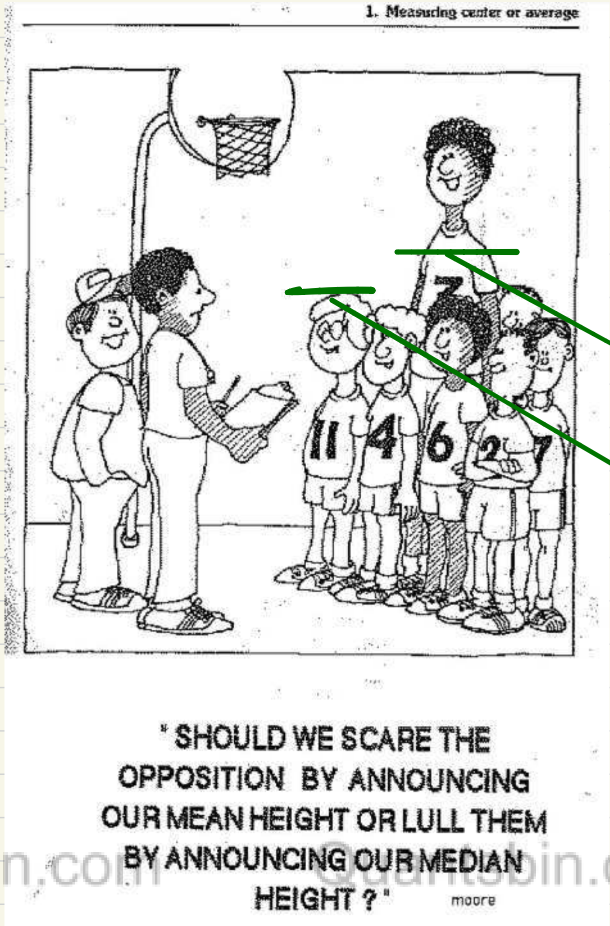
* THERE WILL ALWAYS BE SINGLE VALUE FOR MEAN AND MEDIAN BUT POSSIBLE TO HAVE MORE THAN ONE VALUE FOR MODE

MATHEMATICAL SYMBOLS

inf{T|S} MAX VALUE OF 'T' LESS THAN 'S'

sup{T|S} MIN VALUE OF 'T' GREATER THAN 'S'

IMPACT OF OUTLIERS:



MODE – NO IMPACT

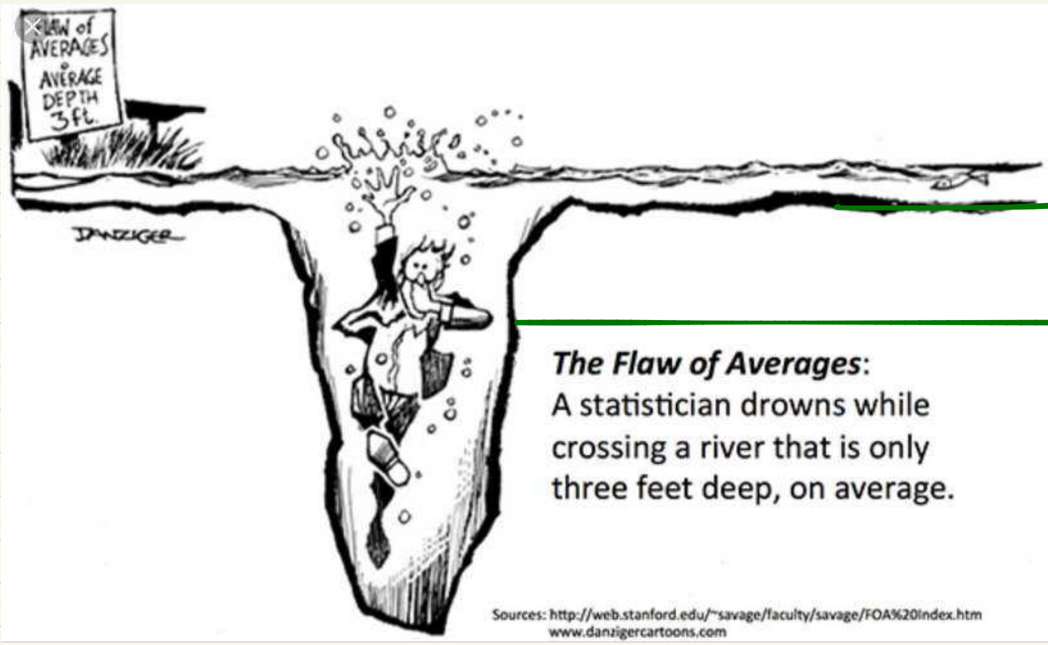
MEDIAN – MAX SHIFT OF ONE POSITION

MEAN – HIGHLY IMPACTED BY OUTLIERS

MEAN

MEDIAN

MEASURE OF CENTRAL TENDENCY ISN'T COMPLETE IN ITSELF



MEDIAN
MODE

MEAN

° Measure of Dispersion

Measures how scattered data is around central point

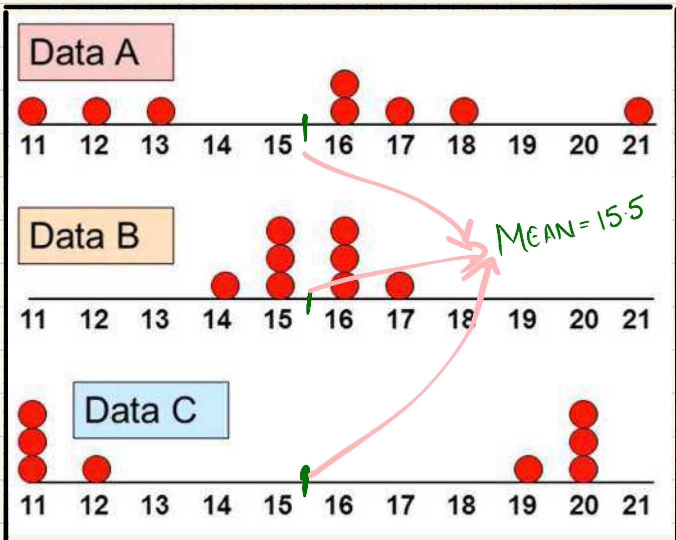
– Range

$X_{MAX} - X_{MIN}$

Data	Range
A	21 − 11 = 10
B	17 − 14 = 3
C	20 − 11 = 9

Least dispersed

Is 'C' less dispersed than 'A'??

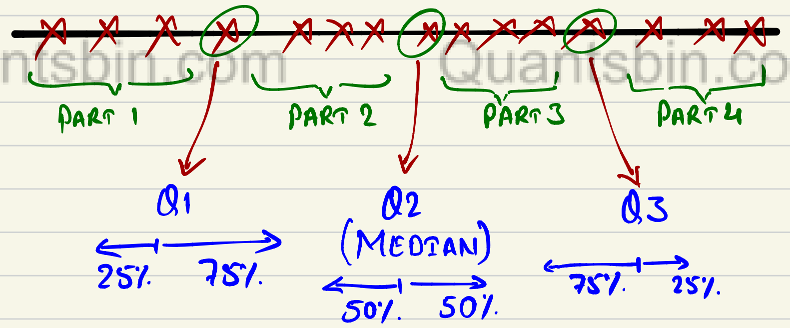


Data 'A', 'B' & 'C' have same mean 15.5

Do they look same??

– Quantiles & Box Plot

Similar to median but divides data into 4 parts

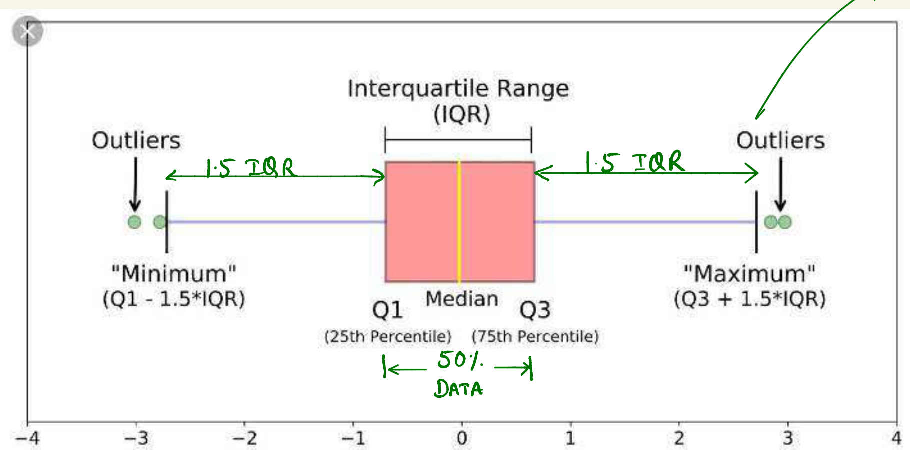


Quantiles are calculated same as median

$Q_1 \Rightarrow \cancel{f_{MID}}\ f_{Q_1} = \dfrac{\sum_{i=1}^{n} f_i}{4}$

$Q_2 \Rightarrow f_{MID} = \dfrac{\sum_{i=1}^{n} f_i}{2}$

$Q_3 \Rightarrow \cancel{f_{MID}}\ f_{Q_3} = \left(\sum_{i=1}^{n} f_i\right) \times \dfrac{3}{4}$



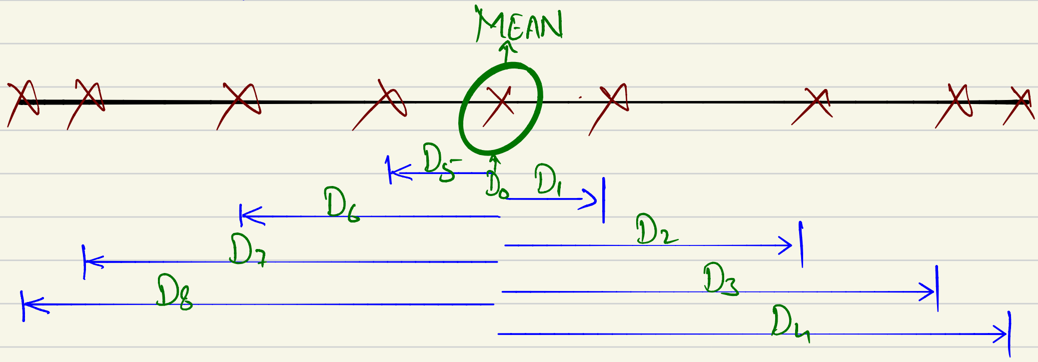
* One of the defination of outliers

Box Plot:

Graphical summarization of data

- MEAN ABSOLUTE DISTANCE (MAD)

$$\frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$ WHERE, $\bar{x}$ = MEAN



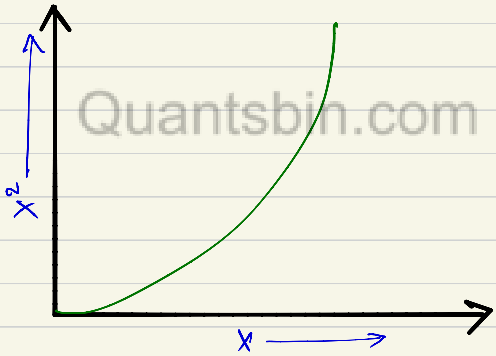
$$MAD = \frac{D_0 + D_1 + D_2 + D_3 + D_4 + D_5 + D_6 + D_7 + D_8}{9}$$

WHY ABSOLUTE? -ve & +ve VALUES SHOULD NOT CANCEL EACH OTHER

WHY NOT ABSOLUTE? MOD/ABSOLUTE IS NON-DIFFERENTIAL FN

- VARIANCE (σ^2)

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



* SQUARE FN IS ALWAYS +VE

$$VARIANCE = \frac{D_0^2 + D_1^2 + D_2^2 + D_3^2 + D_4^2 + D_5^2 + D_6^2 + D_7^2 + D_8^2}{9}$$

* VARIANCE GIVES MORE WEIGHTAGE TO FAR OFF VALUES
* UNIT OF VARIANCE SQ OF UNIT OF DATA

- STANDARD DEVIATION (σ)

$$\sigma = \sqrt{VARIANCE}$$

* STANDARD DEVIATION IS NON-ADDITIVE

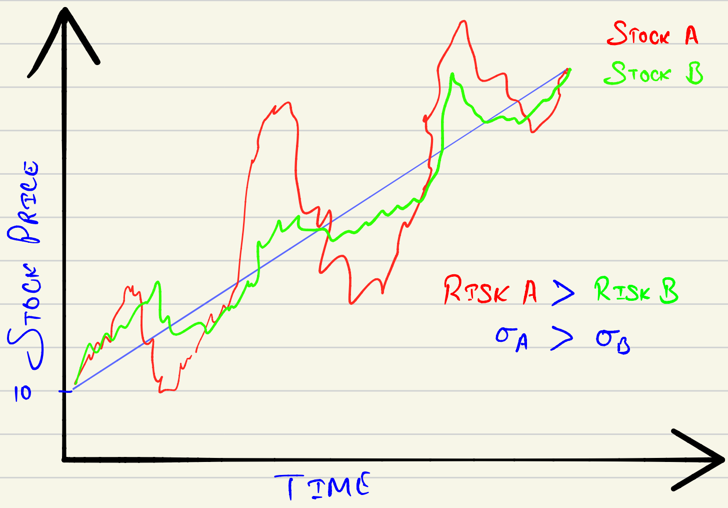
PROCESS	STD. DEV (σ)
A	σ_A
B	σ_B
A+B	$\cancel{\sigma_A + \sigma_B}$

$$\sigma_{A+B}^2 = \sigma_A^2 + \sigma_B^2$$
$$\sigma_{A+B} = \sqrt{\sigma_A^2 + \sigma_B^2}$$

Std. Dev (σ) as Risk

* Imp – For Stock Prices Standard Deviation is Calculated on Returns Not on Prices

$$\sigma_{\text{Stock Price}} = \sqrt{\frac{\sum_{i=1}^{n}\left[\ln\left(\frac{S_i}{S_{i-1}}\right)\right]^2}{n}}$$

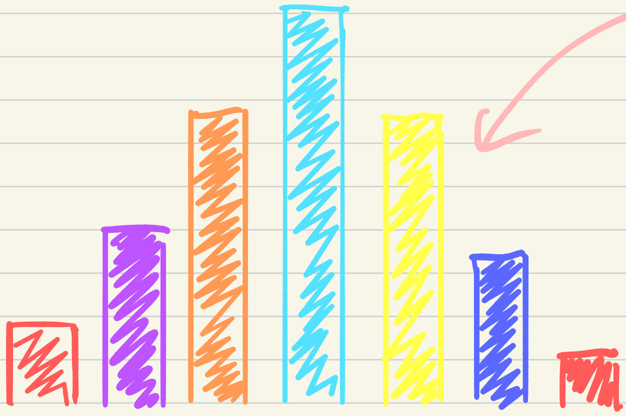


$\ln\left(\frac{S_t}{S_{t-1}}\right) \Leftarrow$ Continuously Compounded Daily Return

Risk as $\sigma_{\text{Stock Price}}$

$\hookrightarrow$ Risk is Uncertainity

• VISUALIZATION



HISTOGRAM

BEST CHART TO LOOK AT DISTRIBUTION OF DATA

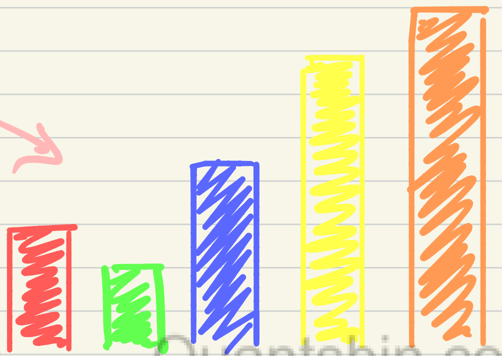
HELPS TO SUMMARIZE LARGE DATA

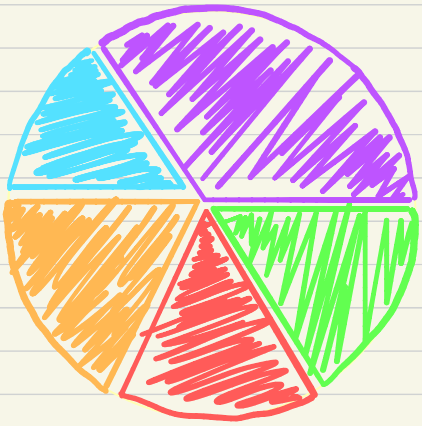
USED TO DESCRIBE SINGLE VARIABLE

BAR CHART

USED TO SHOW CHANGE IN VALUE OF A VARIABLE OVER TIME.

EXAMPLE - SALES NUMBERS OVER DIFFERENT YEARS





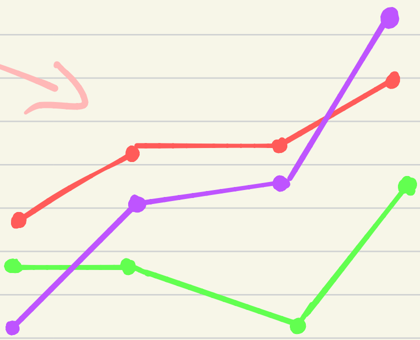
PIE CHART

COMMONLY USED TO PRESENT PERCENT OF EACH KIND OF VARIABLE IN TOTAL DATASET.

LINE CHARTS -

PRESENTS RELATIONSHIP BETWEEN TWO VARIABLE DESCRIBE BY X & Y AXIS

ALSO HELPS IN FINDING TRENDS



CALCULUS

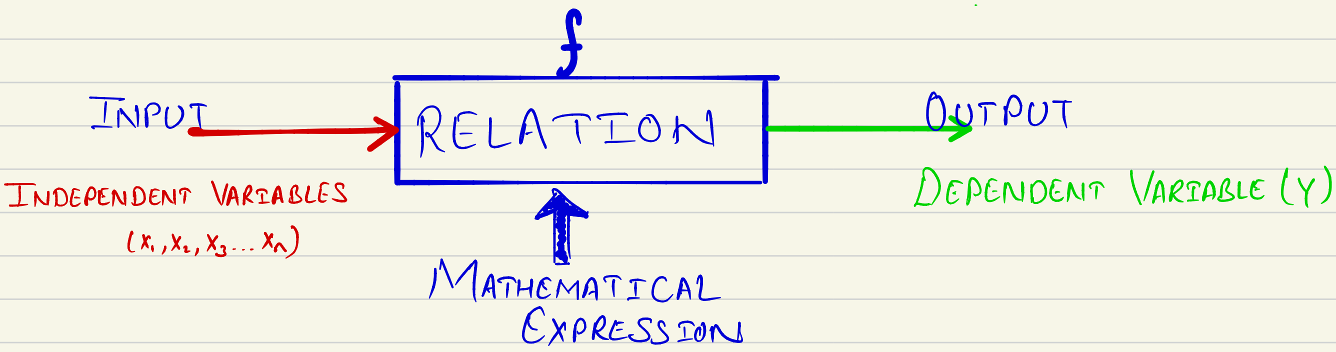
SESSION 1

CONSTANTS, VARIABLES & FUNCTIONS

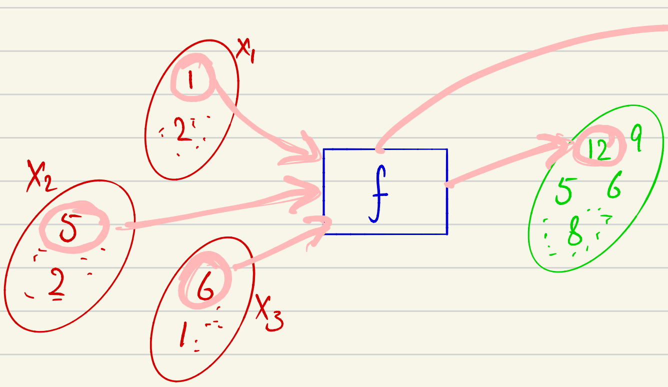
CONSTANTS – VALUES REMAINS SAME. Ex. π, e

VARIABLE – VALUES COULD CHANGE
COMMONLY DENOTED AS 'x'

FUNCTIONS –



$$Y = f(x_1, x_2, x_3, \dots x_n)$$



EXAMPLE f HERE

$x_1 + x_2 + x_3$

RELATION TYPES
- ONE TO ONE ✓
- MANY TO ONE ✓
- ONE TO MANY ✗

ONE TO ONE : $Y = MX + C$
M & C ARE CONSTANT
FOR EACH VALUE OF X THERE IS ONLY ONE CORRESPONDING VALUE OF Y AND VICE VERSA

MANY TO ONE : $Y = X^2$
$X = 2$ OR $X = -2 \implies Y = 4$

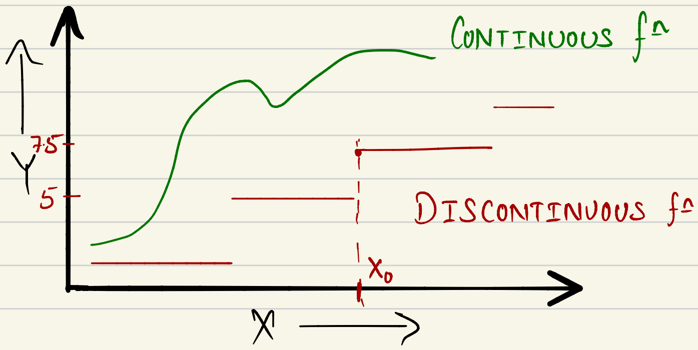
ONE TO MANY : $Y^2 + X^2 = 8$
$X = 2 \Rightarrow Y = 2$ OR $Y = -2$ ✗

LIMIT & CONTINUITY:

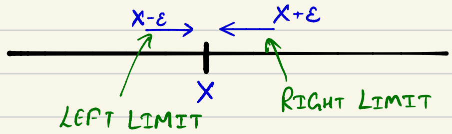
$$Y = f(x)$$

MATHEMATICALLY FOR CONTINUITY LIMIT SHOULD EXIST

I.E. $\lim_{\varepsilon \to 0} f(x+\varepsilon) = f(x-\varepsilon) = f(x)$

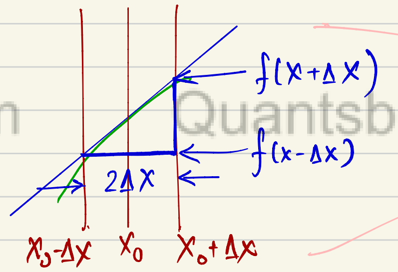


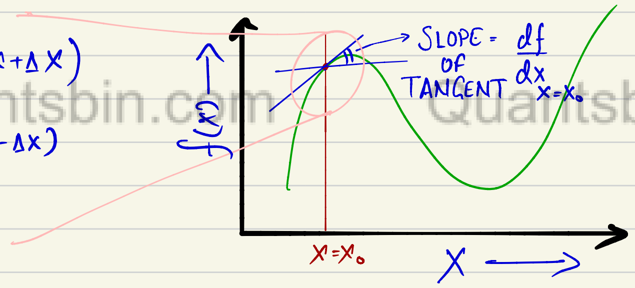
$f(x_0) = 7.5$
$f(x_0+\varepsilon) = 7.5$ (RIGHT LIMIT)
$f(x_0-\varepsilon) = 5$ (LEFT LIMIT)



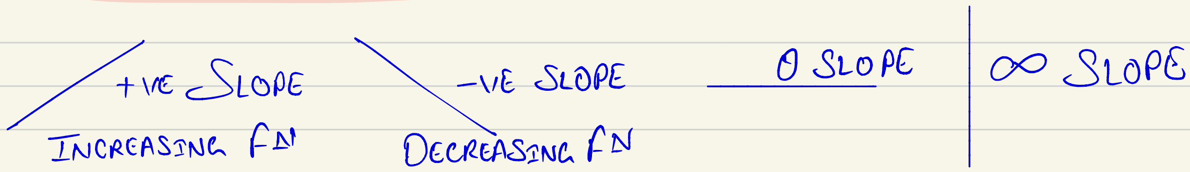
RATE AND SLOPE: DIFFERENTIATION $\frac{df}{dx}$

$Y = f(x)$ $\quad \left.\frac{df}{dx}\right|_{x=x_0}$ CHANGE IN $f(x)$ VALUE FOR UNIT CHANGE IN X WHEN $X = X_0$





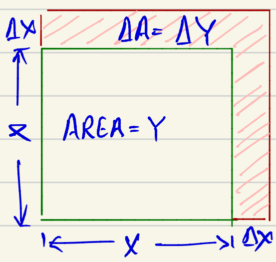
$$\left.\frac{df}{dx}\right|_{x=x_0} \simeq \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x-\Delta x)}{2\Delta x}$$



AREA OF SQUARE $\quad Y = x^2$

$\frac{dY}{dX}$ CHANGE IN AREA FOR SMALL CHANGE IN X

$$\frac{\Delta Y}{\Delta X} \simeq \frac{dY}{dX} \Rightarrow \Delta Y = \Delta X \frac{dY}{dX}$$



TRY FOR $X = 100$ & $\Delta X = 0.1$??

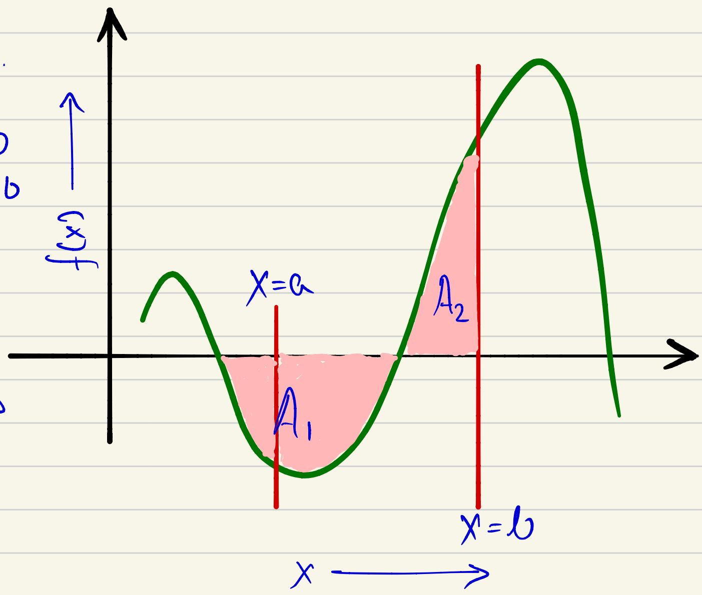
SUMMATION/AREA: INTEGRATION $\int_{x=a}^{x=b} f(x).dx$

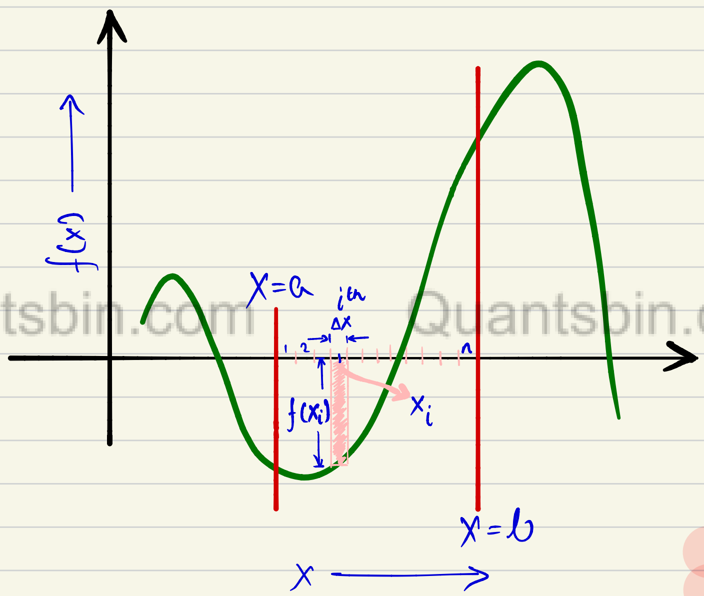
$\int$ IS AREA UNDER THE CURVE.

$\int_{x=a}^{x=b} f(x)dx$ = AREA BETWEEN X-AXIS AND $f(x)$ FROM POINT a TO b

$\Rightarrow \int_{x=a}^{x=b} f(x)dx = -A_1 + A_2$

AS AREA BELOW X-AXIS IS TAKEN AS NEGATIVE





$\int$ AS SUMMATION

$\Delta X = \frac{b-a}{n}$

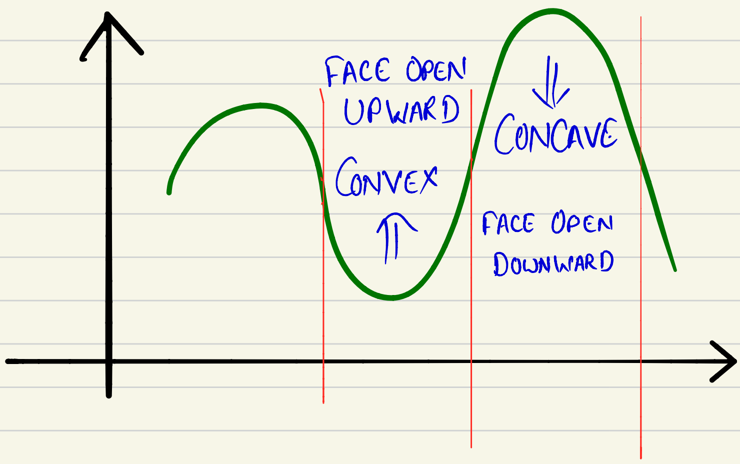
$A_i = f(x_i)\,\Delta X$

TOTAL AREA $= \sum_{i=1}^{n} f(x_i)\,\Delta X$

AS $\Delta X \rightarrow 0$ (VERY SMALL)

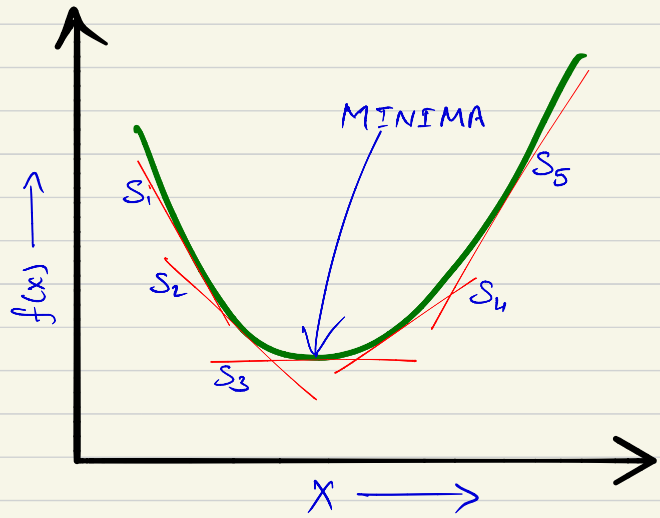
$$\int_{x=a}^{x=b} f(x)dx = \lim_{\Delta X \rightarrow 0} \sum_{i=1}^{n} f(x_i)\,\Delta X$$

CONVEX/CONCAVE

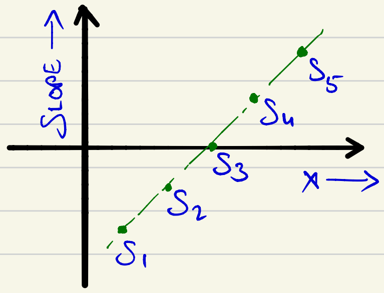


$$\underbrace{S_1 < S_2}_{-VE} < \underbrace{S_3}_{=0} < \underbrace{S_4 < S_5}_{+VE}$$

FOR CONVEX F^N SLOPE INCREASES AS WE MOVE ALONG $X \rightarrow$



SLOPE VS X PLOT



FOR CONVEX F^N SLOPE OF SLOPE IS +VE

I.E. $\dfrac{d}{dx}\left(\dfrac{df(x)}{dx}\right) \geq 0$

Or IF $\dfrac{d^2 f(x)}{dx^2} \geq 0 \Rightarrow$ CONVEX

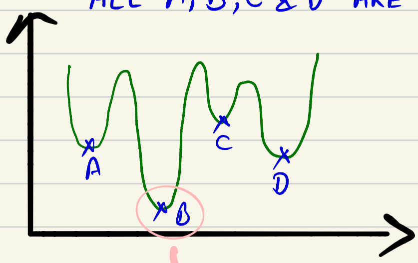
MINIMA –

SLOPE ZERO POINT OF CONVEX F^N

CONDITION FOR MINIMA

$$\frac{df(x)}{dx} = 0 \quad \& \quad \frac{df(x)}{dx} > 0$$

ALL A, B, C & D ARE MINIMA



GLOBAL MINIMA

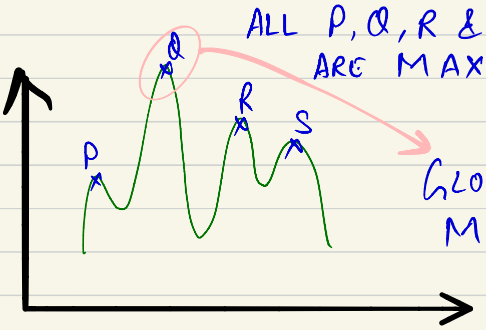
MAXIMA –

SLOPE ZERO POINT OF CONCAVE F^N

CONDITION FOR MAXIMA

$$\frac{df(x)}{dx} = 0 \quad \& \quad \frac{df(x)}{dx} < 0$$

ALL P, Q, R & S ARE MAXIMA



GLOBAL MAXIMA

• Deterministic & Random Events

Deterministic Process

For Deterministic Process Result is Known at the Start of the Process

Random Process

For Random Process we have no knowledge of exact outcome when we start the process.

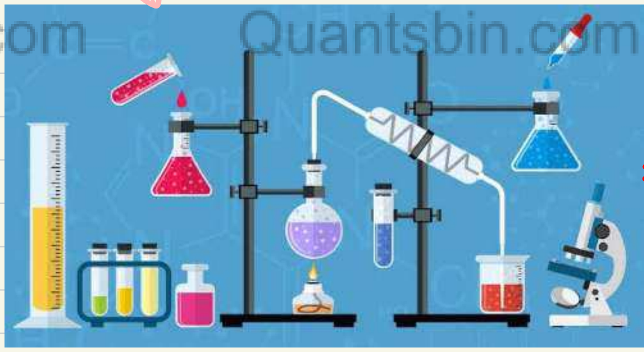
We have knowledge of chances of occurance of various outcomes.

Imp. Termonologies

Experiment : Process Under Observation

E.G. Tossing of Coin

Rolling of Dice



Sample Space : All Possible Outcomes

Event – Any Particular Set of Outcome of our Interest.

Possible Result of Experiment

Probability(p) - Quantifies Chance of Occurance of Each Outcome

$p = 1 \Rightarrow$ Deterministic Process

$p = 0 \Rightarrow$ Outcome Not Possible