statistics

Statistics is the branch of mathematics where we collect, organize, analyse and represent the data for batter decision making

Types of statistics

- Descriptive stats
- Inferential stats

① Descriptive stat.

Descriptive Statistics is a summary that describes or summarizes the collection of information/data. It summarizes the sample data rather than learning about the population that sample data is representing.

② Inferential stats.

Probability, Hypothesis testing, Z-test, t-test, chi-squeue test, Anova test

Descriptive stat.

① measure of center tendency

(a) mean
(b) median
(c) mode

(a) mean =

Dataset = [2, 6, 5, 3, 1, 4]

$$\text{mean} = \frac{2+6+5+3+1+4}{6}$$

$$= 3.5$$

mean of population Data = μ

mean of sample Data = $\bar{x}$

(b) median :-

Dataset = [2, 6, 5, 1, 3, 4]

median = Data sequance

= [1, 2, 3, 4, 5, 6]

median for even no = $\frac{3+4}{2} = 3.5$

[1, 2, 3, 4, 5]

median for odd no = 3

(c) mode :-

Data = [2, 5, 1, 3, 4, 6] x

Data = [2, 5, 1, 3, 3, 4, 6]

mode = 3

Data = [2, 2, 1, 1, 3, 4, 5, 6]

mode = 2, 1

② Measure of Dispersion

ⓐ Variance
ⓑ standard Deviation
ⓒ Range

ⓐ Variance :- It is define as spread of the data from its center value.

population variance

$$\sigma^2 = \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{N}$$

N = population data size

x_i = no. of data point

μ = mean of population

Sample variance

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

$\bar{x}$ = mean of sample

$n-1$ = Sample Datasize

$(n-1)$ is the degree of freedom

(b) Standard Deviation :-

How far the data point from mean.

Population SD

$$\sigma = \sqrt{\sum_{i=1}^{N} \frac{(x_i - \mu)^2}{N}}$$

Sample SD

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

© Range

Range = max value − mini value

Dataset = [1, 7, 17, 19, 23, 25, 28, 44, 45, 50]

Range = 50 − 1

= 49

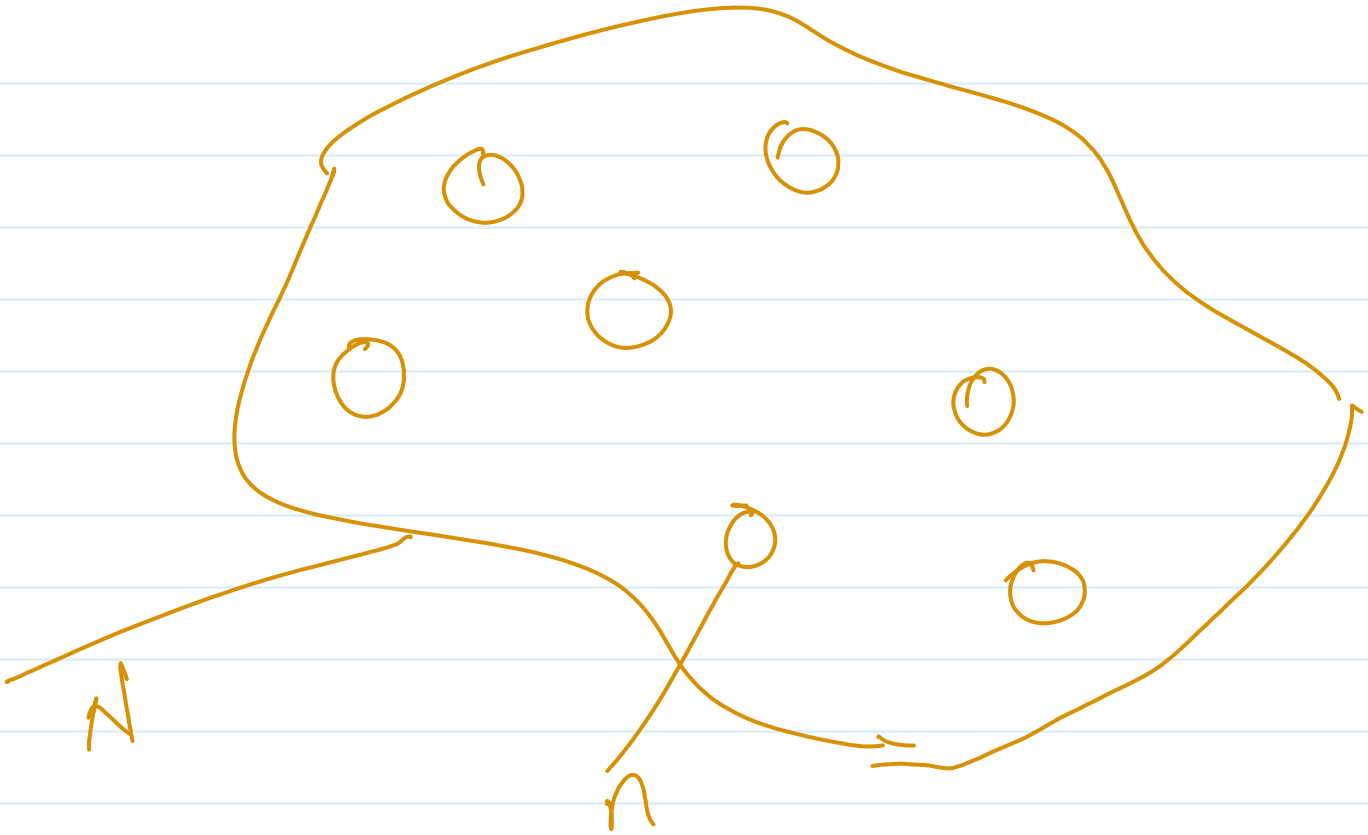
✡ Population & Sample ✡

① Population :- Entire data available for analysis.

Population = N

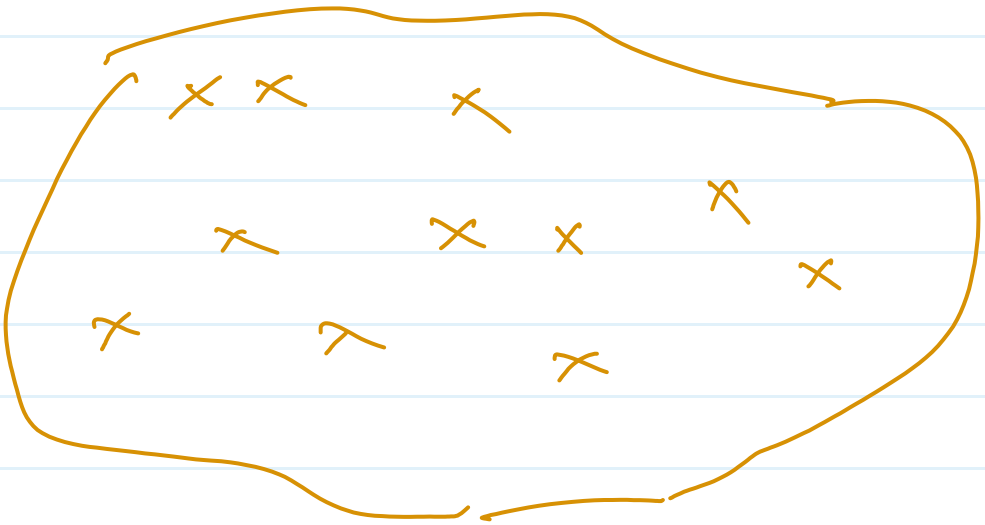
② Sample :- Small chunk, or few datapoint from the popula.

Sample = n

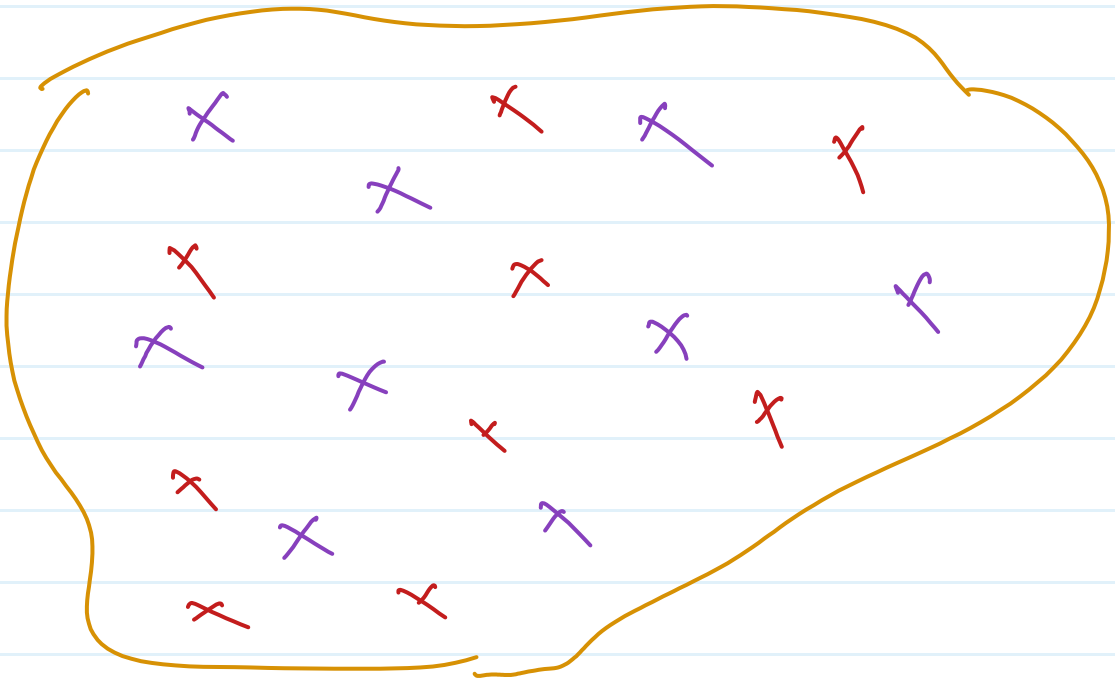


Sampling methods

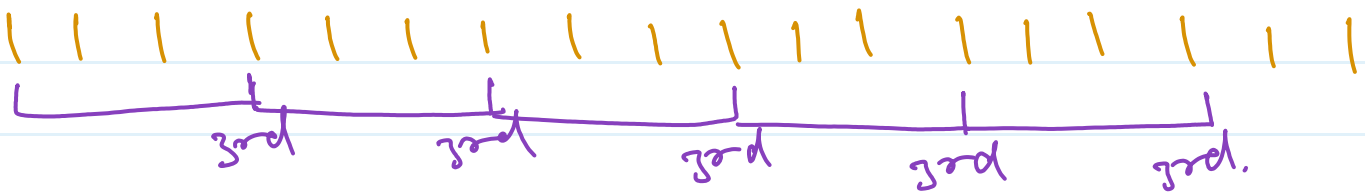
① Simple random sampling



(b) stratified sampling

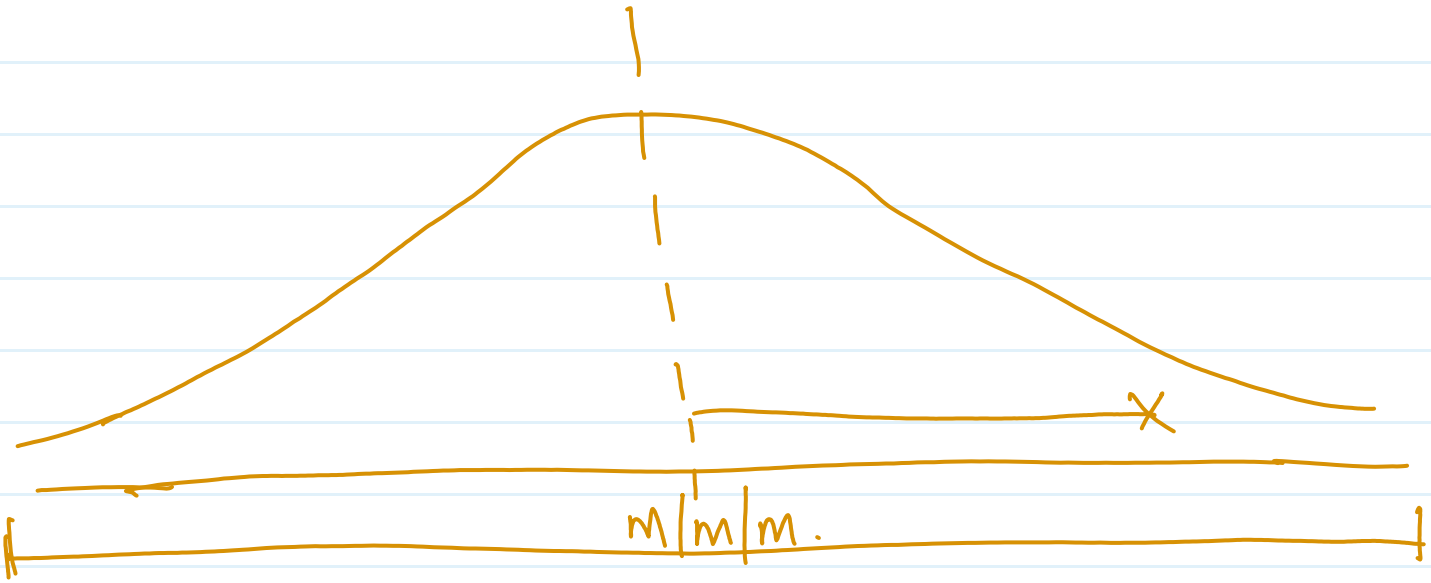


(c) Systematic Sampling



(d) Convinience sampling

with convinience.



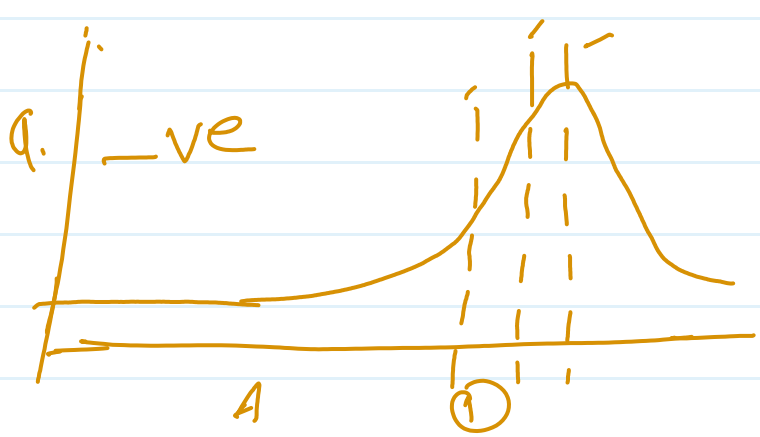
③ measure of shape :-

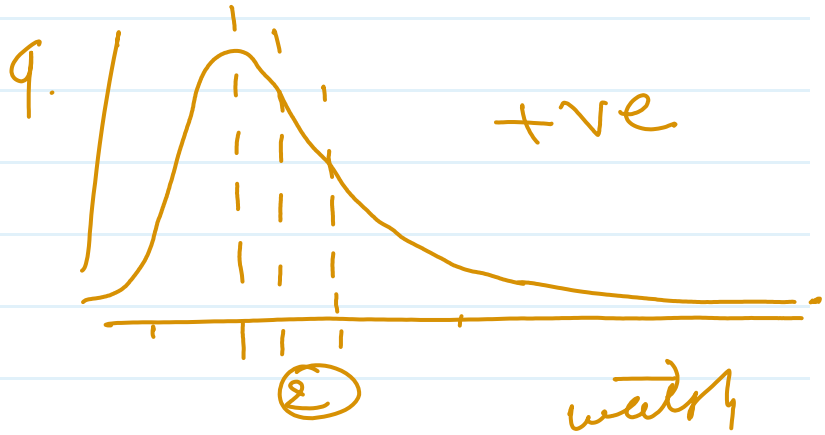
① skewness

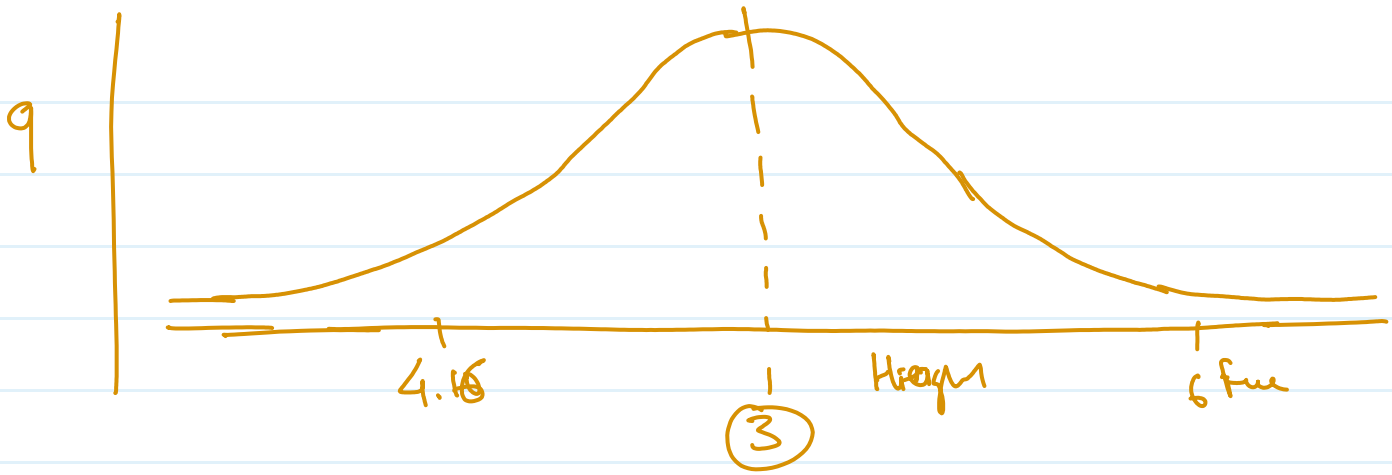
② kurtosis

③ Box and whisker plot

① skewness







① Negative Dist. / Left skewed Dist.

mode > median > mean

② Positive Dist / Right skewed Dist

log normal Dist

mean > median > mode

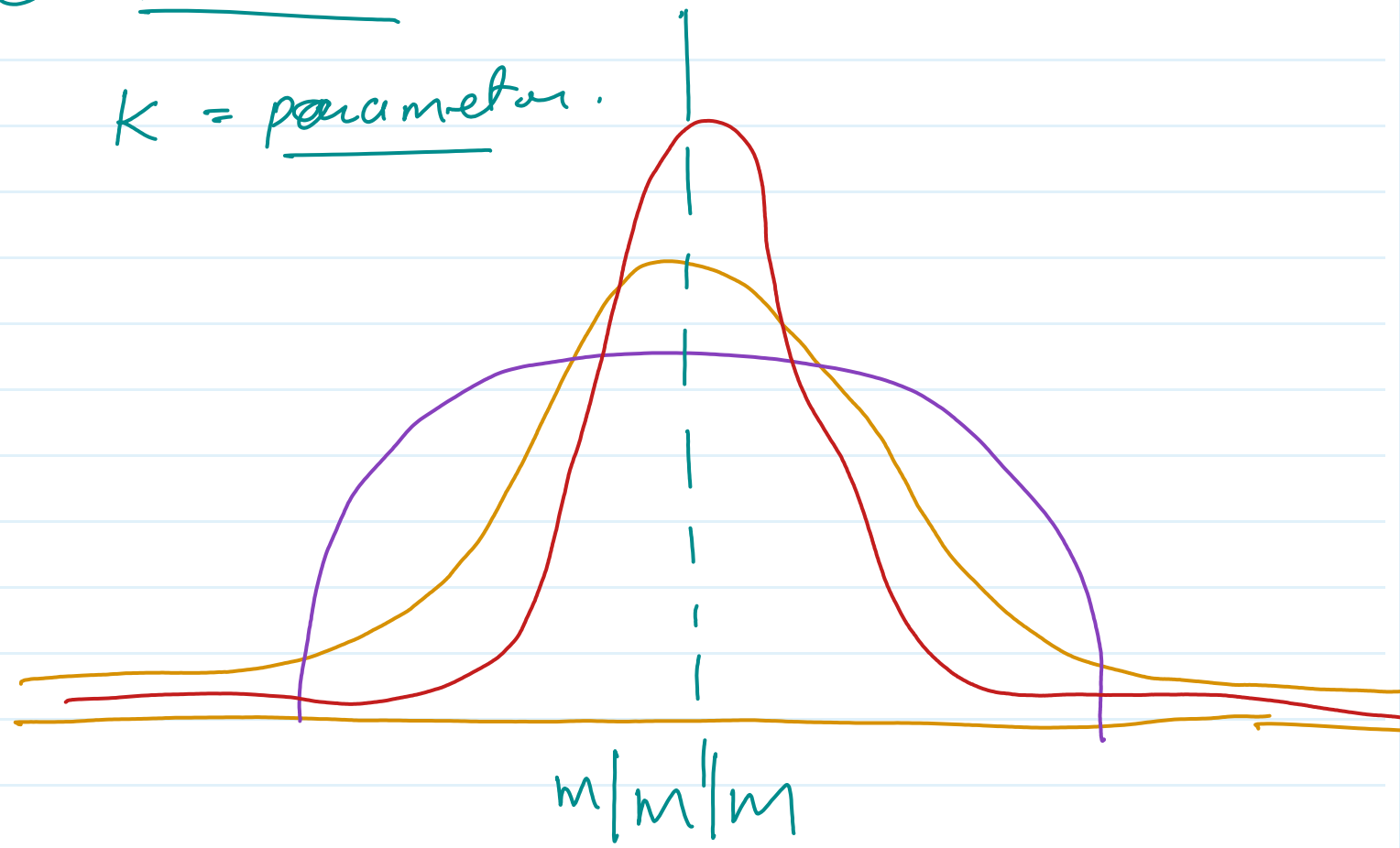
③ Normal Dist. / Gaussian Dist/

Bell curve Dist.

mean = median = mode

② kurtosis

K = parameter.



K = 3

① mesokurtosis

K = 3 =

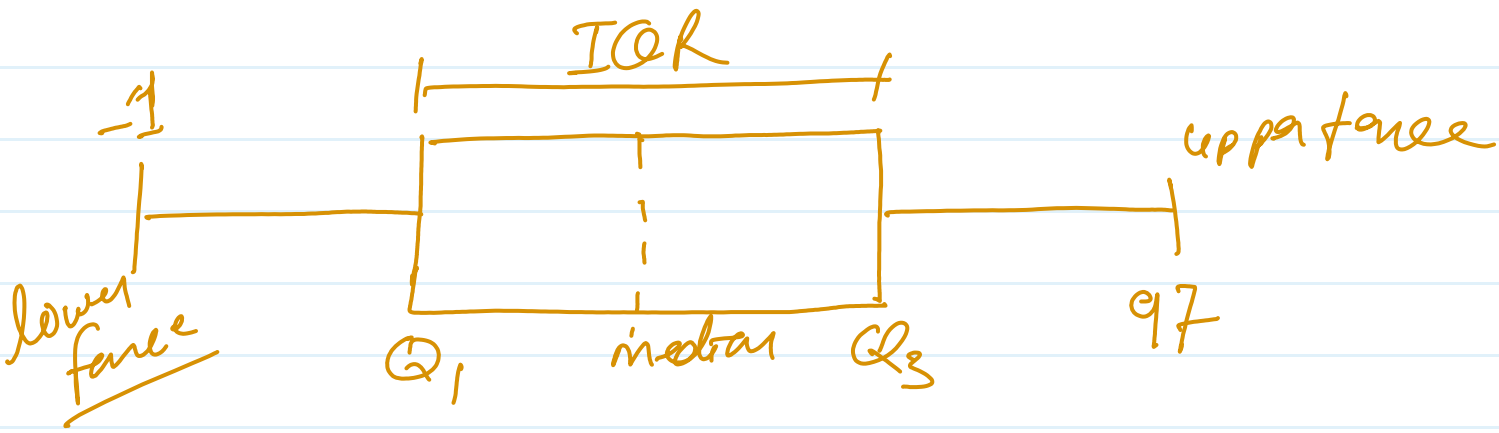
② Leptokurtosis

K > 3 =

③ platikurtosis

K < 3 =

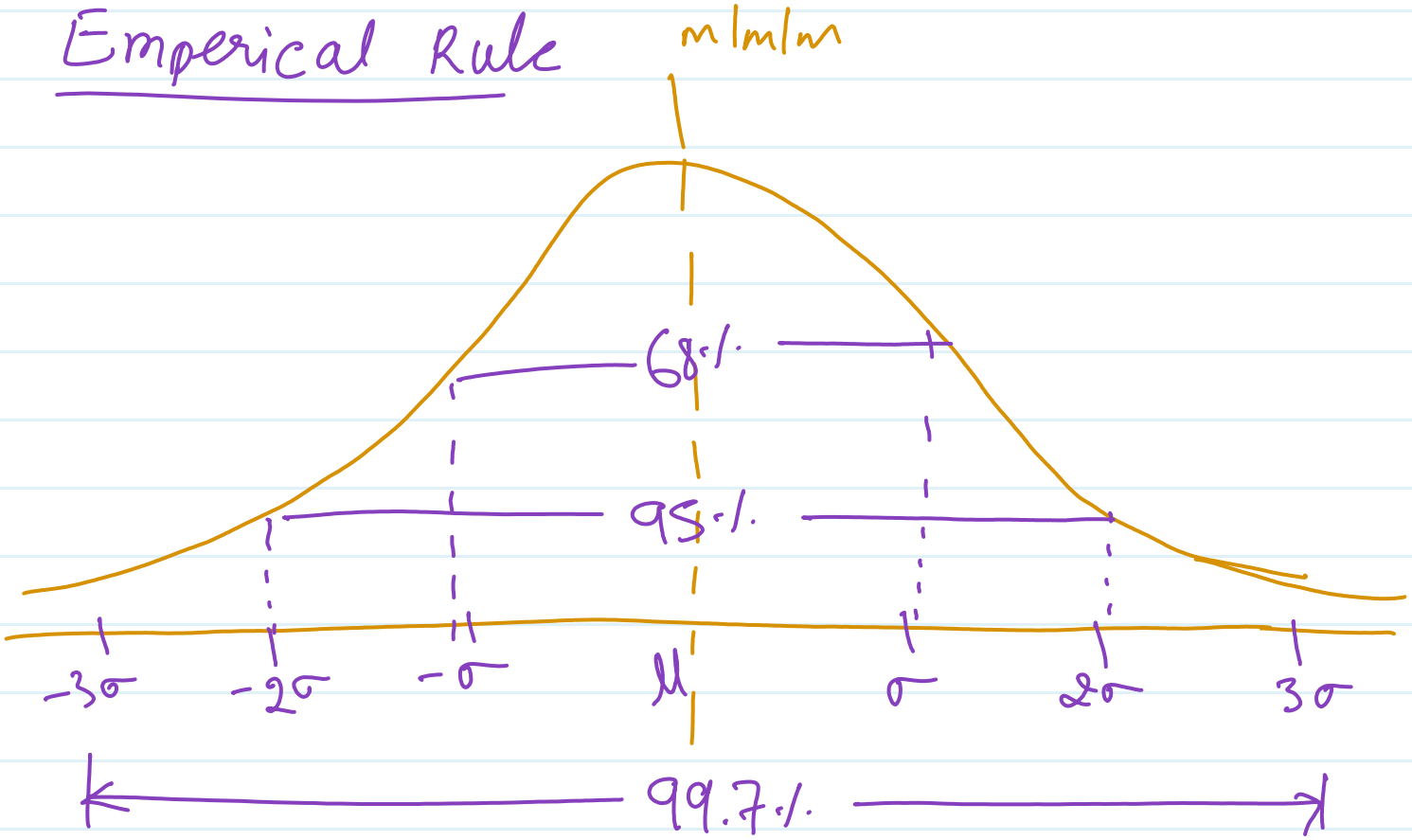
③ Box and whisker plot



Data: [-2, -2.5, 2, 0, 3, 40, 45, 67,
90, 101, 100, 35, 87, 121, -3]

MOJ

Emperical Rule



Emperical Rule / formula

68 - 95 - 99.7

measure of position

1. Quartile
2. percentile
3. Inter quartile range
4. 5 number summary

(1) Quartile

Q_1 — 25%

Q_2 - 50%

Q_3 - 75%

Q_4 = 100

Eg :- 10, 20, 30, 40, 50, 55, 60, 70, 80, 90, 100

min — Q_1 — median Q_2 — Q_3 — max

② percentile

eg:- [2, 2, 3, 4, 5, 5, 5, 6, 7, 8, 8, 8, 8, 9, 9, 10, 11, 11, 12]

percentile of 10th position

$$\text{Percentile} = \frac{16}{19} \times 100$$

$$= 84.2\ \%$$

6th position

$$\Rightarrow \frac{8}{19} \times 100 \Rightarrow 42.1\ \%$$

②. ② Percentile Rank

$$= \frac{\text{Percentile}}{100} \times (n+1)$$

eg:- what value exist at percentile ranking of 25%.

$$\Rightarrow \frac{25}{100} \times (19+1)$$

$$\Rightarrow \frac{\cancel{25}\,5}{\cancel{100}} \times \cancel{20} \Rightarrow 5^{\text{index}}$$

$$\Rightarrow 5^{\text{th}} \text{ number}$$

Eg:- 65%

$$\Rightarrow \frac{\cancel{65}\,13}{\cancel{100}\,5} \times \cancel{20} = 13^{\text{index}}$$

80%

$$\Rightarrow \frac{4000}{500} \times 100$$

$$\Rightarrow \frac{400}{5}\ 80\%$$

$$\rightarrow L \qquad \frac{400}{5}$$

③ Inter quartile Rang

$$IQR = Q_3 - Q_1$$

$Q_1 = 25$

$Q_3 = 75$

$IQR = 75 - 25$

$= 50$

⑤ 5-number Summery

1. min
2. Q_1
3. median
4. Q_3
5. max

Eg:- 13 [2, 2, 3, 4, 5, 5, 6, 7, 8, 9, 12, 15, 79, 88, 91]

=> mn = 2

max = 91

$Q_1 = \frac{25}{100} \times 15+1$

$= \frac{25}{100} \times 16 \Rightarrow 4^{th}$ index

$\boxed{Q_1 = 4}$

$Q_3 = \frac{75}{100} \times 15+1$

$\Rightarrow 12^{th}$ index

$\boxed{Q_3 = 15}$

IQR = $Q_3 - Q_1$

$\Rightarrow 15 - 4$

$\boxed{IQR = 11}$

Lower fence

$$= Q_1 - 1.5(IQR)$$

$$= 4 - 1.5 \times 11$$

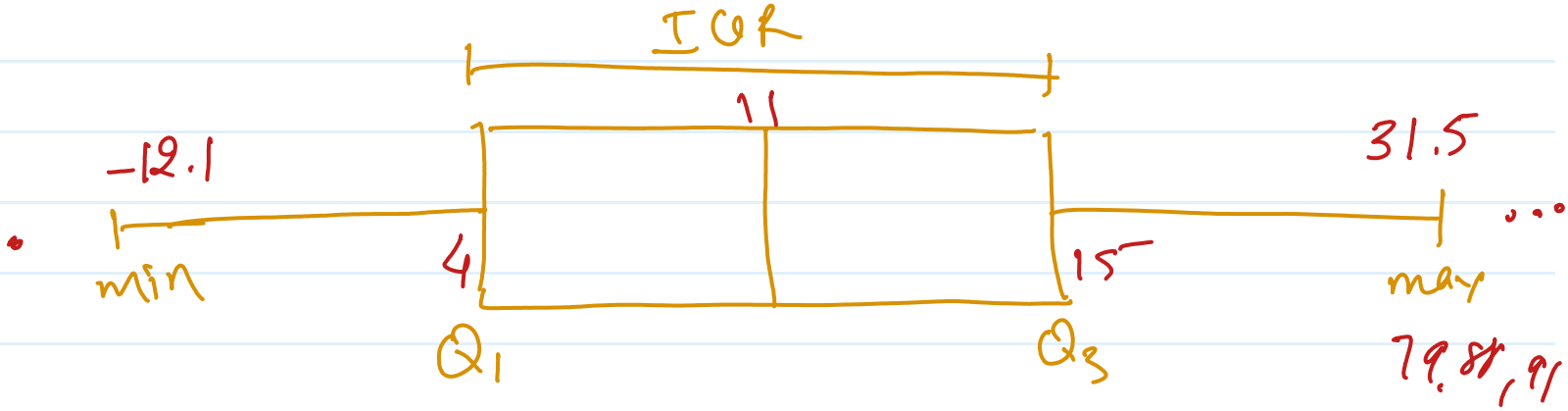
lower fence => -12.5

Upper fence

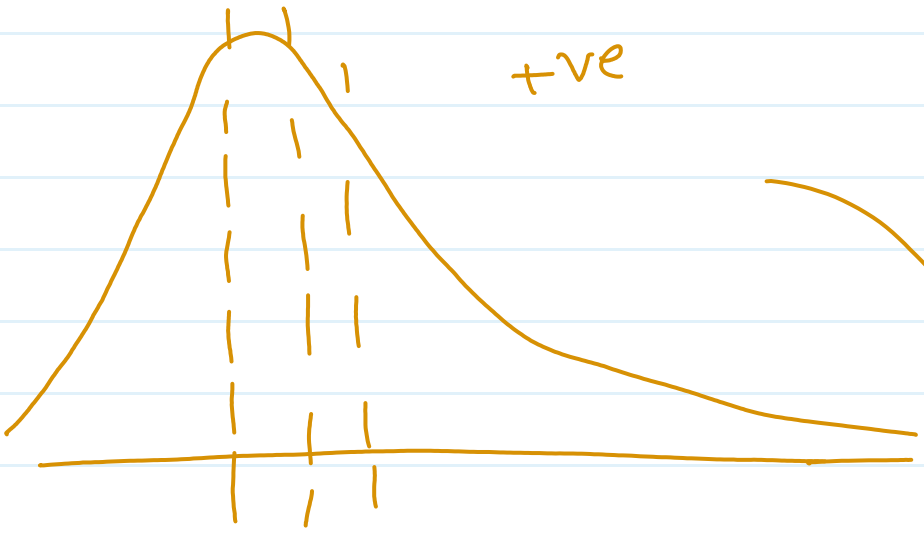
$$= Q_3 + 1.5(IQR)$$

$$= 15 + 1.5(11)$$

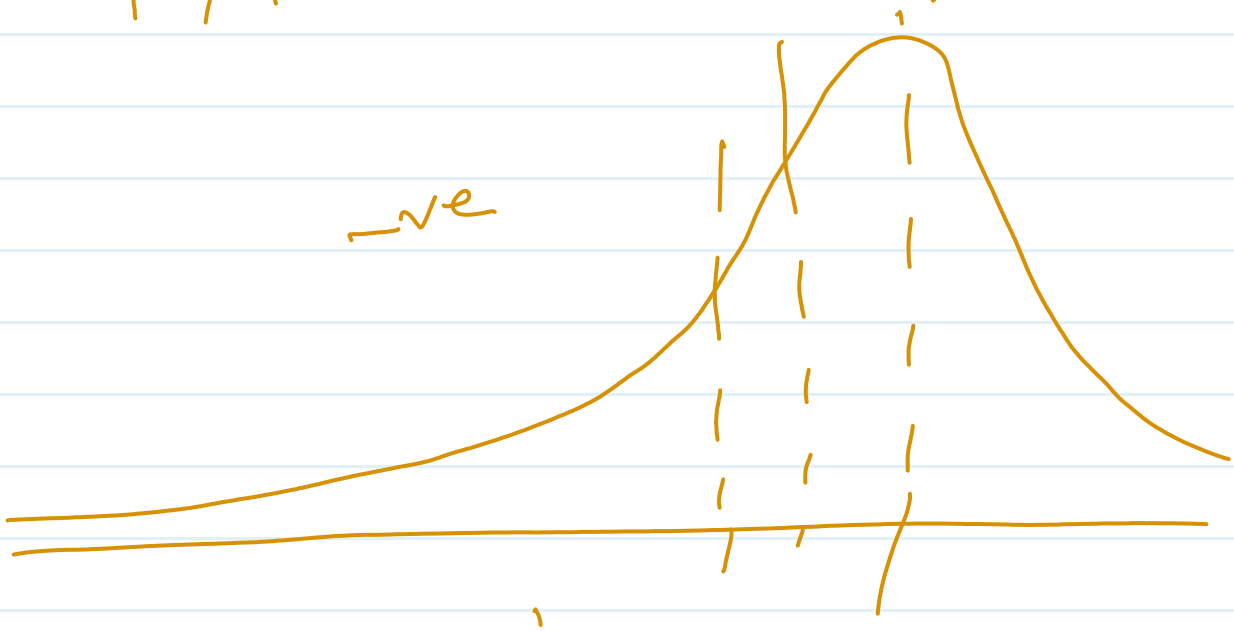
Upper fence = 31.5

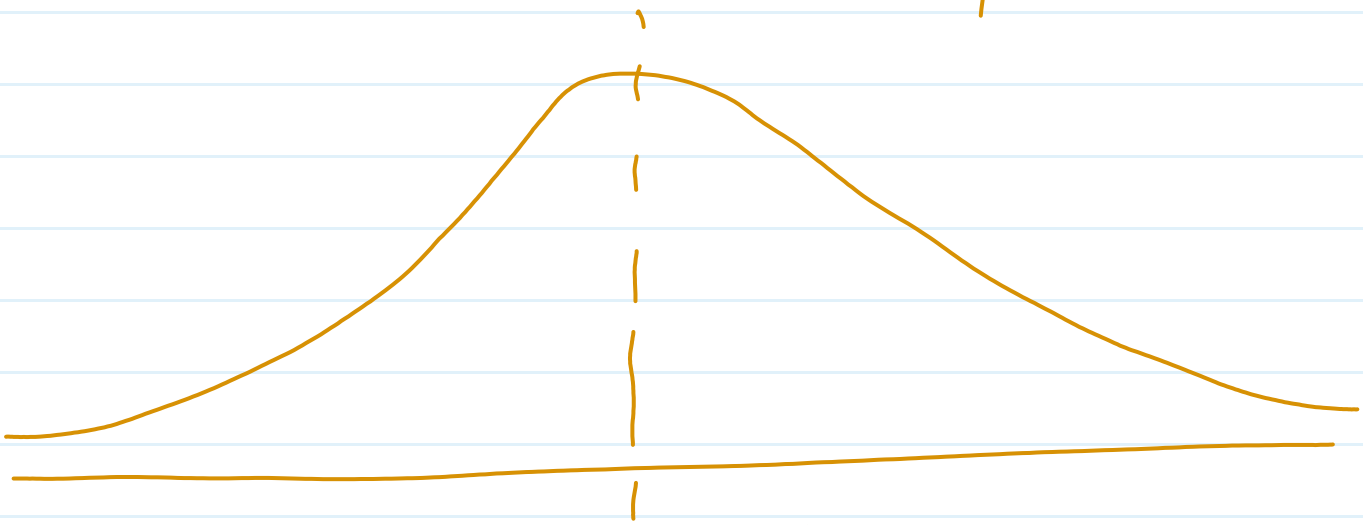


Shape



Asymmetric Dist.





Symmetric Dist.

Variables

$X = 10$

Type of Variable

Quantitative

① Descrete Quan. vari
eg. [Bank a/c, chnd. /home]
[whole number]

② Continous Quan. vari.
Eg [height, weight, Bank balance]
[Decimal number]

Qualitative / Categorical

① Nominal Quali. var.
[True/False, P/F]

② Ordinal Quali. var.
[education:- 10th
12th
Grade.
P.G.
Ph.D.]