* two types of tabulation
* A table has 5 parts

Intro of Statistics

Stats
- singular ⇒ data collected, analysed, presented
- plural ⇒ Just data < Qualitative / Quantitative

lotus — Latin word ⇒ 'status'
Pasta — Italian word ⇒ 'Statista'
Swastik — German word ⇒ ~~'Sta~~ 'statistik'
Barbeque — French word ⇒ 'statistique'

Application
- Eco
- Busin mgmt
- Industry
- Commerce

Limitations:-

I always deals with aggregate
II generally concerned quantitative
III refer module 13.3 (in general future prediction are unlikely to be ... nature)
IV built upon random sampling.

→ data → quantitative info. abt some particular character

Collection of data

data
- Primary → collected for first time by ...
- Secondary → data already collected and then used by diff ...

→ e.g. gender, nationality.

Qualitative characteristics ⇒ attributes
Quantitative — " — ⇒ variables
Money — measureable quantity

- discrete ⇒ personal
- continuous ⇒ Industrial, commercial

Collection of Primary data ⇒ 4 methods
- interview method
- mailed questionnaire method
- observation method
- questionnaire filled and sent by enumerators

* i) Interview Method
  - a) personal interview method
  - b) Indirect — " —
  - c) Telephone — " —

a) Personal — " — ⇒ Best method for natural calamity
⇒ High accuracy, low coverage
A↑, C↓
⇒ meets respondent directly, high respon...

b) Indirect — " —
⇒ cannot meet respondent directly but can connect or meet associated person
⇒ ↑A, ↓C
(here moderate accuracy)

c) Telephone — " — ⇒ max. number of non-respo...
⇒ A↓, C↑, ↓consistent
⇒ quickest method, non-expensive met...

* ii) Mailed Questionnaire method. (highest n... re...)
make a form of ques. and send (in form of mail)
⇒ Response↓, coverage↑, Accuracy↓
⇒ collect widest area

* iii) observation method ⇒ best method
⇒ most accurate method, ↓C ⇒ time

* iv) Questionnaire filled and sent by enumerators.
→ ↓R ↑C

Sources of secondary data
- International sources
- Govt sources
- Pvt & Quasi Govt sources
- Unpublished sources

Density = Population / Area

Scrutiny of data
⇒ to check the data we have collec[ted]
accurate and consistent

classification/org data →

Types of classification or organisation ⇒ 4

→ statistical analysis is concerned wid/dependent

i) Chronological or Temporal or Time Series
ii) Geographical or Spatial series data
iii) Qualitative or ordinal
iv) Quantitative or cardinal.

data → frequency (Qualitative, Quantitative), non-frequency (Time series, Spatial series)

Diagramatic representatⁿ data
⇒ can be used by both educated and uneducated
⇒ less accuracy compared to tabula[r]
⇒ can detect hidden trend.

Line diagram (histogram) ⇒ one dimensional.
Bar diagram ⇒ one dimensional.
(Pie diagram/circle) Pie chart ⇒ two — " —
cube chart ⇒ three — " —

Line diagram ⇒ logarithmic or ratio chart

time series exhibit a wide of fluctuation → ratio chart, logarithmic chart

Types of bar diagram
- vertical: data → time series related, qty
- Horizontal: data → qlty, geographically, spatial

Multiple/Grouped bar diagrams → compare related series

when data is categorial ⇒ pie diagram

Modes of presentation of data
- Textual presentation
- Tabular Presentatⁿ
- Diagrammatic Representation

Quickest method

statistic

caption ↓ upper part of table describes column

al sources
ces
s: Govt. sources
~~ed sources~~

PAGE NO.:
DATE: / /

we have collected
istent
g data →

process of arranging data on basis of characteristic under consideratn into no. of grps or classes ac... similarities

ion ⟹ 4

uid / depend
Time
es d
ata

quicke
meth
statist
** →

...resentation of data

caption
↓
upper part of table describes column/su...

* footnote
→ bottom p...

variable

Discrete	continuous
• finite	• assume any value from given interval.
• countable ∞ no.	• eg height weight, sale, profit.
• eg. no. of petals in flower, no. of misprints of a book contains, no. of road accidents in particular locality	

gram ⇒ one d...

...art ⇒ one ...

...rt ⇒ two

⇒ three

arithmetic

...pes of

...compared to tabulation

hidden ...

educated and uneducated.

data → qlty, geographically, spatial

bar diagram

...ation of data

caption

upper p... of table describe column/su...

* footnote → bottom part of table.

→ table show...

→ when data ... or interv...

upper part
of table
describes
column/sub-co

* footnote
→ bottom part
of table.

→ table should be

→ when data colle
or interval is
Temporal

→

Data

Tabulation

variable

Discrete
↓
• finite
• countable ∞ no.
• eg. no. of petals
in flower. no. of
sprints of a
ok contains,
of road accidents
particular locality

Continuous
↓
• assume
any value
from given
interval.
• eg height
weight, sale
profit.

Data
* qlty
geographically
Spatial

} related

grams

→ can be used by

⇒ less accuracy comp
⇒ can detect hidden tr

histogram
line diagram ⇒ one dimen
bar diagram ⇒ one dimen
pie diagram / pie chart ⇒ two
circle
cube chart ⇒ three

line diagram ⇒ logarithmic or ratio
time series exhibit
a wide fluctuation
ratio chart
logarithmic chart

Types of bar diagram
- vertical
  - data
    - time series
    - qty
  - related
- Horizontal
  - data
    - qlty
    - geographically
    - spatial

multiple / Grouped bar diagrams → comp. relat. seri.

when data is categorical ⇒ pie diagram

Modes of presentation of data

Textual presentation
- → Paragraph
- → merits
  - • simplicity
  - • layman can even present
  - • obs wid exact magnitude
  - → 1st step towards other presentation
- → demerits
  - • dull
  - • monotonus
  - • not preffered
  - • comparison not possible.

Tabular Presentatn or Tabulation
- → systematic presentatn wid rows & columns.
- → (see aside)
- → data must be arranged in such way that comparison
- → presentation appealing
- → complicated data can be presented
- → must for digramatic representation
- → high accuracy

Diagrammatic Representation
- → attractive
- → helpful for both
  - educated
  - uneducated
- → hidden trend ascertained
- → compared to tabulation less accurate.

→ Data coll
data
→ Data
→ Data

quickest → Telephone
method   interview

→ statistical data should be alloted a serial number along wid self explanatory title

** → Tabular representation

Caption
upper part of table describes column/sub-column

Boxhead
entire upper part of table includes column/sub-column and No., units of measurement

stub ~~caption~~
left part of table providing descriptn caption of rows

Body
main part of table containing numerical figures

* footnote
→ bottom part of table.

→ table should be well balanced in length and breadth

→ when data collected in respect of successive time points or interval is called chronological/Time series/Temporal data

→ Data collected region wise Geographical/spatial data

→ Data arranged in respect of attribute → qualitative/ordinal

→ Data ________ of variable → Quantitative or cardinal

Frequency Distribution → tabular representation of statistical data

discrete r.v. ⇒ qualitative characteristic
- discrete
- ungrouped
- simple frequency distri.

continuous variable ⇒ Grouped frequency distri.

Frequency distribution of a variable

I. Range = L − S {for continu variables}
II. Range = No. of class Interval × class length
I. • Tally Mark ⇒ class interval mostly

Class limits
⇒ mini & maxi value of class interval
LCL ⇓ UCL ⇓

continuous ⇒ CB
Non continuous ⇒ CL

Class boundary
⇒ actual class limit of class interval.

continuous variable ⇒ overlapping or mutually exclusive
⇒ class boundaries coincide class limits

Discrete variable ⇒ non overlapping or mutually inclusive
⇒ includes both limit

mid-point / mid value / class mark $= \frac{LCL+UCL}{2}$ or $\frac{LCB+UCB}{2}$

Width / size of class interval
⇒ only boundaries = UCB − LCB

Cumulative frequency
- less than cf
- more than cf

discrete ungrouped — continuous grouped

Frequency Density of class interval
$\frac{\text{frequency}}{\text{class length}}$

Relative frequency and percentage frequency of class interval

add up to unity ⇒ Relative frequency = $\frac{\text{class frequency}}{\text{Total frequency}}$

add up to one hundred ⇒ percentage frequency = $\frac{\text{class frequency}}{\text{Total frequency}} \times 100$ (i.e. expressed in percentage)

Graphical Representation of frequency distribution.

i) Histogram / Area diagram (CB, frequency)
- most convenient
- helps finding mode
- in this CL to be converted into CB and series of adjacent rectangle

• class length can be ⟨ equal (=) / unequal (≠)

class length is usually = but in
of ~~for~~ some condition frequen
directly proportnal to area of

discrete → class mark, frequency
ii) Frequency polygon (line diagram)
continuous
=) usually for single freq. distr
=) can also be used for grouped fr
distri only when width of class int
remains same
=> frequency curve => limiting form of frequency polygo
=> method providing frequency to a
→ (CB, CF)

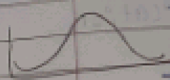
iii) Ogives or cumulative frequency Gra
· To plot cf on graph => ogive
Short term ⟸ ogives => median/quartile/
projection partition

Starts ⟸ less than | more than => starts from
from '0' | type | highest
zero | |
+ type | |
⇓ | ⇓ |
with help of | with help of |
less than cf | more than cf |
on vertical | on vertical |
axis | axis |

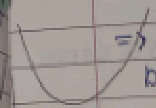
iv) frequency curve .. (smooth, freehand curve
=> smooth curve for which total area
taken to be unity
=> limiting form of histogram/freq polyg
→ can be obtained by drawing smooth & free cu

Frequency curve ⟨ Bell shaped curve / U-shaped curve / J-shaped curve / Mixed curve

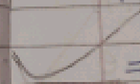
less than => 'S' shaped
ogive



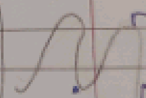
• Bell Shaped curve
=> most common
=> height, weight, mark, profit etc
=> extreme ⇓, center ⇑



• U- Shaped curve
=) exetreme ⇑, center ⇓
=) distributn of Kolkata bound commuters belongs to this type.



• J - shaped curve
=> starts with mini frequency & gradually reaches its max. frequeny at other extremity.
=> distributn of commuters coming to Kolkata from early morning hr to peak morning hour.



Mixed curve
=> combination of all curves

Exclusive series is continuous series

classical def^n of probability → assumption
$\boxed{\text{O/C} \rightarrow \text{mutually exclusive, exhaustive and equally likely.}}$

- Initially probability was branch of mathematics
probability
Theories of testing hypothesis and Estimation are based on Probability

Probability
- subjective
  - → personal judgement
  - → may be biased.
- objective
  - → mathematical operation

Events → results or outcomes of random exp.
experiment → performance that produces certain Result.

- simple [Elementory]
  - ek coin toss
  - cannot be decomposed into further
  - (single)
- composite [compound)
  - → can be splited
  - ek sath 2 coin toss
  - kiye → can be decomposed in 2 or more events
  - [combination)

A simple outcome can be part of multiple outcome.

* Random experiment → result of experiment depend on chance only.
  e.g. tossing a coin

Mutually Exclusive Events or Incompatible
→ occurence of one implies non occurence of other event
→ one event at a time
→ Independent
→ $A \cap B = 0 \Rightarrow$ mutually exclusive.
$P(A \cup B) = P(A) + P(B) \Rightarrow$ mutually exclusive
→ Two events cannot occur simultaneously.

Exhaustive Event
→ atleast one event must occur
→ add all probability then the ans must be ~~to~~ full given set → to find Exhaustive end.
→ $P(A \cup B) = 1$
→ union of both events should give sample space.

Classical probability → Bernoulli & Laplace
└→ priori defⁿ

→ limitations → total no. of events is finite
└→ events are equally likely / equi probable
└→ ited field of application

Equally likely Events OR Mutually symmetric Events OR Equi-Probable event
→ events to occur has equal probability
→ no event is expected to ~~occur~~ occur more frequently.

- $0 \le P(A) \le 1$
- * Probability is also called as chance

$P(A) = 1$ Sure event
$P(A) = 0$ = Immpossible event

$P(A) + P(A') = 1$ $A' / A^c / \bar{A} \Rightarrow A$ compliment
$P(A')' = A$

favourable o/c
No. of unfavourable o/c

odd in favour of event A $= m_A : (m - m_A)$ $= \frac{m_A}{m - m_A}$

unfav / favourable — odd against favour of event A $= \frac{m - m_A}{m_A}$

No. of trial

(No. of outcome in a trial)

$(2)^2 = 4$

* Sample space (S) or (Ω)
- Sum of Probability of event in Sample space = 1.
* $P(A) \ge 0$ for every $A \subseteq S$ (subset)
- Sum of all probabilities of mutually exclusive and exhaustive events = 1
- $P(A) = \frac{n(A)}{n(S)}$

$\boxed{P(B/A)}$ → Probability that event B has occured on Assumption that A has happened

→ Probability of occurrence of atleast one of the events A and B is denoted by → $\boxed{P(A+B)}$

→ Probability of occurence of A as well as B is denoted by → $\boxed{P(AB)}$

$\boxed{P(A) + P(A)^c = 1}$

For $P(AB) = P(A) \cdot P(B)$
→ Two events must be independent

A and B indep.
→ ∴ A^c and B^c indep.

If A and B indep

$\boxed{P(B/A) = P(B)}$

p.m.f = $\boxed{20}$
↓
discrete

p.d.f → continuous

→ Total no. of outcome $= (P)^n$ — n → no. of trials, P → No. of events in one try

* Demorgans law

$P(A \cup B)' = P(A' \cap B') \Rightarrow 1 - P(A \cup B)$

$P(A \cap B)' = P(A' \cup B') \Rightarrow 1 - P(A \cap B)$

$P(A \cap B') = P(A) - P(A \cap B)$

$P(B \cap A') = P(B) - P(A \cap B)$

→ union of probability
$P(A \cup B)$ → occurence of atleast one event
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) +$
$P(A \cap B \cap C)$

→ intersection
$P(A \cap B)$ → simultaneous occurence of A and B
$P(A \cap B) = P(A) + P(B) - P(A \cup B)$

→ $A - B$ → set A but not B
i.e $A - B = \{x : x \in A \text{ and } x \notin B\}$

→ $B - A$ → set B but not A
i.e $B - A = \{x : x \in B$ and ~~not~~ $x \notin A\}$.

* Events mutually exclusive → $P(A \cap B) = 0$
$\therefore P(A \cup B) = P(A) + P(B)$
$P(A \cap B \cap C) = 0$ → $P(A \cup B \cup C) = P(A) + P(B) + P(C)$

* Two events exhaustive
$P(A \cup B) = 1$
$P(A \cup B \cup C) = 1$

* events equally likely
$P(A) = P(B) = P(C)$.

Addition Thm.

Thm 1 $P(A \cup B) \rightarrow P(A+B) = P(A) + P(B)$

Thm 2 → extension of Thm 1.

Thm 3 $P(A \cup B) \rightarrow P(A) + P(B) - P(A \cap B)$

ne event or bo

$-P(B\cap C)-P(A\cap$

of A and B

$)+P(C)$

— Probability of occurence of two event A and B Simultaneously → Compound Probability / Joint Probability
→ $P(A\cap B)$

$P(A/B) \rightarrow \frac{P(A\cap B)}{P(B)}$ → Probability of A when event B has already occured.
$P(B)>0$

$P(B/A) \Rightarrow \frac{P(A\cap B)}{P(A)}$ $P(A)>0$

If events A, B are independent

$P(A\cap B) = P(A)\times P(B)$

$P(A\cap B\cap C) = P(A)\cdot P(B)\times P(C)$.

If A, B → indep. then $\left\{\begin{array}{l} B' \text{ and } A' \\ A \text{ and } B' \\ A' \text{ and } B' \end{array}\right\}$ indep.

$P(A\cap B) = P(A)\times P(B/A)$.

Expected value of random variable

$\mu = E(x) = \Sigma p_i x_i$

$E(x^2) = \Sigma p_i x_i^2$

$V(x) = E(x^2) - \mu^2$

Theoretical Distributions

Theoretical Distribution

- Discrete (pmf)
  - Binomial
  - Poisson
- $p+q=1$
- Continuous (pdf)
  - Normal

→ probability distribution existing only in Theory
→ To make short term projections
→ statistical analysis is possible only on basis of Theoretical probability distribution.
→ probability distribution posses all characteristics of an observed distribution

Binomial Distribution → pmf.
→ n is large
(n, p) , $X \sim B(n, p)$ → only 2 o/c < success / failure

→ derived from Bernoulli process
→ trial is neither certain nor impossible

* characteristics of Bernoulli.
  i) Each trial is associated with two mutually exclusive and exhaustive. the occurence one is success and non-occurence failure
  ii) indep. trails
  iii) $p+q=1$ ↳ probability
  iv) trials are finite +ve integer.

$$f(x) = {}^{n}C_{x} \cdot p^{x} \cdot q^{n-x}$$
$$n > 0 \quad , \; p, q \geq 0$$
$$\Sigma f(x) = 1$$

ons

uous (p.d.f)

al.

Theory.

basis of

eristics

xclusive

and

→ Biparametric distribution

→ $\mu = np$ (mean)

→ Binomial Distrib. < unimodal / Bi-modal

→ μ_0 → largest integer
$(n+1)p$ (non integer)
$(n+1)p$ & $(n+1)p-1$ (integer)

mode	
$(n+1)p$	= integer (bimodal)
$(n+1)p$	→ non integer unimodal

→ variance → npq

- npq numerically less than / = 1 $\quad {}^nC_r = \frac{n!}{r!(n-r)!}$
  $\therefore npq < np$
- variance < mean

- variance 'X' attains max value $p=q=0.5$ & max value is $\frac{n}{4}$

- $X \sim B(n_1, p)$
  $Y \sim B(n_2, p)$
  $(X+Y) \sim B(n_1+n_2, p)$ only when p is same.

Poisson Distribution.

→ introduced in 1837.

→ $\boxed{\begin{matrix} p \to 0 \\ n \to \infty \end{matrix}}$

→ $m = n$, $p = \lambda$ → integer

→ probability of having more than one success in this time interval is very low

$$\frac{e^{-m} \cdot m^x}{x!}$$

→ $e = 2.71828$

→ uniparametric distribution wid only one parameter 'm'

→ mean → $m = \mu$

→ variance → $\sigma^2 = m$

→ Poison

Biomodal   unimodal

μ_0 = largest interger m (non integer)
= m and m-1

* n↑ (v.large) p↓ → $m = np$
then $B(n,p) \cong P(m)$

→ $X \sim P(m_1)$
$Y \sim P(m_2)$
$X+Y \sim P(m_1+m_2)$

→ Poisson distribution is limiting form of Binomial Distribution.

e.g. (1) distri. of no. of printing mistake per pge of large book
(2) —— 11 —— road accidents on busy road per minute
(3) —— 11 —— radio active element per min. in a fusion process
(4) —— 11 —— demands per min. for health centre & so on.

→ method of moments can be applied to fit a poisson distribution to an abs frequency distribution.

Normal / Gau

→ continuou

→ height,

→ p.d.f

→ most im

→ Karl Gau

→ Biparame

→ parame

$X \sim$

$f(x)$

imp

→ origin 20

→ $y = f(x)$

→ bell shape

→ curve do no touch x-a

→ mean =

→ Area = (u

→ unique m

→ 2 = parts distrib

→ symme

→ B.D is a

→ mean a
1st Qu
3rd Q

→ Norma to con

Normal / Gaussian Distribution

→ continuous Random variable

→ height, weight

→ p.d.f $\int_{-\infty}^{+\infty} f(x) = 1$

→ most imp & universally accepted

→ Karl Gauss → Gaussian distri.

⇒ Biparametric

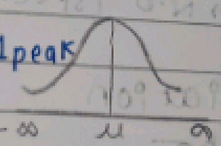
⇒ parameters μ and σ^2

$X \sim N(\underset{\text{mean}}{\mu}, \sigma^2)$

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

imp

- origin 200 years back
- $y = f(x)$
- bell shaped, 1 peak
- curve do not touch x-axis



$-\infty$ to μ = μ to ∞ = 0.5

- mean = mode = median, skewness → zero
- Area = (unity) 1
- unique mode
- 2 = parts, curve → symmetric and corresponding distribution → symmetrical
- symmetric $x = \mu$
- B.D is also symmetric abt $P = 0.5$
- mean deviation of ND $\sigma\sqrt{\frac{2}{\pi}} \cong 0.8\sigma$

1st Quartile $Q_1 = \mu - 0.675\sigma$

3rd Quartile $Q_3 = \mu + 0.675\sigma$

∴ Quartile Deviation → 0.675σ

→ Normal curve changes its curvature from concave to convex and from convex to concave.

$P(\mu - \sigma < x < \mu + \sigma) = 0.68128$
$P(\mu - 2\sigma < x < \mu + 2\sigma) = 0.9546$
$P(\mu - 3\sigma < x < \mu + 3\sigma) = 0.9973$

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

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

* No. of trials for BD ↑, P is moderate we tend to use N.D.

* P.D. for m↑ → N.D.

* t, chi-square, F → N.D (specific conditions)

Quartile Deviation
$QD = 0.675\sigma$
mean Deviation
$MD = 0.8\sigma$

* QD : MD : SD = 10 : 12 : 15
QD < MD < SD

* Standard normal Distribution
→ cont. r.v   z follows ~~...~~ std N.D.

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

- z mean = mode = median = 0
- SD of z 1
- std N.D = 0
- pts of inflexion −1 and 1

→ Symmetrical $z = 0$

→ If $\bar{x}$ denotes AM of random sample of size n drawn from normal population then

$$z = \frac{\sqrt{n}\,(\bar{x} - \mu)}{\sigma} \sim N$$

when to use

→ when probability calculation outside the popular interval.

- Results of ODI matches → Binomial distribution
- wages of ~~[illegible]~~ workers of factory → Normal distrib.
- parameter is characteristics of population.

for any $\sqrt[4]{\ }$: $\sqrt{\ }$ 2 times
$\sqrt{\ }$ for eg $8^{1/4}$ $\sqrt{\sqrt{3}}$

$\boxed{(x)^{1/n} = ?}$

→ $\sqrt{\ }$ 12 times
→ −1
→ × n
→ +1
→ $\boxed{x = x}$ 12 times.

for reciprocal of any no just do ÷ =

for sq of any no. x =
cube x = = and so on

call trick for log
But base 10

1.) Type no. & press $\sqrt{\ }$ 15 times
2.) subtract 1
3.) Divide resulting no. by
"0.000070274"
OR
Multiply by
"14230"

$\log 2 = 0.3010$ | $a^0 = 1$
$\log 3 = 0.4771$

$\log_b a$ → $a > 0$
$b \neq 1$
$b > 0$

→ $\log_a (mn) = \log_a m + \log_a n$
→ $\log_a \left(\frac{m}{n}\right) = \log_a m - \log_a n$
→ $\log_a m^n = n \log_a m$
→ $\log_{a^b} m^n = \frac{n}{b} \log_a m$
→ $\log_a a = 1$
→ $\log a^1 = 0$
→ $\log_b a = \frac{1}{\log_a b}$
→ $\log_b a = \frac{\log_m a}{\log_m b}$