

statistics

1 HOUR / 2 MARKS

SK AGRAWAL

topics

▶ **Introduction to Statistics**

- ▶ - Data classifications and processing, graphical representation of data, frequency distributions
- ▶ - Measures of central tendency, dispersion and skewness
- ▶ - Elementary theory of probability and probability distributions, sampling and sampling distributions
- ▶ - Simple test of significance, regression and correlation, multiple correlation coefficient
- ▶ - Time series
- ▶ - Index numbers

What is it

Statistics

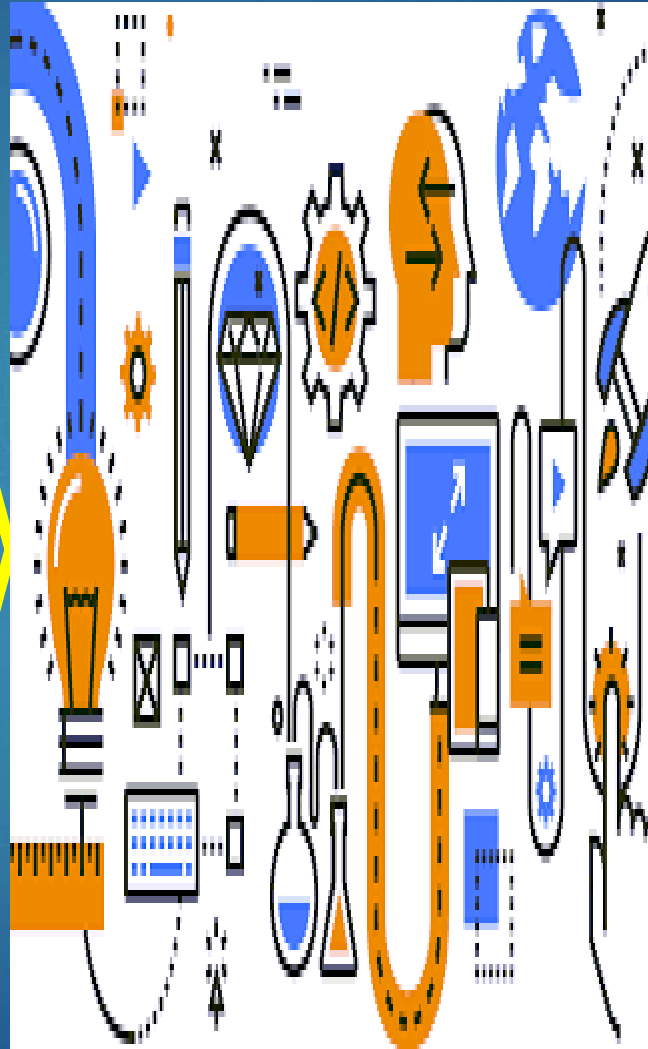
- ▶ is the sub section of Mathematics
- ▶ It deals with large quantity of similar numbers
- ▶ It also helps us to understand the larger population / data
- ▶ It helps to compare different set of similar data
- ▶ It helps us to understand the data over a period of time.

Dr Deming was a professor of Maths and Statistics in USA. Japan called him to help improve the quality of Japanese products. He used STATISTICS for the same.

SIX SIGMA – is also a statistics term.



Random Data



Analysis



*Sensible
Results*

Collate Data

- ▶ See the data, and have a feel of it.
- ▶ Take the smallest value
- ▶ Take the largest value
- ▶ Make small intervals
- ▶ Tabulate.

1,2,1,5,3,8,5,2,4,5,8,4,5,3,6,7,4,2,1,6,9,3,5,4

Minimum = 1, Maximum = 9




Bands

1-3

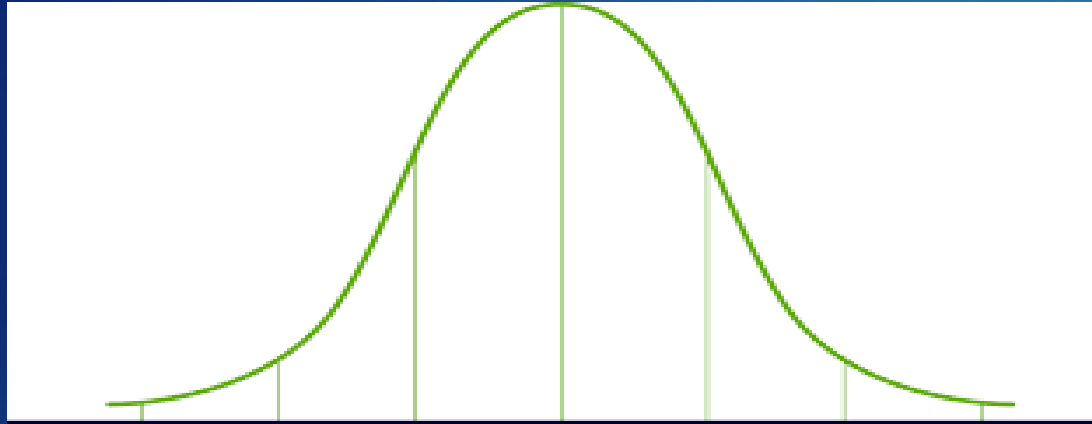
4-6

7-9

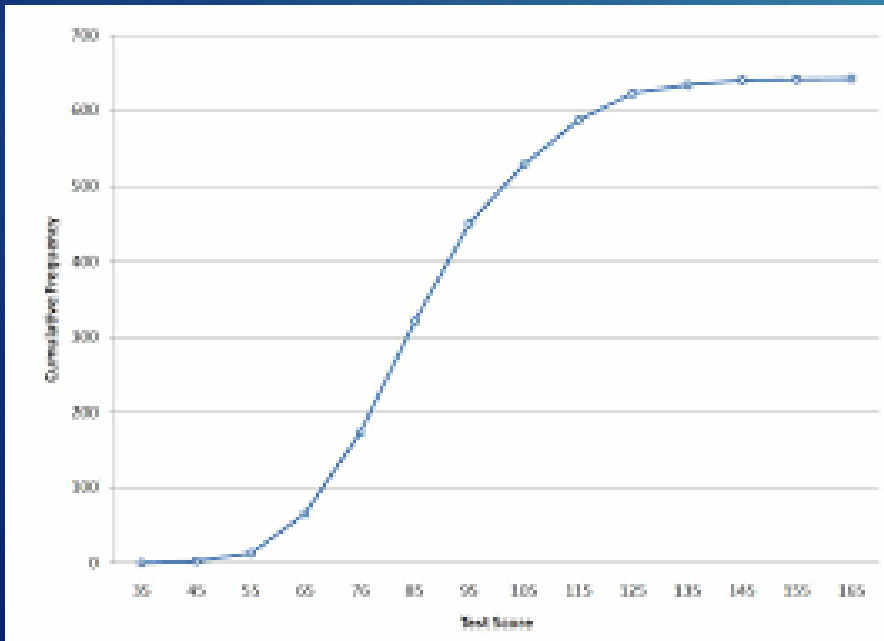
We get the data table

Range	Tally	Number
1-3		7
4-6		9
7-9		6

Different type of data representation

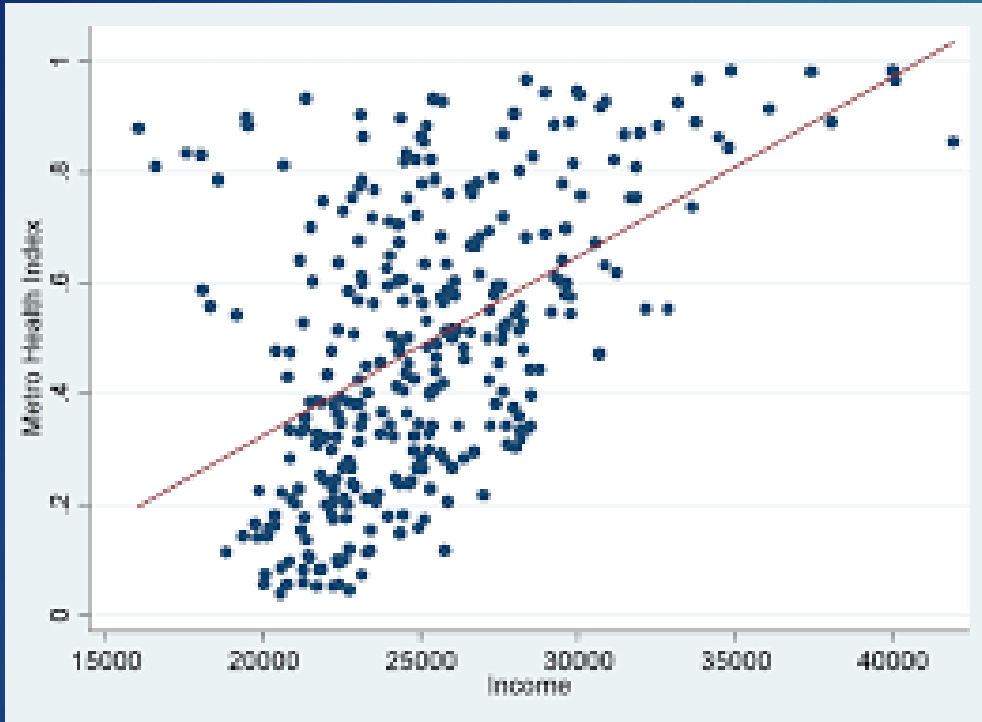


Frequency Distribution

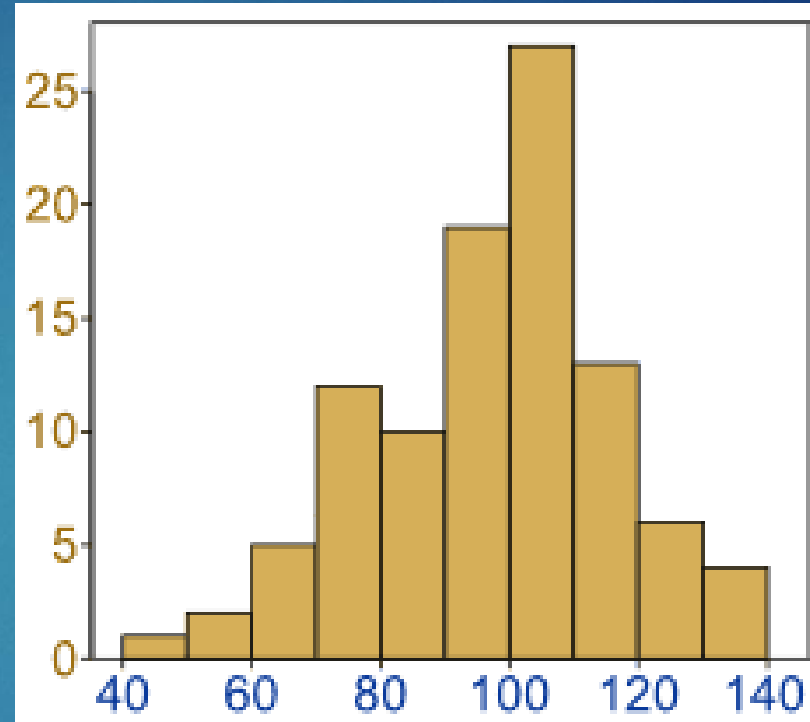


Cumulative Frequency Distribution

Graphical Representation



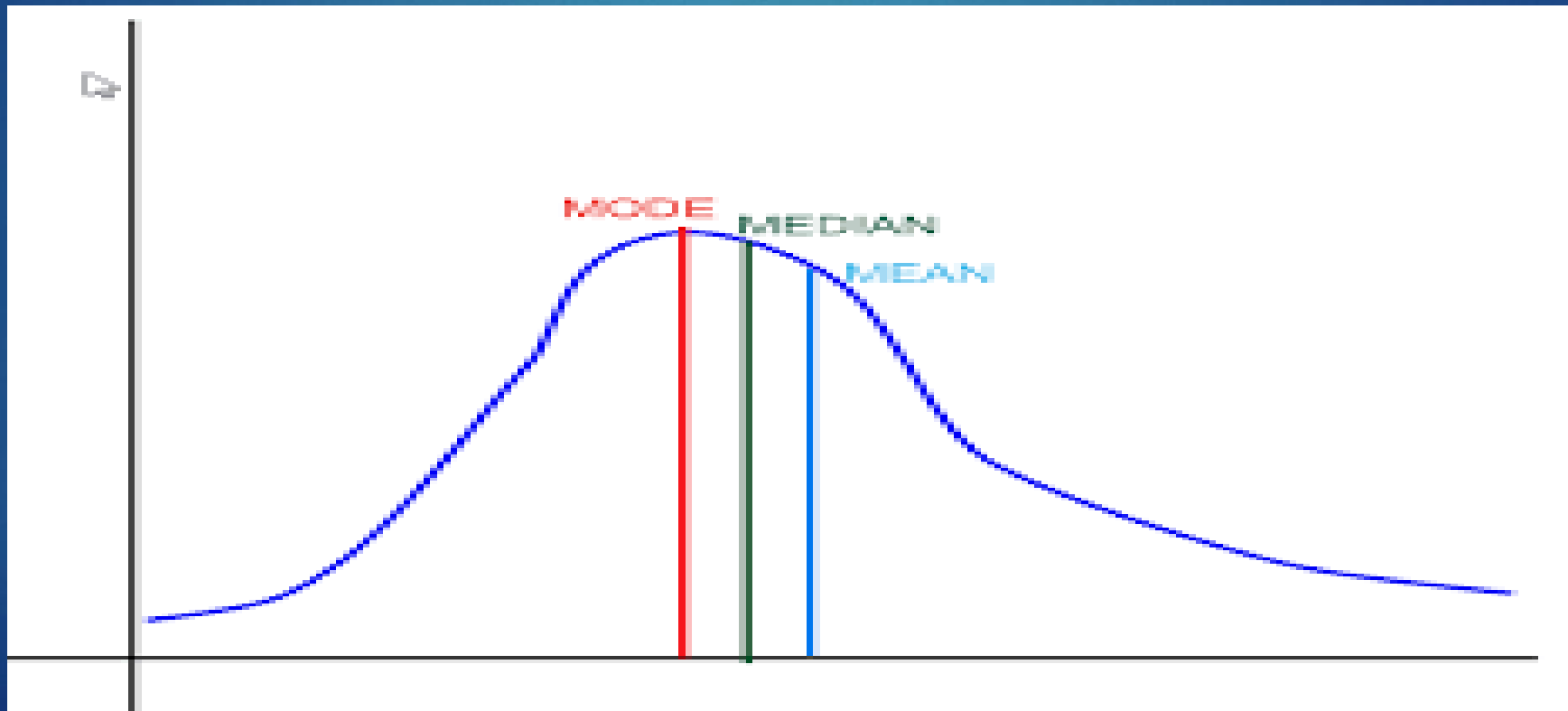
Scatter Diagram



Histogram

Basic Measurements

- ▶ Mean (Average) = Total of all values / number of values.
- ▶ Median = Central Point of the distribution
- ▶ Mode = maximum Point of the distribution.

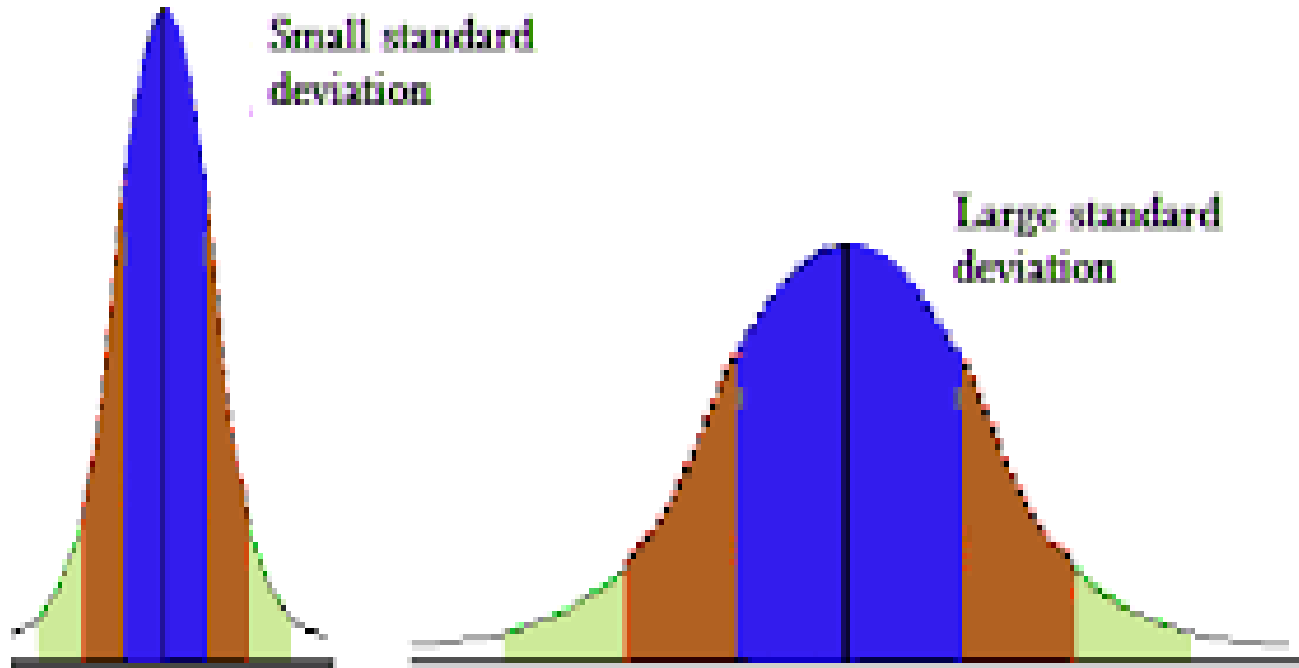


Formula's

- ▶ Means = Sum of all the values / number of values
 - ▶ Excel command :: Average
- ▶ Median – the centre point.
 - ▶ Put the numbers from low to high.
 - ▶ Count the number = say n
 - ▶ $n/2$ or $(n+1)/2$ – is the median value
 - ▶ Excel Command :: Median
- ▶ Mode :: The most common point.
 - ▶ Please see which number maximum no of times.
 - ▶ In excel command :: Mode

Dispersion

11



Standard Deviation = how much is the spread of the graph or distribution.

In a normal distribution

- Mean +/- sigma = 68% of data
- Mean +/- 2 sigma = 96% of data
- Mean +/- 3 sigma = 99.7% of data
- Mean +/- 6 sigma = 99.9999% of data

Standard Deviation

12

$$SD = ((x-\text{mean})/n) \wedge 0.5$$

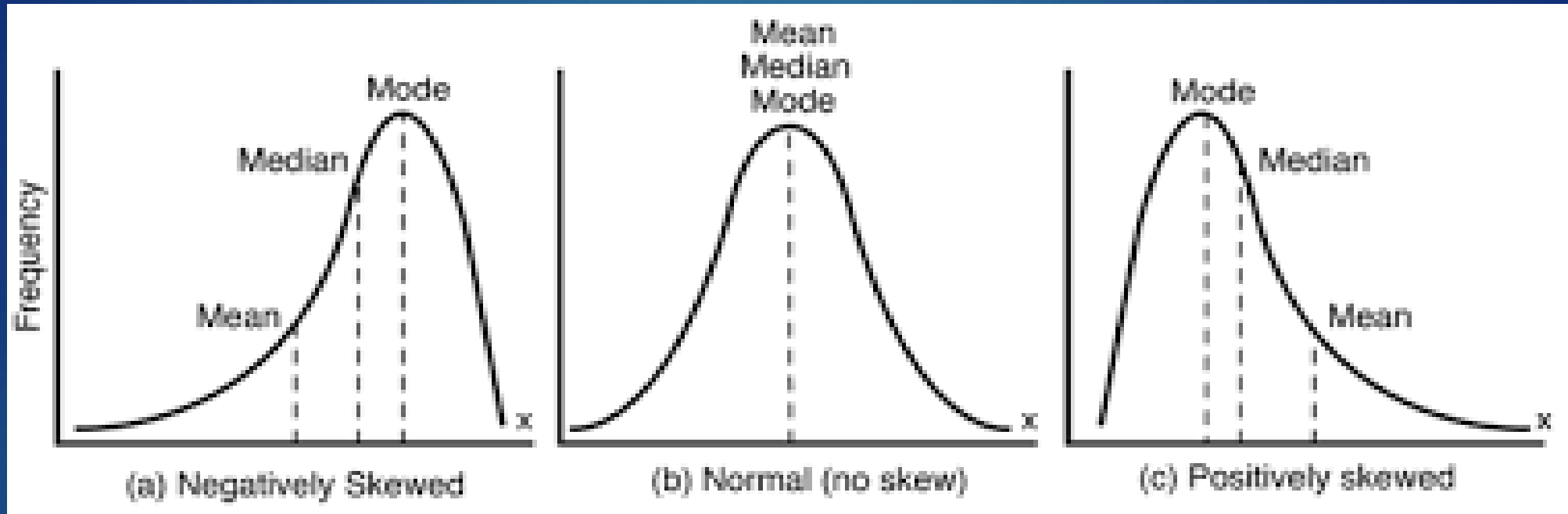
X = the data

N = number of data

This is quite an important data and is used in all types of data distributions

Skewness

13



By SK Agrawal | 9873592082 for DJF

How Much data is shifted towards one side.

Skewness

DISTRIBUTION.	EQUATION.
Bernoulli distribution.	$\frac{1-2p}{\sqrt{p(1-p)}}$
Beta distribution.	$\frac{2(b-a)}{(2+a+b)} \sqrt{\frac{1+a+b}{ab}}$
Binomial distribution.	$\frac{1-2p}{\sqrt{np(1-p)}}$
Chi square distribution.	$2\sqrt{\frac{2}{r}}$
F distribution.	$\frac{2(2n+m-2)}{m-6} \sqrt{\frac{2(m-4)}{n(m+n-2)}}$
Negative binomial.	$\frac{2-p}{\sqrt{r(1-p)}}$
Poisson Distribution.	$y^{-1/2}$

Difference between sigma & skewness

15

Sigma	Skewness
Measure how much data is spread	Measures if the data is uniform or not.
Measures amount of variations	Measures how much it is left or right of centre
How much data is distributed	Which side it is distributed

Probability

Probability is the branch of mathematics concerning numerical descriptions of how likely an event is to occur or how likely it is that a proposition is true. Probability is a number between 0 and 1, where, roughly speaking, 0 indicates impossibility and 1 indicates certainty.

Probability

Chance of an event happening.

Coin Toss	Heads	Tail
1	$\frac{1}{2}$	$\frac{1}{2}$
2	$\frac{1}{2}, \frac{1}{2}$	$\frac{1}{2}, \frac{1}{2}$
3	$\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$	$\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$

No of spin	Minimum 1 head
3	HHH, HHT, HTH, THH, HTT, THT, TTH, TTT
4	HHHH, HHHT, HHTH, HTHH, THHH, HHTT... And so on

Probability of minimum 2 heads in 4 toss is ??

H	H	H	H
H	H	H	T
H	H	T	H
H	H	T	T
H	T	H	H
H	T	H	T
H	T	T	H
H	T	T	T
T	H	H	H
T	H	H	T
T	H	T	H
T	H	T	T
T	T	H	H
T	T	H	T
T	T	T	H
T	T	T	T

Probability – one more case

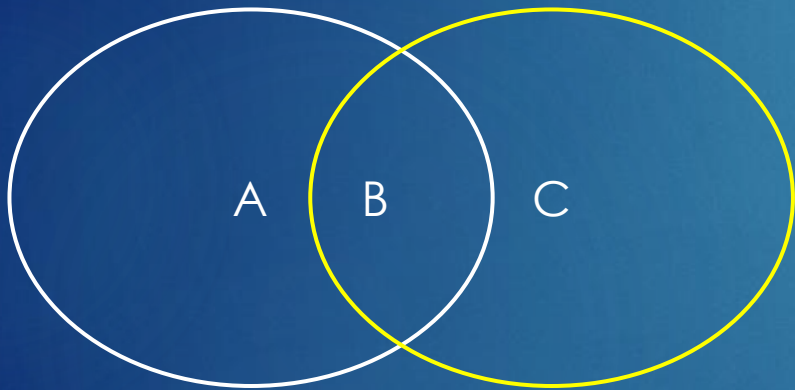


- ▶ Two dice...
- ▶ Probability of.. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 & 12.... Please calculate chance of getting a 7.

Throw	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	
Dice 1	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	5	5	5	5	5	5	6	6	6	6	6	6	6
Dice 2	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	6
Total	2	3	4	5	6	7	3	4	5	6	7	8	4	5	6	7	8	9	5	6	7	8	9	10	6	7	8	9	10	11	7	8	9	10	11	12	

Terms of Probability

- ▶ Mutually Exclusive Events
 - ▶ ONLY one can happen. Coin toss, Dice throw.
- ▶ Probability = Number of favourable events / total possible events.



In this case, the first circle = $A+B$
Second Circle = $B+C$

Union = TOTAL of the two = $A+B+C$
Interception = B

Laws of probability

- ▶ The addition law
 - ▶ For mutually exclusive event
 - ▶ $P(A) \text{ \& } P(B)$ are mutually exclusive
 - ▶ $P(A \cup B) = P(A) + P(B)$
 - ▶ If $P(A) \text{ \& } P(B)$ are not mutually exclusive.
 - ▶ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- ▶ Multiplication Law
 - ▶ Probability of A happening – provided B happens.
 - ▶ $P(A \cap B) = P(A \cap B) / P(B)$

More on probability

23

- ▶ Discrete – goes in steps – like toss of a coin
- ▶ Continuous – like – temperature in a day. Can have smaller values.

Probability Functions....

- Machine failure
- Accident rates

Probability Distribution

- ▶ Binomial Distribution
 - ▶ Very common
 - ▶ Discrete
 - ▶ Eg toss of a coin
- ▶ Poisson Distribution
 - ▶ No. of calls in a day / Trucks arriving at a toll booth.
 - ▶ They are defined by an integer.
 - ▶ They are discrete random variable.
- ▶ The normal distribution
 - ▶ Continuous distribution
 - ▶ Good for sampling
 - ▶ Close to actual life situation.
 - ▶ One peak, Mean / mode / median at the centre.

Sampling & Test of Significance

25

- ▶ Sampling is when – a small set of data is use to predict the outcome of the full batch.
 - ▶ Exit Polls – to see the number of seats.
 - ▶ Destructive Testing – gears, ball bearings
- ▶ *When we test a sample... is the data good enough representation of the whole?*
- ▶ *Can we take the decision based on that?*
- ▶ *How much is the error possible?*

Sample Selection

26

- ▶ Will be decided on
 - ▶ Uniformity of the over all data – screws manufacturing or election results
 - ▶ Minimum numbers (Usually 30 plus)
 - ▶ Criticality of the parameter to be checked.
 - ▶ Number of parameters to be checked.

Test of Significance

27

- ▶ Does the sample data – represents the full in a reliable way?
- ▶ Reliability – is the test of significance.

H_0 = null hypothesis – what we expect. H_a = alternative hypothesis – what we want to test and see.
ONLY H_0 or H_a – one will be true.

How much variation – will make the either one acceptable?

Example..

Chocolate factory – candy weight 5 gm. QA engineer says that the machine is NOT OK and the data has changed.

Day 0 – average of 30 p/c – 5.07 gms. Day 4 – average – 5.34 gms , Day 7 average – 6.22 gms.

H_0 = 5 gms. H_a – not equal to 5 gms.

Day 0 – may be OK

Day 4 – may be NOT OK

Day 7 – NOT OK.

H_0 – is wrong, and H_a is right.

The variation is quite large and is **SIGNIFICANT** (Important)

Correlation & Regression

28

	Correlation	Regression
Data sets	Not necessary to be related	Some relationship may be there
Range of value	(-1) to (+1). (-1) means – one data goes up and other comes down. (+1) means both move is same direction.	A best fit line is made. Usual is root mean square value and optimization.
Example	Positive – height and weight, El Niño and rain fall	Best fit lines, can be straight or curved lines.
Names of measurement	Pearson method, Spearman Rank,	

Both are used to predict future outcomes if a certain data point occurs.

Multiple Correlation & Regression Analysis

29

- ▶ Use of more than ONE independent variable to increase the accuracy of the estimates.
- ▶ This is a higher level of analysis.
- ▶ Helpful where we can do complex maths.
- ▶ We may not get the relationship with one, so, two or more are used.

Time Series

30

- ▶ When a measurement is carried out over a period of time
- ▶ Time interval to be same.
- ▶ Measurement data to be same in same units.

This gives the historic perspective of the actual situation.

Example:

Per Capita Income of a country .. Tells us how the country is growing in prosperity'

X – Charts on the machine... tells us how the machine is behaving in it's accuracy.

Stock price – how is it changing over time.

Index

- ▶ Very important number
- ▶ When several data points are joined in a common number, with relevant importance.
- ▶ It is a weighted average.

NIFTY Index.

Shares prices of 50 shares.

Each share has a fixed weight.

Total 50 shares are such that they represent the over all Indian Economics

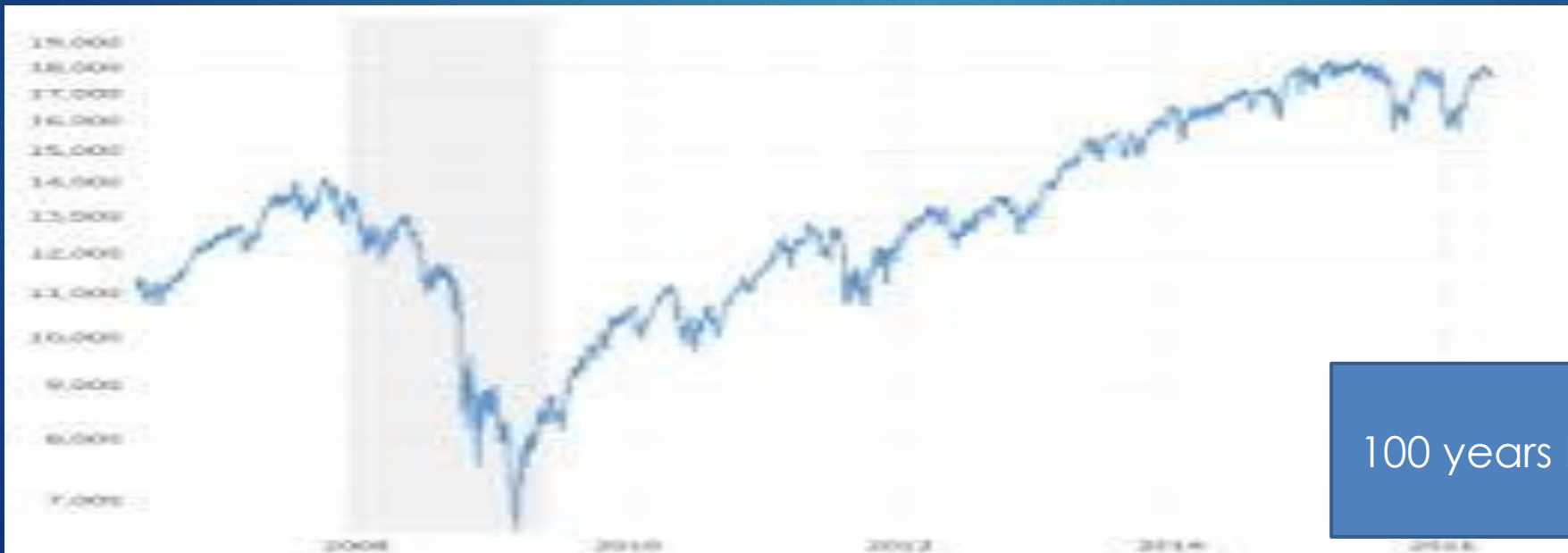
Some Important Indexes

32

- ▶ Public Indexes
 - ▶ Consumer Price Index
 - ▶ Whole sale price Index
 - ▶ Property price Index
 - ▶ Construction Cost Index.
 - ▶ Commodity Price index.
- ▶ Private Index – made by a company for it's use.
 - ▶ Labour productivity index.
 - ▶ WIP Index
 - ▶ RMC / Sale index

Index as Time Series

Index are very commonly used in time series as they give a fair description of the situation and a better decision making data set.



100 years Dow Jones

thanks

SK Agrawal

9873592082

ska@agrawalnassociates.com