CA Foundation - Quantitative Aptitude:
"Ultimate Exam Preparation Guide"

Important Formula Flowchart

Ratio and Proportion, Indices and Logarithms

RATIO:

- **Duplicate ratio** of $a : b$ is $a^2 : b^2$.
- **Triplicate ratio** of $a : b$ is $a^3 : b^3$.
- **Sub-Duplicate ratio** of $a : b$ is $\sqrt{a} : \sqrt{b}$.
- **Sub-Triplicate ratio** of $a : b$ is $\sqrt[3]{a} : \sqrt[3]{b}$.
- **Inverse ratio** of $a : b$ is $b : a$.
- **Compounded ratio** of $a : b$ and $c : d$ is $ac : bd$.

PROPORTION:

- **Cross-Product rule:** If $a : b = c : d$ then $ad = bc$
  i. e., Product of extreme terms = Product of middle terms
- If $a : b = b : c$ then $b^2 = ac$
  This proportion is **Continuous Proportion.**
  b is Mean Proportion where $b = \sqrt{ac}$ = G. M of a and c
- **Invertendo:** If $a : b = c : d$ then $b : a = d : c$
- **Componendo:** If $a : b = c : d$ then $(a + b) : b = (c + d) : d$
- **Dividendo:** If $a : b = c : d$ then $(a - b) : b = (c - d) : d$
- **Componendo and Dividendo:** If $a : b = c : d$ then $(a + b) : (a - b) = (c + d) : (c - d)$
- **Alternendo:** If $a : b = c : d$ then $a : c = b : d$
- **Addendo:** If $a : b = c : d = e : f = \ldots$ then $(a + c + e + \cdots) : (b + d + f + \cdots)$
- **Subtrahendo:** If $a : b = c : d = e : f = \ldots$ then $(a - c - e - \cdots) : (b - d - f - \cdots)$

INDICES:

- $a^m = a \times a \times a \times a \times \ldots\ldots$ m times
- $a^0 = 1$, where $a \neq 0, \infty$
- $a^{-1} = \frac{1}{a}$
- $a^{-m} = \frac{1}{a^m}$
- $a^m \times a^n = a^{m+n}$
- $\frac{a^m}{a^n} = a^{m-n}$
- $(a^m)^n = a^{mn}$

LOGARITHM:

- $\log_a(m + n) = \log_a m + \log_a n$
- $\log_a\left(\frac{m}{n}\right) = \log_a m - \log_a n$
- $\log_a m^n = n \log_a m$

- $\log_a m = \frac{\log_b m}{\log_b a}$
- $\log_a a = 1, \log_a 1 = 0$
- $\log_x(y) \times \log_y(x) = 1$
- $\log_{a^b}(x) = \frac{1}{b}\log_a(x) = \log_a(x)^{\frac{1}{b}}$
- $\log_{a^b}(x)^y = \frac{y}{b}\log_a(x)$
- $a^{x \log_a b} = b^x$
- $a^{\log_a(x)} = x$

Equations

QUADRATIC EQUATION:

- Equation: $ax^2 + bx + c = 0$
- Roots of Quadratic equation; $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$
- Sum of roots $= -\frac{b}{a} = -\frac{\text{Coefficient of } x}{\text{Coefficient of } x^2}$
- Product of roots $= \frac{c}{a} = \frac{\text{Constant term}}{\text{Coefficient of } x^2}$
- To construct a quadratic equation:

$$x^2 - (\text{Sum of roots})x + \text{Product of roots} = 0$$

Nature of Roots:

- If $b^2 - 4ac = 0$, then the roots are real and equal.
- If $b^2 - 4ac > 0$, then the roots are real and unequal (or distinct).
- If $b^2 - 4ac < 0$, then the roots are imaginary.
- If $b^2 - 4ac$ is a perfect square $(\neq 0)$, then the roots are real, rational and unequal (distinct).
- If $b^2 - 4ac > 0$, but not a perfect square then the roots are real, irrational and unequal.

CUBIC EQUATION:

- For cubic equation $ax^3 + bx^2 + cx + d = 0$
- Sum of roots $= \alpha + \beta + \gamma = -\frac{b}{a}$
- Sum of roots taken two at a time $= \alpha\beta + \beta\gamma + \gamma\alpha = \frac{c}{a}$
- Product of roots $= \alpha\beta\gamma = -\frac{d}{a}$
- Cubic Equation $= x^3 - (\alpha + \beta + \gamma)x^2 + (\alpha\beta + \beta\gamma + \gamma\alpha)x - \alpha\beta\gamma = 0$

- $\alpha^2 + \beta^2 = (\alpha + \beta)^2 - 2\alpha\beta$
- $\alpha^3 + \beta^3 = (\alpha + \beta)(\alpha^2 + \beta^2 - \alpha\beta)$
- $(\alpha + \beta)^2 - (\alpha - \beta)^2 = 4\alpha\beta$

Mathematics of Finance

SIMPLE INTEREST:

- $S.I = Pit$
- $A = P + S.I = P + Pit = P(1 + it)$

where, A = Amount, i = Annual interest rate in decimal, t = time in years

COMPOUND INTEREST:

- Amount $= P\left[1 + \frac{r}{m}\right]^{t \times m}$
- C.I. $= A - P = P\left[1 + \frac{r}{m}\right]^{t \times m} - P = P\left\{\left[1 + \frac{r}{m}\right]^{t \times m} - 1\right\}$
- S.V (Scrap Value) $=$ Cost value $(1 -$ Rate of Depreciation$)^{\text{time}}$
- Effective Rate, $E = \left(\left(1 + \frac{r}{m}\right)^{m} - 1\right) \times 100$
- Difference between S.I. and C.I.:

For 2 years	For 3 years
$= Pr^2$	$= Pr^2(r + 3)$

ANNUITY:

- **For Regular Annuity**:
- Future value $(F.V) = R\left\{\frac{(1+i)^n - 1}{i}\right\}$

  where, R = Regular periodic payment, $i = \frac{r}{m}$ and $n = t \times m$ (Total payments)
- Present value $(P.V) = R\left\{\frac{1 - (1+i)^{-n}}{i}\right\}$

- **For Annuity Immediate**:
- $F.V = R\left\{\frac{(1+i)^n - 1}{i}\right\} \times (1 + i)$
- $P.V = R\left\{\frac{1 - (1+i)^{-n}}{i}\right\} \times (1 + i)$

APPLICATION:

- **Perpetuity:**
  - Ordinary Perpetuity $(P.V) = \frac{R}{i}$
  - Immediate Perpetuity $(P.V) = \frac{R}{i}(1 + i) = \frac{R}{i} + R$
- **Growing Perpetuity:**
  - Regular $(P.V) = \frac{R}{i - g}$
  - Immediate $(P.V) = \frac{R}{i - g}(1 + i)$
- **Sinking Fund:** $F.V. = R\left[\frac{(1+i)^n - 1}{i}\right]$

- **Net Present Value:** NPV = Present value of all cash inflow - Present value of all cash outflow
  - If NPV > 0, Accept Proposal
  - If NPV < 0, Reject Proposal
- **Leasing:** Cost of Asset $=$ Rent $\left\{\frac{1 - (1+i)^{-n}}{i}\right\}$
- **Valuation of Bond:** Price at which Bond should be purchased = Present value of all future interest + Present value of Bond's maturity value
- **Compound Annual Growth Rate (CAGR):** $CAGR\ (t_0, t_n) = \left[\frac{V(t_n)}{V(t_0)}\right]^{\frac{1}{t_n - t_0}} - 1$

  where, $V(t_0)$ = Beginning Period; $V(t_n)$ = End Period

PERMUTATIONS AND COMBINATIONS

FUNDAMENTAL PRINCIPLE OF COUNTING:

- **Multiplication Rule**:

If one thing can occur in 'm' different ways, and another thing can occur in 'n' different ways, then total number of ways of doing both things simultaneously is '$m \times n$' ways.

- **Addition Rule**:

If one thing can occur in 'm' different ways, and another thing can occur in 'n' different ways, then either of these things can occur in '$m + n$' ways.

PERMUTATION:

- Number of permutations of $'n'$ different things taken $'r'$ at a time is ${}^{n}P_{r} = \frac{n!}{(n-r)!}$ where $r \leq n$
- Number of permutations of $'n'$ different things taken $'r'$ at a time and repetition is allowed is n^{r}.
- Number of circular permutations of n different things is $(n-1)!$.
- When clockwise and anticlockwise arrangements are not different, the number of circular permutation of n different things is $\frac{(n-1)!}{2}$.
- Number of permutations of n distinct objects taken r at a time when a particular object is not taken in any arrangement is ${}^{n-1}P_{r}$
- Number of permutations of r objects out of n distinct objects when a particular object is always included in any arrangement is: $r.\ {}^{n-1}P_{r-1}$
- Permutations when some of the things are alike, taken all at a time; $P = \frac{n!}{n_1!n_2!n_3!}$

COMBINATION:

- Number of ways in which selection of r things out of n different things is done is ${}^{n}C_{r} = \frac{n!}{r! \times (n-r)!}$
- Combinations of n different things taking some or all of n things at a time $= 2^{n} - 1$.
- Combination of n things taken some or all at a time when n_1 things are alike of one kind, n_2 things are alike of second kind & n_3 things are alike of third kind:

$$= (n_1 + 1) \times (n_2 + 1) \times (n_3 + 1) - 1$$

- If we have to select the combination such that r_1 things to be selected from n_1 and r_2 things to be selected from n_2 the, total selections are: ${}^{n_1}C_{r_1} \times {}^{n_2}C_{r_2}$.

- ${}^{n}C_{0} = {}^{n}C_{n} = 1$
- ${}^{n}C_{r} = {}^{n}C_{n-r}$
- ${}^{n+1}C_{r} = {}^{n}C_{r} + {}^{n}C_{r-1}$
- ${}^{n}P_{r} = {}^{n-1}P_{r} + r \cdot {}^{n-1}P_{r-1}$

SEQUENCE AND SERIES

ARITHMETIC PROGRESSION (A.P):

- nth term of an arithmetic progression: $t_n = a + (n-1)d$

  where, a = first term, d = common difference & n = position of the required term
- Sum of n terms of an arithmetic progression:
  - $S_n = \frac{n}{2}[2a + (n-1)d]$
  - $S_n = \frac{n(a+l)}{2}$ where, a = first term, l = last term & d = common difference

GEOMETRIC PROGRESSION (G.P):

- nth term of a geometric progression:
  $t_n = ar^{n-1}$, where, a = first term, r = common ratio
- Sum of n terms of an arithmetic progression:
  - If common ratio, $r > 1$, then $S_n = \frac{a(r^n - 1)}{r - 1}$
  - If common ratio, $r < 1$, then $S_n = \frac{a(1 - r^n)}{1 - r}$
- Sum of Infinite Geometric series: $S_n = \frac{a}{1 - r}$, $r < 1$

- **Arithmetic Mean (A):** $A = \frac{a + b}{2}$
- **Geometric Mean (G):** $G = \sqrt{ab}$

SETS, RELATION AND FUNCTION

SETS:

- Number of sets for n number of elements: 2^n
- Number of proper sets: $2^n - 1$

- **Cardinal Number:**

  For the finite sets A and B, then
  - $n(A \cup B) = n(A) + n(B) - n(A \cap B)$
  - $n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cap B) - n(B \cap C) - n(C \cap A) + n(A \cap B \cap C)$
  - $n(B - A) = n(B) - n(A \cap B)$

RELATION:

- Domain (R) = {a: (a, b) ∈ R}
- Range (R) = {b: (a, b) ∈ R}
- **Reflexive Relation**: (a, a) ∈ R for all a ∈ A.
- **Symmetric Relation**: (a, b) ∈ R ⇒ (b, a) ∈ R for all a, b ∈ A.
- **Transitive Relation**: (a, b) ∈ R and (b, c) ∈ R ⇒ (a, c) ∈ R, for all a, b, c ∈ A.

A relation that is Reflexive, Symmetric, and Transitive is called an **Equivalence relation**.

FUNCTION:

- **One to One function**: For $x_1 \neq x_2$ for $x_1, x_2 \in X \Rightarrow f(x_1) \neq f(x_2)$
- **Many to One function**: For $x_1 \neq x_2$ for $x_1, x_2 \in X \Rightarrow f(x_1) = f(x_2)$
- **Into Function**: The mapping f is said to be into if there is at least one element in Y that has no pre-image in X.
- **Onto Function**: The mapping f is said to be onto if every element in Y has at least one pre-image in X.

LIMIT AND CONTINUITY

LIMITS:

Let $\lim_{x \to a} f(x) = l$ and $\lim_{x \to a} g(x) = m$ where l and m are finite quantities

i) $\lim_{x \to a} \{f(x) + g(x)\} = \lim_{x \to a} f(x) + \lim_{x \to a} g(x) = l + m$

ii) $\lim_{x \to a} \{f(x) - g(x)\} = \lim_{x \to a} f(x) - \lim_{x \to a} g(x) = l - m$

iii) $\lim_{x \to a} \{f(x) \cdot g(x)\} = \lim_{x \to a} f(x), \lim_{x \to a} g(x) = lm$

iv) $\lim_{x \to a} \frac{f(x)}{g(x)} = \left\{\frac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)}\right\} = \frac{1}{m}$ if $m \neq 0$

v) $\lim_{x \to a} c = c$ where c is constant

vi) $\lim_{x \to a} cf(x) = c \lim_{x \to a} f(x)$

vii) $\lim_{x \to a} F\{f(x)\} = F\left\{\lim_{x \to a} F(x)\right\} = F(l)$

- $\lim_{x \to 0} \frac{(e^x - 1)}{x} = 1$
- $\lim_{x \to 0} \frac{a^x - 1}{x} = \log_e a (a > 0)$
- $\lim_{x \to 0} \frac{\log(1+x)}{x} = 1$
- $\lim_{x \to \infty} \left(1 + \frac{1}{x}\right)^x = e$
- $\lim_{x \to a} \frac{x^n - a^n}{x - a} = na^{n-1}$
- $\lim_{x \to 0} \frac{(1+x)^n - 1}{x} = n$

CONTINUITY:

A function y = f(x) is said to be continuous at x = a if and only if

- *f(a)* exists i.e., f is defined at *x = a*
- $\lim_{x \to a^-} f(x) = \lim_{x \to a^+} f(x) = f(a)$
- $\lim_{x \to a} f(a) = f(a)$

Basic Applications of Differential and Integral Calculus in Business and Economics

DIFFERENTIATION:

- $\frac{d}{dx}(f(x)) = f'(x)$
- $\frac{d}{dx}(x^n) = nx^{n-1}$
- $\frac{d}{dx}(e^{ax}) = ae^{ax}$
- $\frac{d}{dx}(\log x) = \frac{1}{x}$
- $\frac{d}{dx}(a^x) = \log a \,.\, a^x$
- $\frac{d}{dx}[c(\text{a constant})] = 0$
- $\frac{d}{dx}[c\, f(x)] = c\, f'(x)$

❖ **Application:**

- Cost function, C(x): C(x) = V(x) + F(x)
- Average cost (AC or C) $= \frac{C(x)}{x}$
- Average variable cost (AVC) $= \frac{V(x)}{x}$
- Average Fixed Cost (AFC) $= \frac{F(x)}{x}$
- Marginal Cost: $\frac{dC}{dx}$.
- Revenue Function: $R(x) = P.x$
- Marginal Revenue: MR $= \frac{dR}{dx}$

- Profit Function: P(x) = R(x) - C(x)
- Marginal Profit: $\frac{dP}{dx}$

INTEGRATION:

- $\int x^n \, dx = \frac{x^{n+1}}{n+1} + c, n \neq -1$
- $\int dx = x, \text{since } \int 1.dx = \int x^0 \, dx = \frac{x^1}{1} = x$
- $\int e^x \, dx = e^x + c,$
- $\int e^{ax} \, dx = \frac{e^{ax}}{a} + c$
- $\int \frac{1}{x} \, dx = \log x + c$
- $\int \frac{1}{ax+b} \, dx = \frac{1}{a} \log |ax + b| + c$
- $\int a^x \, dx = \frac{a^x}{\log_e a} + c$
- $\int c \, f(x) \, dx = c \int f(x) \, dx$, where c is constant
- $\int \{ f(x)dx \pm g(x)\} dx \, dx = \int f(x)dx \pm \int g(x)dx$
- Definite integral; $\int_a^b f(x) = F(b) - F(a)$

❖ **Properties of Definite Integration:**

- $\int_a^b f(x) \, dx = \int_a^b f(t) \, dt$ (The definite integral is independent of the variable used)
- $\int_a^b f(x) \, dx = -\int_b^a f(x) \, dx$
- $\int_a^b f(x) \, dx = \int_a^c f(x) \, dx + \int_c^b f(x) \, dx, a < c < b$
- $\int_o^a f(x) \, dx = \int_o^a f(a - x) \, dx$
- $\int_a^b f(x) \, dx = \int_a^b f(a + b - x) \, dx$
- $\int_{-a}^a f(x) \, dx = 2 \int_o^a f(x) \, dx$
- $\int_{-a}^{a} f(x)dx = 2\int_0^a f(x)dx$ if f(x) is an even function i.e., $f(-x) = f(x)$
  $= 0$ if f(x) is an odd function i.e., $f(-x) = -f(x)$

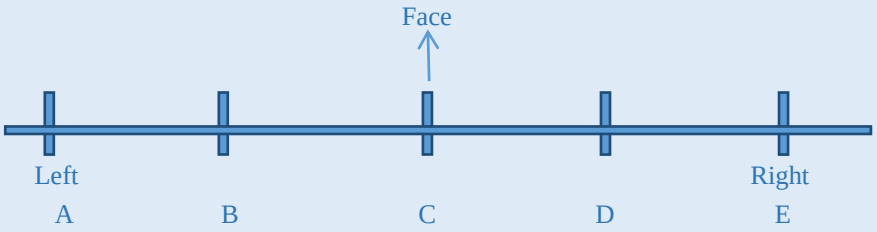
Seating Arrangements

Process of making an arrangement of a group of people to sit according to a pre-planned manner.

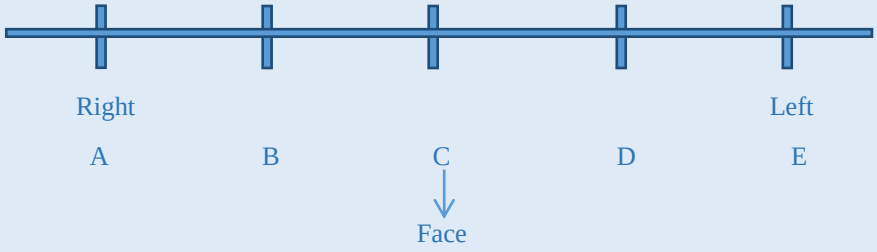
Types:

❖ **Linear Arrangement:** Here, the arrangement of the people is linear i.e. you have to arrange them in a line.

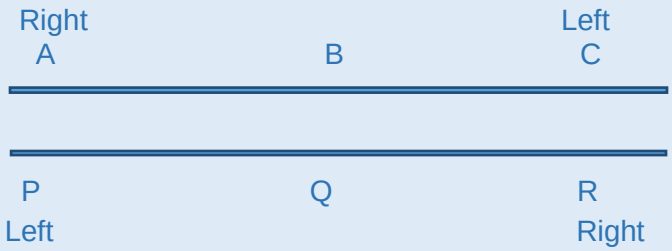
(a) When the direction of face is not mentioned: In this case, we take the directions our-self as base and the diagram will be as follow.



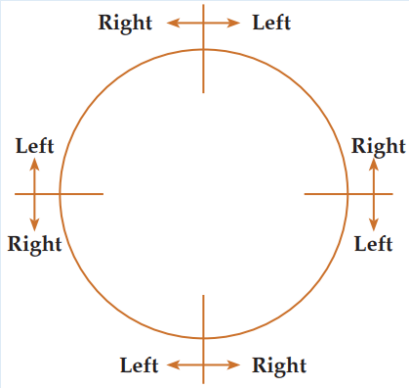
(b) When the direction of face is towards you and the diagram will be as follow;



❖ **Double row arrangement:** There will be two group of persons. Here, we arrange one group in one row and the other group in other row. The people in these rows normally face each other.



❖ **Circular Arrangement:** In this types of arrangement, people or objects are arranged in a circle or around a round table. The main aspects to consider are:



1. Moving to the left follows a clockwise direction.
2. Moving to the right follows an anti-clockwise direction.

- The illustration represents four individuals, but the directional rules remain the same regardless of the number of people.
- For rectangular and sequential arrangements, directions are considered similarly to those used in a two-row sequence.

Blood Relations

Some Important relations:

✧ Children of same parent: Siblings

✧ One's husband or Wife: Spouse

◆ Relatives on Mother's side: Maternal

- Relatives on Father's side: Paternal

✧ Mother's or father's father: Grandfather (Maternal grandfather/ Paternal grandfather)
✧ Mother's or father's mother: Grandmother (Maternal grandmother /Paternal grandmother)
✧ Mother's or father's brother: Uncle
✧ Mother's or father's sister: Aunt
✧ Wife's father or husband's father: Father-in-law
✧ Wife's mother or husband's mother: Mother-in-law
✧ Sons of Wife: Daughter-in-law
✧ Daughter's husband: Son-in-law
✧ Husband's or Wife's brother: Brother-in-law
✧ Husband's or Wife's sister: Sister-in-law
✧ Brother's or Sister's daughter: Niece
✧ Brother's or Sister's son: Nephew
✧ Uncle's or Aunt's son or daughter: Cousin or first cousin
✧ Son or Daughter of the first cousin: Second cousin
✧ Father of grandfather or Father of grandmother: Great grandfather
✧ Mother of grandfather or Mother of grandmother: Great grandmother

Statistical Description of Data

❖ Statistics is originated from
- Latin word: Status
- Italian word: Statista
- German word: Statistik
- French word: Statistique

Statistics in Singular Sense: It's the statistical tool that is employed for collecting, analysing and presenting data, leading finally to drawing statistical inferences about some important characteristics. Ex; Mean, Dispersion etc.

Statistics in Plural Sense: Defined as data qualitative as well as quantitative, usually collected with a view to have statistical analysis.Ex; Numerical data relating to national income, literacy etc.

Application of Statistics:
→ Economics
→ Business Management
→ Commerce & Industry

Limitation of Statistics:
→ Deals in aggregates only
→ Does not deal in qualitative Data
→ Not always true
→ Expert consultation required

Data: It's the collection of quantitative information about some particular characteristics.
Variable: It's a measurable quantity
- Discrete variable: Variables that can only take on a finite number of values, Ex; No. of accidents etc
- Continuous variable: Variable that has an infinite number of possible values. Ex; Weight etc.
- Attribute: A qualitative characteristic is Attribute. Ex; Nationality of a person etc.

❖ **Types of Data**:

	Primary Data	Secondary Data
Meaning	Collection of original data for the first time.	Compilation of existing data
Who collects/compiles?	Collected by the investigator or his agents	Compiled by persons other than who collected the primary data
Cost	Relatively more costly	Relatively less costly
Use	Useful for the purpose of enquiry	Made to suit the purpose of enquiry
Personal	Possibility of personal prejudice in its collection	No possibility of personal prejudice in its compilation since such data are already collected.

❖ **Methods of collecting Primary Data:**

- **Interview data:** 1) Personal Interview 2) Indirect Interview 3) Telephonic Interview

Investigator directly contacts the persons about whom information is to be collected.

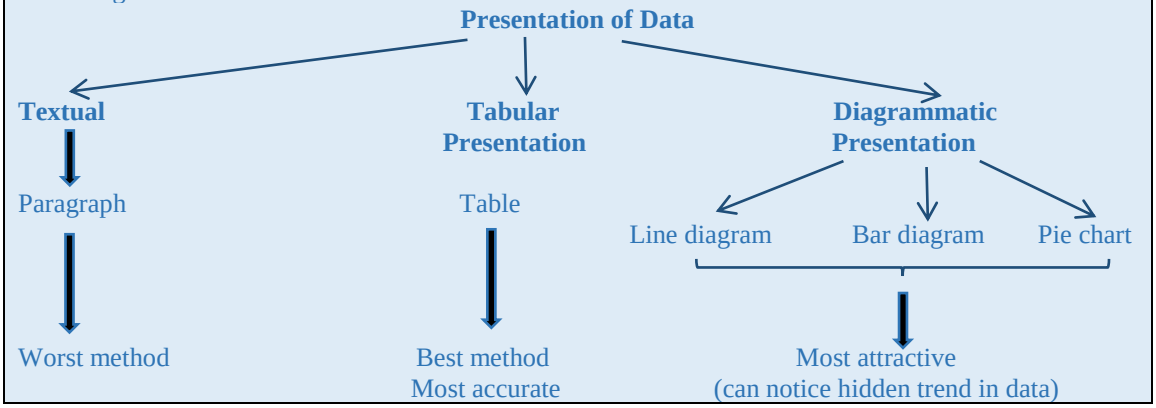
- **Mailed Questionnaire method:** It's a form of questionnaire which is mailed to targeted individuals, which has a collection of questions on a particular topic asked to them as a part of interview which is used for conducting research on that topic. It is very low costly.
- **Observation method:** Used for research to gather data about people, objects, events etc.

  → Accurate Data
  → Time consuming
- **Questionnaires filed and sent by enumerators:** It involves sending schedules through the enumerators. The enumerators contact the informants, get replies to the questions contained in a schedule and fill them in their own handwriting in the questionnaire form. It is costly and time consuming method.

SCRUTINY:

Classification of Data:

- **Chronological or Temporal**: On the basis of time
- **Geographical or Spatial:** On the basis of Location (Place)
- **Qualitative (or Ordinal):** On the basis of characteristics or attributes like social status etc.
- **Quantitative or Cardinal:** On the basis of magnitude (a) discrete (b) continuous such as height, weight etc.



(i) **Textual method**: Method, communication researchers use to describe and interpret the characteristics of a recorded or visual message.
(ii) **Tabular method:**
1) Table Number
2) Title (which explains content of tables)
3) Headnote (Given in brackets information about the units)
4) Caption - Describes column & Subcolumns

Box Head ⬇ Upper part of table includes column & sub-column number

5) Stubs - Row head
6) Body of Table - Numeric information
7) Source of Data
8) Footnote - Specific feature which is not explained

Diagrammatic Method:

- ✧ Line diagram or Histogram
- ✧ Bar diagram
- ✧ Pie diagram

❖ **Some Statistical terms:**

- ◆ **Range**: Difference between the largest and smallest observations in the given data.
- ◆ **Class**: A large number of observations varying in a wide range are usually classified in several groups according to the size of their values and the groups are bounded by limits, then each group is called a class.
- ◆ **Class limit:** The smallest and largest possible values in each class of frequency distribution table.
- ◆ **Length or class interval:** Difference between the lower limit and upper limit of the class.

❖ **Methods used in classification of Data:**

- ◆ **Exclusive Method**: Value of the upper limit of a class is excluded from that class interval but included in the lower limit of the next class interval.
- ◆ **Inclusive Method**: Value of the upper limit of a class is included in that very class interval.

❖ **Class Boundary:** Actual class limit of a class interval.

✧ Lower class boundary = LCB = Lower limit class $-\frac{1}{2}$D

✧ Upper class boundary = LCB = Lower limit class $-\frac{1}{2}$D

❖ **Class-mark or Mid point or Mid value:** Central value of the class interval.

Mid -value or class $= \frac{\text{LCb} + \text{UCB}}{2}$

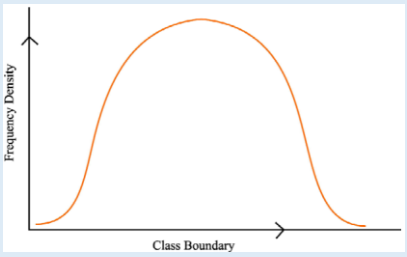
❖ **Frequency Density:** Frequency density $= \frac{\text{Frequency of class interval}}{\text{Length or size or width of class interval}}$

Graphical Representation of a Frequency Curve

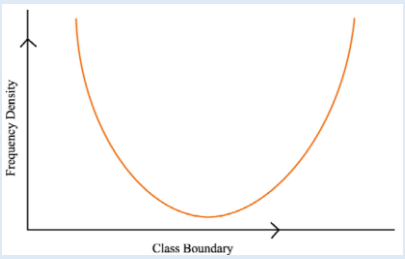
Histogram	Frequency Polygon	Ogive
2-D	1-D	1-D

❖ **Frequency curve:**

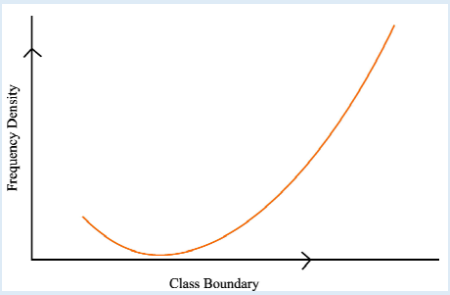
✧ Bell-shaped curve:



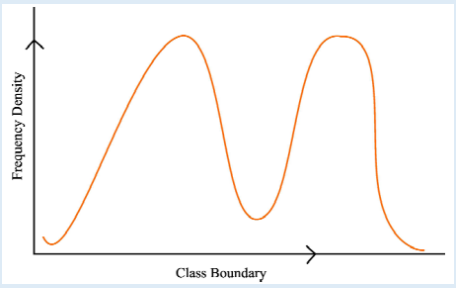
✧ U-shaped curve:



✧ J-shaped curve:



✧ Mixed curve:



SAMPLING

Branches of Statistics

Descriptive Stats
Gathering, Classifying, Summarizing data from samples

Inferential Stats
It involves drawing conclusion, generalization or making prediction from the gathered data

Population:
→ Group of People
→ Group of objects
→ Group of events & observations

Ex: Height of male students, Blood Pressure of females between ages 40 to 60 years & Temperature in June.

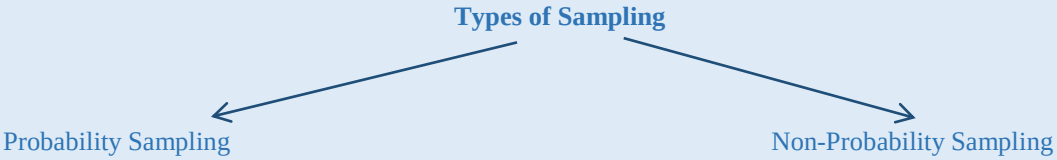
❖ **Sample:** Sample is a subset of Population
"Small group of elements selected from populations."
Number of elements in a sample is called the sample size.

❖ **Parameter:** Measurable characters of populations are called Parameters.
Ex: Population mean $= \mu$
Population variance $= \sigma^2$

❖ **Principles of Sampling**: Sampling is the procedure of selecting elements for a sample from population so that inferences can be drawn about population from sample.

Some Basic Principles of Sampling:

- **Law of Statistical Regularity**: This law suggest that if a large sample is taken randomly from population, it will possess almost same characters of population.
- **Law of Inertia of large numbers:** This law is the corollary of "Law of Statistical Regularity". This law says that "Larger the size of sample, more accurate the results are."
- **Principle of Optimization:** Maximum efficiency at minimum cost can be achieved only when appropriate "Sampling design" is selected.
- **Principle of Validity: A**ccording to this law, sampling design is valid only if it is possible to obtain accurate estimates about population.



❖ **Population Sampling (Random Sampling):** When elements are randomly selected for sample.
This can be done in 4 ways:
1) **Sample Random Sampling: (SRS):** In this sampling, each member of population has an equal chance of being selected in sample.

2) **Systematic Sampling: (Quasi Random sampling):** In this sampling, every member of the population is assigned a number. This first member is randomly selected and then instead of choosing other randomly, we chose them in regular intervals.

3) **Stratified Sampling:** In this sampling, members are divided into subgroups called strata based on gender, Age and Income etc. After that, members are selected using Random or Systematic sampling from each subgroup.

4) **Cluster Sampling (Multi stage Sampling):**

5) When population size is large, divide the population in subgroups (each subgroup has similar characteristic of the whole population). Then some groups are selected randomly and then members are selected from them for sample

❖ **Non-Population Sampling (Random Sampling):** Each member is not selected randomly, so valid inferences can not be made in this type of sampling.

- **Purposive or Judgement sampling**: Based on the opinion of expert. Ex: Indian idol
- **Convenience sampling:** Those elements are selected which are easily accessible to researcher. Ex: Asking your students to complete survey regarding services provided by universities.
- **Volunteer response sampling:** People who are themselves ready to conduct the survey collect the sample data.
- **Snowball Sampling:** First select some members, then with the help of them select some more & process continues.

Errors in Sample survey

Sampling Errors | **Non-Sampling Errors**

❖ **Sampling Errors:** Difference between Sample statistics & population parameter because sample was not the true representative of population.
→ Faulty sampling method
→ Faulty Demarcation of sampling units
→ Replacing sampling unit with unsuitable unit
→ Wrong choice of statistic

❖ **Non-Sampling Errors:** These are human errors, census & sampling both can have these errors.
→ Lapse of memory
→ Preference for certain digits
→ Wrong measurements
→ Untrained interviewer
→ Biased opinion

❖ **Population:** Aggregate of all units under consideration. Population size (Denoted by 'N')

- **Finite Population:** Finite countable elements
- **Infinite Population**: Uncountable elements

✧ Sample size is denoted by 'n'.
✧ Detailed & complete list of all sampling units is known as Sampling frame.

- Population mean $(\mu) = \frac{\sum_{i=1}^{N} x_i}{N}$
- Population Porportion $(P) = \frac{X}{N}$
- Population S.D. $(\sigma) = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$

❖ **Statistics**: Measurable characters of sample

$\hat{\mu} =$ Sample mean $= \bar{x} = \frac{\Sigma x_i}{n}$

$\hat{x} =$ Sample S. D $= S = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$

$\hat{p} =$ Sample proportion $= p = \frac{x}{n}$

Total no. of samples with replacement $= (N)^n$
Total no. of samples without Replacement $= {}^{N}C_n$

❖ **Sampling Distribution**:

- Standard Error of mean $(\bar{x}) = \frac{\sigma}{\sqrt{n}}$ for SRSWR

- Standard Error of mean $= \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ for SRSWOR

  SRSWR → Simple Random sampling with Replacement
  SRSWOR → Simple Random sampling without Replacement

Standard Error for Proportion:

❖ $SE(P) = \sqrt{\frac{pq}{n}}$ for SRSWR

❖ $\mathrm{SE}(P) = \sqrt{\frac{Pq}{n}}\sqrt{\frac{N-n}{N-1}}$ for SRSWOR

$\sqrt{\frac{N-n}{N-1}}$ = finite Population multiplier

✧ Expectation of Sampling Distribution $= E(x) = \sum p_i x_i$

✧ Variance of Sampling Distribution $= \sum p_i {x_i}^2 - (\sum p_i x_i)^2$

CENTRAL TENDENCY & DISPERSION

CENTRAL TENDENCY:

It identifies a single value as representative of an entire distribution.

❖ **Measures of Central Tendency**:
- ✧ Arithmetic Mean
- ✧ Median
- ✧ Mode
- ✧ Geometric Mean
- ✧ Harmonic Mean .

➢ **Arithmetic Mean:**

$\mathrm{AM} = \frac{\text{Sum of all observations}}{\text{Total no.of observations}}$

Methods

Direct Method

$\bar{X} = \frac{\sum f_i x_i}{N}$

where $N = \Sigma f_i$

Shortcut method

Assume mean method

If $x_i - A = d_i$ then $\bar{x} = A + \frac{\sum f_i d_i}{N}$

Step deviation method

If $\frac{x_i - A}{h} = v_i$ then $\bar{x} = A + \frac{\sum f_i v_i}{N} \times h$

- ✓ If all observations are same (let k) then mean will also be k.
- ✓ Sum of deviations from the arithmetic mean is always zero i.e., $\sum x_i - \bar{x} = 0$
- ✓ Sum of squares of Deviation is minimum when deviations are taken from their arithmetic mean i.e., $\sum (x_i - k)^2$ is minimum when $k = \bar{x}$
- ✓ Combined mean, $\bar{x}_{12} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2}$

- ✓ Arithmetic mean changes with change of origin & charge of scale i.e., If $y_i = ax_i + b$, then $\bar{y} = a\bar{x} + b$.

➢ **Median:**

It in the middle observed value of the data when all the observed values are arranged wither in ascending order or in descending order.

- ✓ Median changes with charge of origin & change in scale i.e if $y = a + bx$, then median of $y = a + b$(median of x)
- ✓ Sum of absolute deviation is minimum when deviations are taken from the median i.e. $\Sigma|x_i - k|$ is minimum when $k =$ median

◆ **Partition values (Fractiles):**

Data	Median	Quartiles	Deciles	Percentiles
Ungrouped & Discrete Data	$M = \left[\left(\frac{n+1}{2}\right)\right]^{th}$ term	$Q_k = \left[k\left(\frac{n+1}{4}\right)\right]^{th}$ term	$D_k = \left[k\left(\frac{n+1}{10}\right)\right]^{th}$ term	$P_k = \left[k\left(\frac{n+1}{100}\right)\right]^{th}$ term
Continuous data	$M = l + \left[\frac{\frac{N}{2} - Cf}{f}\right] \times h$	$Q_k = l + \left[\frac{\frac{kN}{4} - CF}{f}\right] \times h$	$D_k = l + \left[\frac{\frac{kN}{10} - CF}{f}\right] \times h$	$P_k = l + \left[\frac{\frac{kN}{100} - CF}{f}\right] \times h$

➢ **Mode:**

An observation with highest frequency.

- Discrete Series: The observation having maximum frequency.
- Continuous Series:

$$\text{Mode} = l + \left\{\frac{f_1 - f_0}{2f_1 - f_2 - f_0}\right\} \times h$$

- ✓ Mode also changes with change of origin or change in scale.

Relation between Mean, Median and Mode:

Mean - Mode = 3(Mean - Median) or 3 Median = 2 Mean + Mode

➢ **Geometric Mean:** It's the n^{th} root of the product of n observations.

- For individual series: $GM = (x_1 \times x_2 \times \cdots x_n)^{\frac{1}{n}}$
- For Discrete & Continuous series: $GM = \left[(x_1)^{f_1} \times (x_2)^{f_2} \times \cdots \times x_n^{f_n}\right]^{\frac{1}{N}}$

- ✓ If all items are same (Let k) then $GM = k$
- ✓ $\log(GM) = \frac{\Sigma \log x_i}{N}$ or $GM = AL\left[\frac{\Sigma \log x_i}{N}\right]$
- ✓ $GM(xy) = GM(x) \times GM(y)$
- ✓ $GM\left(\frac{x}{y}\right) = \frac{GM(x)}{GM(y)}$
- ✓ Combined Geometric mean; $G = [(G_1)^{N_1} \times (G_2)^{N_2}]^{\frac{1}{N_1 + N_2}}$

➢ **Harmonic Mean:** Reciprocal of Average of Reciprocal of all 'N' observations

- For individual series; $HM = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{N}{\Sigma\left(\frac{1}{x_i}\right)}$

- For Discrete & Continuous series; $HM = \frac{N}{\Sigma\left(\frac{f_i}{x_i}\right)}$
  - ✓ If all items are same (let k) then $HM = K$
  - ✓ Combined $HM = \frac{N_1 + N_2}{\frac{N_1}{H_1} + \frac{N_2}{H_2}}$

➢ **Relation between AM, GM & HM:**

✧ $AM \geq GM \geq HM$

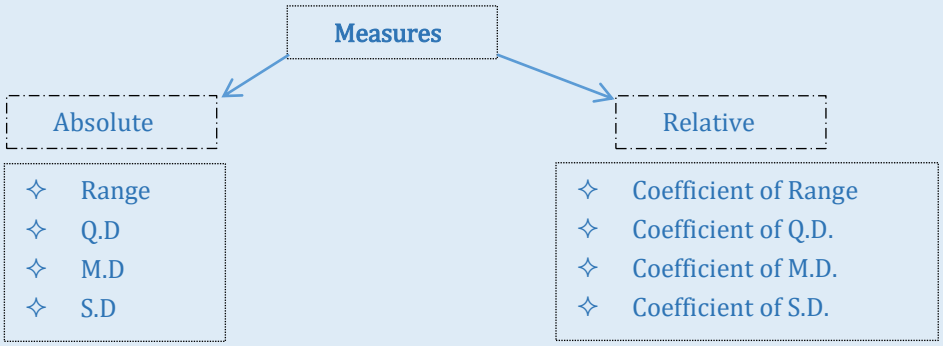
✧ If all items are different, then $AM > GM > HM$

✧ For any two items a & b, then $AM \times HM = GM^2$

DISPERSION:

The degree of scatterness or spread or variation of the variable about a central value.

❖ **Measures of Dispersion**:



➢ **Range:** Difference between largest and smallest observation; $R_x = L - S$

✧ Coefficient of Range $= \frac{L - S}{L + S} \times 100$

- ✓ Range does not change with origin.
- ✓ Range changes with scale i.e., If $y_i = a + bx$, then $R_y = |b| \times R_x$

➢ **Quartile Deviation:**

✧ $\text{Inter} - \text{Quartile Range} = Q_3 - Q_1$

✧ $\text{Semi} - \text{Quartile Range (Quartile Deviation)} = \frac{Q_3 - Q_1}{2}$

✧ Coefficient of $Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$ or $\frac{QD}{\text{Median}} \times 100$ (Only for symmetrical distribution)

- ✓ QD does not change with origin.
- ✓ QD changes with scale i.e., If $y_i = a + bx$, then QD of $y = |b| \times$ QD of x

➢ **Mean Deviation:** The average of absolute deviations taken from mean, median or mode.

MD

About Mean

$MD_x = \frac{\sum f_i |x_i - \bar{x}|}{N}$

About Median

$MD_m = \frac{\sum f_i |x_i - m|}{N}$

✧ Coefficient of $M.D = \frac{MD}{\text{Mean}} \times 100$ or $\frac{MD}{\text{Median}} \times 100$

- ✓ MD does not change with origin.
- ✓ MD changes with scale i.e., If $y = a + bx$, then $MD_y = |b| \times MD_x$

➢ **Standard Deviation & Variance:**

- ◆ Variance $= \sigma^2$ and S. D. $= \sigma \Rightarrow$ S. D. $= \sqrt{\text{variance}}$ or Variance $= (\text{S.D.})^2$
- ◆ S. D. $(\sigma) = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{N}}$
- ◆ S. D. $= \sqrt{\frac{\sum f_i x_i^2}{N} - \left(\frac{\Sigma f_i x_i}{N}\right)^2}$
- ◆ S. D. $= \sqrt{\frac{\sum f_i {d_i}^2}{N} - \left(\frac{\sum f_i d_i}{N}\right)^2}$, where $d_i = x_i - A$
- ◆ S. D $= \sqrt{\frac{\Sigma f_i {u_i}^2}{N} - \left(\frac{\Sigma f_i u_i}{N}\right)^2} \times h$, where $\frac{x_i - A}{h} = u_i$

- ✓ S. D. does not change with origin.
- ✓ S. D. changes with change of scale i.e., If $y_i = a + bx_i$, then $SD_{y_i} = |b| \times SD_{x_i}$ & Variance of $y_i = b^2 \times$ variance of x_i

✧ Coefficient of variance(C.V) $= \frac{\sigma}{\bar{x}} \times 100$

✧ S.D. of first 'n' natural number $= \sqrt{\frac{n^2 - 1}{12}}$

✧ S.D. of two number a & $b = \frac{|a - b|}{2}$

✧ Q. D : MD : SD $= 10 : 12 : 15$

✧ Combined S.D. $= \sqrt{\frac{N_1(\sigma^2 + {d_1}^2) + N_2(\sigma_2^2 + d_2^2)}{N_1 + N_2}}$ where, $d_1 = \overline{x_{12}} - \overline{x_1}$, $d_2 = \bar{x}_{12} - \bar{x}_2$ & $\overline{x_{12}} = \frac{N_1\overline{x_1} + N_2\overline{x_2}}{N_1 + N_2}$

PROBABILITY

The chances of occurrence of an event is called Probability. Probability of an event is denoted by P(E).

- ✧ $0 \leq P(E) \leq 1$
- ✧ $P(E) + P(\text{not } E) = 1$
- ✧ $P(\text{not } E) = 1 - P(E)$

❖ **Random experiment:** Any experiment produce more than one possible outcomes & where the result can not be predicted. Ex; Tossing a coin, Throwing two dice together etc.

❖ **Sample Space:** Set (collection) of all possible outcomes which are associated to a random experiment.

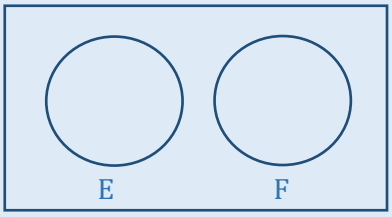
❖ **Event:** It is a subset of sample space
→ Simple Event: It contains only one element.
→ Compound Event: It contain more than one element.

→ Impossible Event: It can never (empty) happen i.e., $P(E) = 0$
→ Sure Event: It will definitely happen i.e., $P(E) = 1$

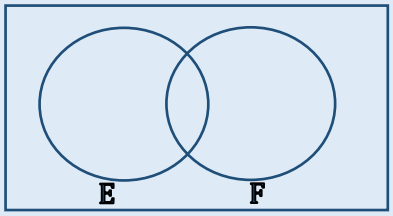
→ Equally Likely Events: Two or more events with same probability.

→ Complimentary Events: Not happening of an event is known as complimentary event. It is denoted by E' or $\bar{E}$.

❖ **Mutually Exclusive Events:** When selection of one event results in rejection of the events or when two events cannot happen simultaneously or when two events have nothing or common.



$E \cap F = \phi$
(Mutually Exclusive)



$E \cap F \neq \phi$
(Not Mutually Exclusive)

❖ **Mutually Exhaustive Events:** Two or more events are mutually exhaustive if their union makes sample space i.e., $A \cup B = S \Rightarrow P(A \cup B) = 1$.

➢ **Classical (Priori) Definition of Probability**:
Assumes
→ All outcomes are known
→ All outcomes are equally likely
→ All outcomes are mutually exclusive

◆ $P(E) = \frac{Total\ no.of\ favorable\ outcomes}{Total\ no.of\ possible\ outcomes}$

➢ **Relative Frequency Definition of Probability:**

$n \Rightarrow$ No. of times an experiment is repeated & $n \to \infty$ (n is a big member)
$f_A \Rightarrow$ An event 'A' is repeated f_A times
Then Probability of event A

$$P(A) = \lim_{n\to\infty} \frac{f_A}{n}$$

➢ **Axiomatic (Modern) Definition of Probability:**
Axiom: Any rule which don't require proof

Under this axiomatic approach:
→ $P(A)$ is always non negative i.e., $P(A) \geq 0$
→ Probability of sample space is a sure event i.e., $P(S)=1$
→ If A & B are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$

➢ **Odds in Favour & Against:**
Odds in favour of $A = A : \bar{A}$
Odds against of $A = \bar{A} : A$

$$P(A) = \frac{A}{A + \bar{A}} \ \& \ P(\bar{A}) = \frac{\bar{A}}{A + \bar{A}}$$

❖ **Set Theory in Probability:**

- $P(A \text{ or } B) = P(A \cup B) = P(\text{alteast one set}) = P(A) + P(B) - P(A \cap B)$
- $P(A \text{ and } B) = P(A \cap B) = P(A) + P(B) - P(A \cup B)$
- $P(A \text{ but not } B) = P(\text{Only } A) = P(A - B) = P(A \cap \bar{B}) = P(A) - P(A \cap B)$
- $P(B \text{ but not } A) = P(\text{Only } B) = P(B - A) = P(B \cap \bar{A}) = P(B) - P(A \cap B)$
- $P(\text{Neither A nor B}) = P(A' \cap B') = P(A \cup B)' = 1 - P(A \cup B)$
- $P(\text{Not A or not B}) = P(A' \cup B') = P(A \cap B)' = 1 - P(A \cap B)$

❖ **Conditional Probability:**

- $P\left(\frac{A}{B}\right) =$ Probability of event of A when event B has already occurred
- $P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}$, where $P(B) \neq 0$
- $P\left(\frac{A}{\bar{B}}\right) = \frac{P(A \cap \bar{B})}{P(\bar{B})}$, where $P(\bar{B}) \neq 0$
- $P\left(\frac{\bar{A}}{\bar{B}}\right) = \frac{P(\bar{A} \cap \bar{B})}{P(\bar{B})}$, where $P(\bar{B}) \neq 0$

❖ **Compound Probability theorem:**

$$P(A \cap B) = P(A) \times P\left(\frac{B}{A}\right)$$

$$P(A \cap B \cap C) = P(A) \times P\left(\frac{B}{A}\right) \times P\left(\frac{C}{A \cap B}\right)$$

❖ **For Independent Events:**

- A and B are independent events if $P(A \cap B) = P(A) \times P(B)$

  'or'
- $P\left(\frac{A}{B}\right) = P(A)$ and $P\left(\frac{B}{A}\right) = P(B)$

❖ **Total Probability theorem:** $P(A) = P(E_1)\, P\left(\frac{A}{E_1}\right) + P(E_2)\, P\left(\frac{A}{E_2}\right)$

❖ **Random Variable & Probability Distribution:**

- Mean $(\mu) = E(x) = \sum x_i P_i$
- Variance $(\sigma^2) = E(x) = \sum P_i (x_i - \mu_i)^2$ or $\sum P_i x_i^2 - (\sum x_i P_i)^2$ or $E(x^2) - [E(x)]^2$
- S.D $= \sqrt{\text{Variance}}$

  For a constant k,
  - ✓ E (k) = k
  - ✓ E(x + y) = E(x) + E(y)
  - ✓ E(kx) = k E(x)
  - ✓ If x and y are Independent, E(x.y) = E(x). E(y)

THEORETICAL DISTRIBUTION

The tabulation of different values of random variables and their corresponding probabilities.

❖ **Binomial Distribution:** Also known as Bernoulli Trials

According to Bernoulli, If an experiment is performed 'n' times, then n = no. of trials

- No. of trials is a finite positive integer
- Outcomes of each trials are categorized as success (P) & failure (Q)
- In each trial, probability of success & failure remain same (Trials are independent)

  $p + q = 1$ & $q = 1 - p$
- If x is a random variable, then $x = 0, 1, 2, 3, \ldots, n$

  $x \sim B(n, p)$

  $f(x = r) = {}^{n}C_r\, p^r q^{n-r}$

  for r=0,1, 2, 3,⋯ n

- $P(x = 0) + P(x = 1) + \cdots + P(x = n) = 1$

✓ Binomial Distribution is biparametric.

✓ Mean $(\mu) = np$, Variance $(\sigma^2) = npq$, ***S.D.*** $(\sigma) = np$

→ Maximum variance $= \frac{n}{4}$ when $p = q = \frac{1}{2}$

✓ Additive property

If $x \sim B(n_1, p)$

$y \sim B(n_2, p)$

then $x + y \sim B(n_1 + n_2, p)$

✓ Mode of Binomial Distribution depends on the value of $(n + 1)p$

- If $(n + 1)p$ is non-integer, then Mode = $[(n + 1)p]$ (greatest integer)
- If $(n + 1)p$ is integer, then there are two modes
  - ✧ First mode $= (n + 1)p$
  - ✧ Second mode $= (n + 1)p - 1$

✓ This is used when ' n ' is small & p is not small.

❖ **Poisson Distribution:**

- Used when probability of success in a small time interval is very small.
  → n is big & p is small
- Poisson Distribution is uni-parametric. (m is the only parameter)
- For Random variable x;
  $x \sim P(m)$
- $P(x = r) = \frac{e^{-m}\, m^r}{r!}$
  for $r = 0, 1, 2, 3, \ldots \infty$
- $P(x = 0) + P(x = 1) + P(x = 2) + \cdots = 1$

✓ Mean $=$ m $=$ np, Variance $=$ m $=$ np, **S.D** $= \sqrt{m}$

✓ Mode $= \begin{cases} [m] \text{ (integral part)} & \text{if m is non} - \text{integer} \\ m \text{ and } m - 1 & \text{if m is integer} \end{cases}$

✓ Additive property

If $x \sim P(m_1)$

$y \sim P(m_2)$

then $x + y \sim P(m_1 + m_2)$

❖ **Normal Distribution:**

- Also known as Gaussian Distribution
- Most important & universally accepted continuous probability Distribution function.
- It is a bi-parametric Distribution where mean (μ) & variance (σ^2) are two parameters which decides the shape normal distribution curve.
  $\mu \Rightarrow$ It will tell central value, $\sigma^2 \Rightarrow$ It decides the spread

✓ Normal curve is a symmetrical curve, skewness $=$ zero

✓ Area under this curve is taken 1.

$\int_{-\infty}^{\infty} f(x) = 1$

✓ $\int_{-\infty}^{\mu} f(x) = 0.5$

$\int_{\mu}^{\infty} f(x) = 0.5$

✓ Both tails of the curve never touches horizontal axis (x-axis).

✓ In Normal Distribution, Mean $=$ Median $=$ Mode

✓ Mean Deviation $= 0.8\sigma$

✓ Quartile deviation $(QD) = 0.675\sigma$

$Q_1 = \mu - 0.675\sigma$

$Q_3 = \mu + 0.675\sigma$

✓ Mean = Median = Mode = $\frac{Q_3 + Q_1}{2}$

✓ **Point of inflexion**; $\mu - \sigma$ & $\mu + \sigma$

✓ **Additive Property:**

If $x \sim N(\mu_1, \sigma_1^2)$ and $y \sim N(\mu_2, \sigma_2^2)$

then $x + y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

✓ **Standard Normal Variate**:

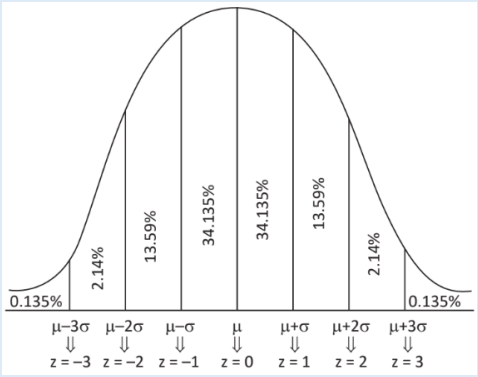
If $\mu = 0$ & $\sigma = 1$

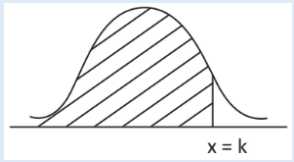
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ then $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

This is known as standard normal variate; $z = \frac{x - \mu}{\sigma}$

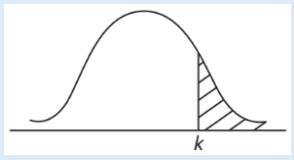
✓ For Standard Normal variate; $\mu = 0$ & $\sigma = 1$

- ✧ Mean = Median = Mode = 0
- ✧ MD = $0.8\sigma = 0.8$
- ✧ $QD = 0.6756 = 0.675$
- ✧ $Q_1 = -0.675$ and $Q_3 = 0.675$
- ✧ Point of Inflexion = -1 & 1



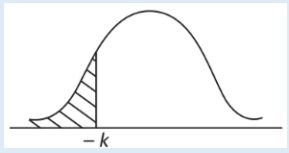
✓ 

$P(x < k) = \phi(k)$

✓ 

$P(x > k) = 1 - P(x < k)$

$= 1 - \phi(k)$

✓ 

$P(x < -k) = 1 - P(x < k)$

$\phi(-k) = 1 - \phi(k)$

CORRELATION AND REGRESSION

Correlation: The statistical tool which studies the relationship between two or more variables. It is denoted by "*r*" or *r*(*x*, *y*).

Ex; The price and demand of commodity are inversely correlated i.e., negatively correlated.

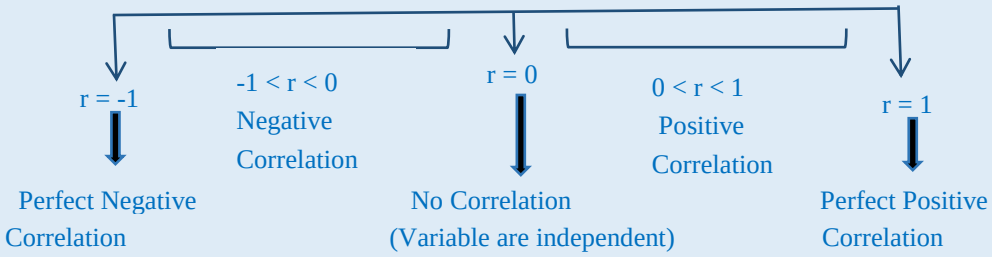
- ✧ Correlation can be negative, positive or zero.
- ✧ $-1 \leq r \leq 1$

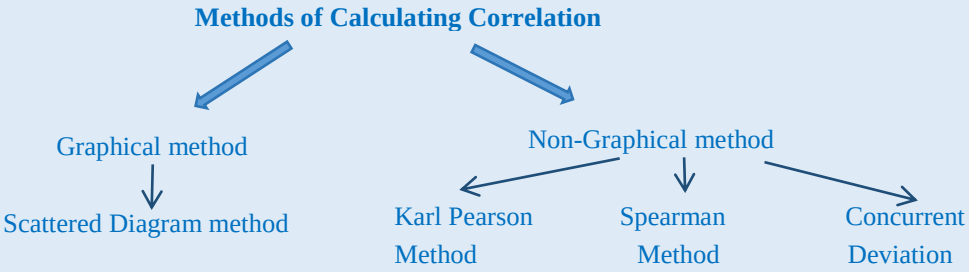
❖ For **Direct/Positive** Correlation between x & y;

$x \uparrow y \uparrow$ or $x \downarrow y \downarrow \Rightarrow r$ is positive

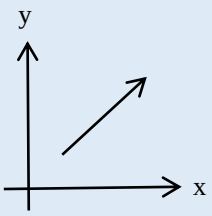
❖ For **Inverse/Negative** Correlation between x & y;

$x \downarrow y \uparrow$ or $x \uparrow y \downarrow \Rightarrow r$ is negative

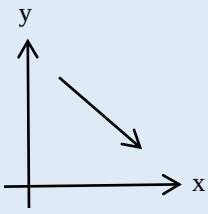




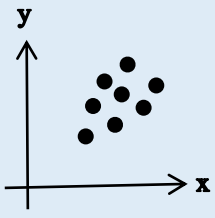
❖ **Scattered Diagram method:**



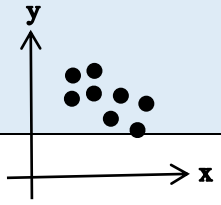
For straight line (Lower left to upper right)
r = 1



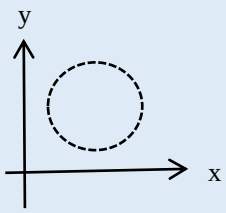
For straight line (Upper left to lower Right)
r = -1



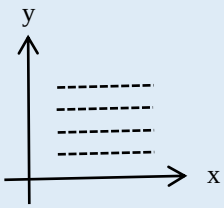
No straight line
0 < r < 1



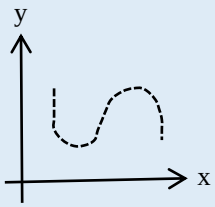
No straight line
-1 < r < 0



$r = 0$



$r = 0$



$r = 0$

❖ **Karl Pearson's Coefficient of Correlation:**

- $r = \frac{cov\cdot(x,y)}{\sigma_x\,\sigma_y}$ where, $cov\,(x,y) = \frac{\sum (x_i - \bar{x})\cdot(y_i - \bar{y})}{N}$
- $r = \frac{\sum(x_i - \bar{x}).(y_i - \bar{y})}{N.\sigma_x\cdot\sigma_y}$
- $r = \frac{\sum(x_i - \bar{x})\cdot(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \times \Sigma(y_i - \bar{y})^2}}$
- $r = \frac{\sum xy - \frac{\sum x \times \sum y}{N}}{\sqrt{\sum x^2 - \frac{(\Sigma x)^2}{N}}\sqrt{\sum y^2 - \frac{(\Sigma y)^2}{N}}}$

→ Cov (x, y) can be any real number (Negative, Positive & zero).
→ Cov (x, y) does not change with origin but it changes with change of scale.
→ Correlation does not change with change of origin

i.e., If $u_i = x_i - A$ & $v_i = y_i - B$ then, $r = \frac{\Sigma uv - \frac{\Sigma u \times \Sigma v}{N}}{\sqrt{\Sigma u^2 - \frac{(\Sigma u)^2}{N}}\sqrt{\Sigma v^2 - \frac{(\Sigma v)^2}{N}}}$

→ Correlation does not change with change of scale (Provided scale is positive).

- When scale is negative, the sign of correlation may change but magnitude does not change. Ex;

✧ If $r\,(x,y) = 0.5$ and $u_i = 2x_i$ & $v_i = 4y_i$ then $r(u_i, v_i) = (+)(+)0.5 = 0.5$

✧ If $r(x,y) = 0.6$ and $u_i = -2x_i$ & $v_i = 3y_i$ then $r(u_i, v_i) = (-)(+)0.6 = -0.6$

❖ **Spearman's Rank correlation:** This method is used for qualitative characters & level of agreements and disagreements between opinions of Judges.

- When no numbers repeat; $r = 1 - \frac{6\Sigma D^2}{N^3 - N}$
- When some numbers repeat; $r = 1 - \frac{6\left[\sum D^2 + \frac{1}{12}(m_1{}^3 - m_1) + \frac{1}{12}(m_2{}^3 - m_2)\right]}{N^3 - N}$

❖ Concurrent Deviation method:

$r = \pm\sqrt{\pm\left(\frac{2c - m}{m}\right)}$ where, c is total number of concurrent deviations and $m = n - 1$

- Coefficient of Determination: $r^2 = \frac{\text{Explained variance}}{\text{Total variance}}$
- Coefficient of Non-Determination: $1 - r^2$

Regression:

There are two Linear Regression Lines:

	When y depends on x	When x depends on y
Standard form	$y = a + bx$	$x = a + by$
How to Find ?	Formula: $y - \bar{y} = b_{yx}(x - \bar{x})$	Formula: $x - \bar{x} = b_{xy}(y - \bar{y})$
Slope of line	Slope $= b_{yx}$	Slope $= \frac{1}{b_{xy}}$

b_{yx} and b_{xy} are known as Regression coefficients.

❖ Calculation of Regression Coefficients:

We know,

$b_{yx} = \frac{\text{cov}(x,y)}{\sigma_x^2}$

$b_{yx} = \frac{\sum(x_i - \bar{x})\cdot(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$

$b_{yx} = \frac{\sum xy - \frac{\Sigma x \Sigma y}{N}}{\sum x^2 - \frac{(\Sigma x)^2}{N}}$

$b_{yx} = r\frac{\sigma_y}{\sigma_x}$

$b_{xy} = \frac{\text{cov}(x,y)}{\sigma_y^2}$

$b_{xy} = \frac{\sum(x_i - \bar{x})\cdot(y_i - \bar{y})}{\sum(y_i - \bar{y})^2}$

$b_{xy} = \frac{\sum xy - \frac{\Sigma x \Sigma y}{N}}{\sum y^2 - \frac{(\Sigma y)^2}{N}}$

$b_{xy} = r\frac{\sigma_x}{\sigma_y}$

- $r = \pm\sqrt{b_{xy} \times b_{yx}}$

  → 'r' will be positive if both b_{xy} and b_{yx} are positive.

  → 'r' will be negative if both b_{xy} and b_{yx} are negative.

  $\Rightarrow b_{xy} \times b_{yx} \leq 1$
- $r \leq \frac{b_{xy} + b_{yx}}{2}$

✧ If Regression line y on x is: $ax + by + c = 0$, then $b_{yx} = -\frac{a}{b}$

✧ If Regression line x on y is: $Ax + By + C = 0$, then $b_{xy} = -\frac{B}{A}$

- Two Regression lines intersect each other at $(\bar{x}, \bar{y})$.
- Regression coefficients does not change with change of origin;

  If $u_i = x_i - A$ & $v_i = y_i - B$, then $b_{yx} = b_{vu}$ & $b_{xy} = b_{uv}$
- Regression coefficients changes with change of scale;

  $u_i = ax_i + b$ & $v_i = cy_i + d$

  $b_{vu} = \frac{c}{a} \times b_{yx}$

  'or' $b_{vu} = \frac{\text{Scale of y}}{\text{Scale of x}} \times b_{yx}$
- When two lines are perpendicular, then $r = 0$

- When two lines are coincident, then $r = 1$

INDEX NUMBER

Ratio of two or more time periods involved, one of which is the base time period. Ex; NSE, BSE etc

Methods of Construction of Index Number:

- **Simple Aggregative method:** $P_{01} = \frac{\sum P_1}{\sum P_0} \times 100$
- **Simple Relative method:**
  - $P_{10}(AM) = \frac{\sum \left(\frac{P_1}{P_0} \times 100\right)}{N}$
  - $P_{01}(GM) = AL\left[\frac{\sum \log\left(\frac{P_1}{P_0} \times 100\right)}{N}\right]$
- **Weighted Aggregative method:** $P_{01} = \frac{\sum P_1 W}{\sum P_0 W}$ where, W is weight
  - **Laspeyre's (Base year Quantity):** $P_{01} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$
  - **Paasche's (Current Quantity):** $P_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100$
  - **Fisher's (Geometric Mean of L & P):**

$$P_{01} = \sqrt{L \times P}$$

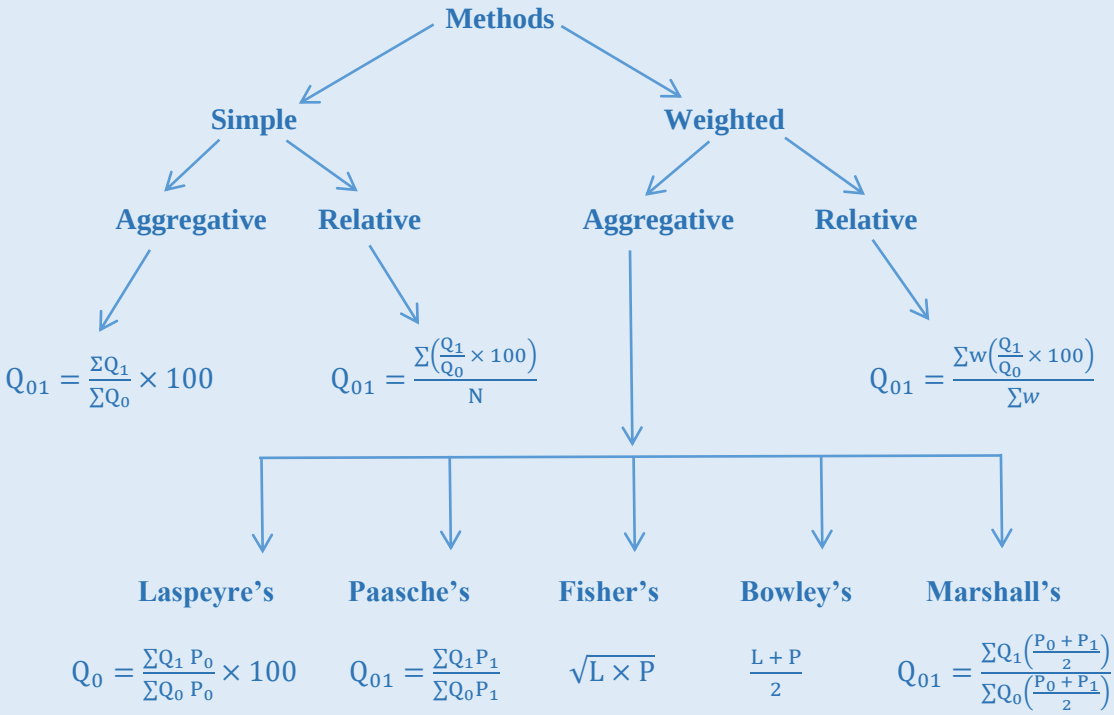
$$P_{01} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100$$

  - **Marshall Edgeworth:** $P_{01} = \frac{\sum p_1\left(\frac{q_0+q_1}{2}\right)}{\sum p_0\left(\frac{q_0+q_1}{2}\right)}$
  - **Dorbish & Bowley:** $P_{01} = \frac{L+P}{2} = \left[\frac{\frac{\sum p_1 q_0}{\sum p_0 q_0} + \frac{\sum p_1 q_1}{\sum p_0 q_1}}{2}\right] \times 2$

Weighted Relative Method:

$$P_{01}\,(AM) = \frac{\sum w_i\left(\frac{P_1}{P_0} \times 100\right)}{\sum w_i} \;\&\; P_{01}(GM) = AL\left[\frac{\sum w_i \log\left(\frac{P_1}{P_0} \times 100\right)}{\sum w_i}\right]$$

Quantity Index (Volume index):



✧ **Value Index:** $V_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_0} \times 100$

✧ **Group Index:** $\frac{\sum \omega_i I_i}{\sum \omega_i}$

❖ **Deflated Value (Real Value):** Deflated value $= \frac{\text{Current value}}{\text{Price Index}}$

❖ **Shifting of Base year:** Shifted Price Index $= \frac{\text{Original Price index}}{\left(\text{Index of the year where it has to be shifted}\right)} \times 100$

Test of Adequacy:

1) **Unit Test:** Formula should be independent of units of price & quantity, Due to charge in units index number should not change.

→ This test is satisfied by all methods except "Simple Aggregative method".

2) **Time Reversal Test:** Formula should work both ways "Forward" & "Backward"

$$\{P_{01} \times P_{10} = 1\}$$

3) Factor Reversal Test:

$$P_{01} \times Q_{01} = V_{01}$$

or

$$P_{01} \times Q_{01} = \frac{\sum P_1 q_1}{\sum P_0 q_0}$$

Satisfied by "fishers"

4) **Circular Test** → Extension of Time reversal

$$P_{01} \times P_{12} \times P_{21} = 1$$

Satisfied by Simple Price Relative (GM) & Weighted Aggregative (Fixed Weight)