Stat in DA

1. What is Statistics?

→ It is a science of Collecting, Organizing, analyzing data for better decision making.

2. What is Data?

→ facts or pieces of information that can be measured

ex: Iq of class
{ 28, 90, 50, ... }
age of students
{ 24, 22, 21, ... }

Types of Stats

1. Descriptive Statistics :-
It consist of organising and summarizing data.

2. Inferential Statistics :-
Using Data, we can make condision using some techniques.

Ex: class → 20 students
1st sem maths { 76, 70, 90, 53 ... }

1. what is the avg of class? (Descriptive)

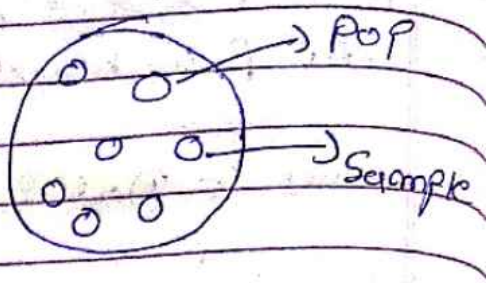
2. 7th sem (Inferential)

Sample and Population

Population = N

↓

Whole dataset is known as population



Sample = n

~~Small subsets~~ Small subsets of data taken from population → Sample

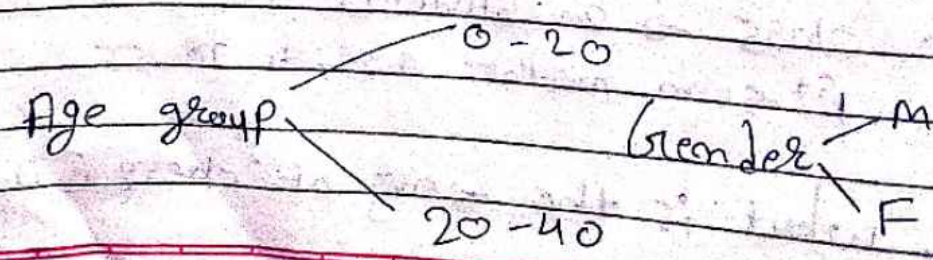
Sampling Techniques:-

1. Simple Random Sampling:-

Every member of population has an equal chance of getting selected in sample (n)

2. Stratified Sampling:-

Splitting data into non-overlapping groups



3. Systematic Sampling:

4. Convenience Sampling: surveying only those people who have knowledge of that particular domain.

Ex: Exit poll? (Random)

Survey of house hold (Each woman) [Convenience]

2.

Variables

↳ It is a property that can hold / store take any value

Age: {8, 10, 15, 20, 25, ...}

Marks: {70, 80, 95, ...}

Types:

1. Qualitative :- Categorical Values

Based on some characteristics, we can derive categorical values.

Eg: 0 - 10 ⟶ low ⟶

10 - 50 Avg

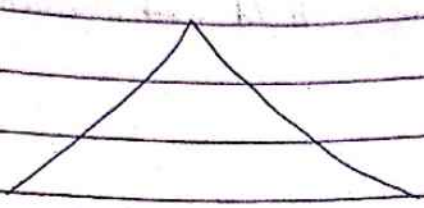
50 - 70 good

2. Quantitative:-

Numerical value (measurable numerically)

Height: {162, 159, 155...}

Weight: {59, 65, 79...}



Discrete (int)	Continuous (Float)
Whole No	Decimal No
1. No. of students	1. Height {165.2, 167.9...}
2. No. of bank ac.	2. weight {165.5, 160.9...}

1. Blood presure → Cantinuous / Discrete

2. Mathical stuties → Qualitative

3. River length → Conti

4. Song length → Conti

5. Gender → Qualitive (Category)

Variable measurement Scales

order matters

1. Ordinal : ordered [rank, graduation
2. Nominal : Categorical values / colors, classes, degrees)
3. Interval : [No zero / absolute point] (order as well as value matter)
4. Ratio :- Zero means nothing

Interval

20°C : 40°C
1°C : 2°C

Ratio

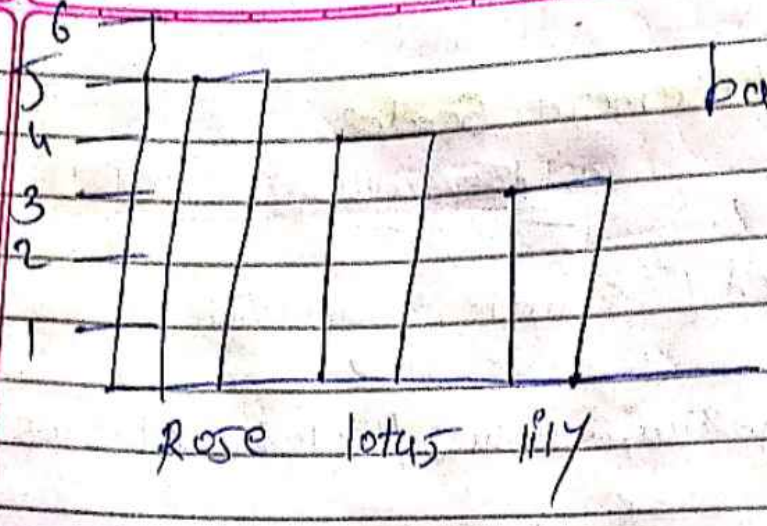
20kg : 40kg
1 : 2

Frequency

data : Flowers

[Rose, lily, lotus, Rose, lose, rose, Rose, lotus, lily, lotus, lily, lotus

Flowers	Frequency	Cumulative
Rose	5	5
lily	3	8
lotus	4	12



bar graph / chart

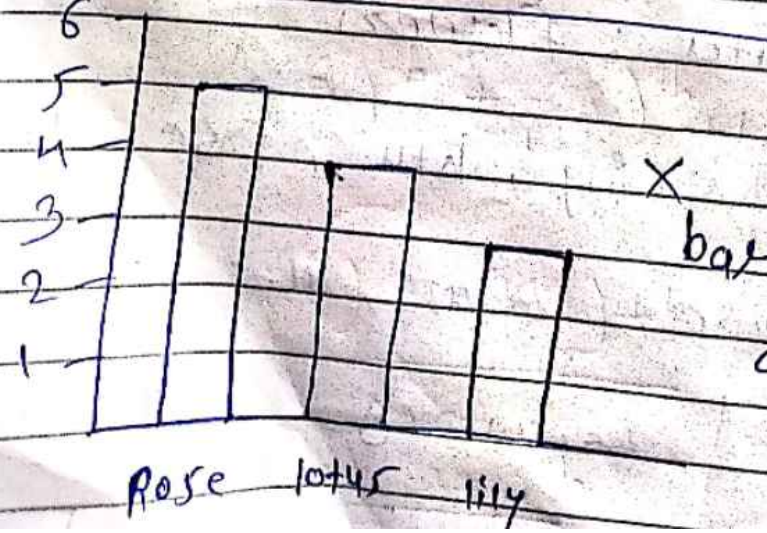
Histogram

marks : [2, 5, 12, 15, 21, 28, 35, 35, 34, 36, 39, 42]

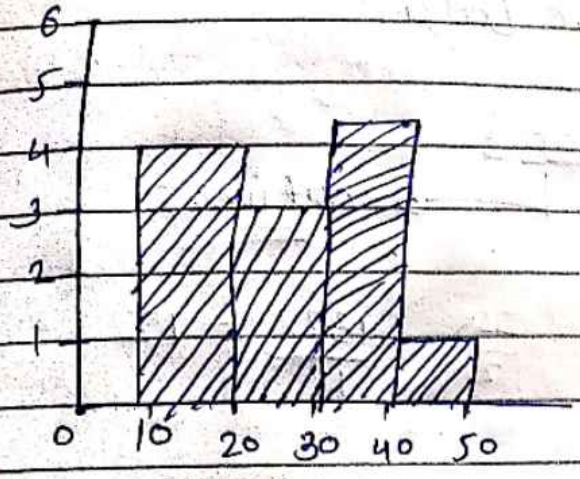
bin =

[0-10] - 2
[10-20] - 2
[20-30] - 3
[30-40] - 4
[40-50] - 2

⟹ continuous



X bar graph / chart



→ Measure of Central Tendency

→ Measure of Dispersion

→ Distribution

Measure of Central Tedency

Avg → mean

Pop

$$\mu = \frac{\sum xi}{N}$$

Sam

$$\bar{x} = \frac{\sum xi}{n}$$

{ 2, 3, 5, 3, 2, 1, 3 }

$$\frac{2+3+5+3+2+1+3}{7}$$

mean :- It refers to the measure used to the distribution.

of the Data.

{1, 1, 2, 2, 3, 3, 4, 5, 5, 6, 100}
└ outliers

$\frac{32}{10} \rightarrow 3.2 \rightarrow \frac{132}{11} = 12$

Median → middle Value

↳ as ascending order

↳ data → even
→ odd

Even:-

$$\frac{\left(\frac{n}{2}\right)^{th} + \left(\left(\frac{n}{2}\right)^{th} + 1\right)^{th}}{2}$$

dataset: {11, 12, 13, 14, 15, 16}

$n = 6$

$$\left(\frac{6}{2}\right)^{th} + \left(\frac{6}{2}\right)^{th} + 1 \quad \left| \quad \frac{3^{th} + 4^{th}}{2}\right.$$

$$\frac{\left(\frac{6}{2}\right)^{th} + \left(\frac{6}{2}\right) + 1}{2} \quad \left| \quad = \frac{13+14}{2}\right.$$

$$\frac{3^{th} + (3+1)^{th}}{2} \quad \left| \quad = \frac{27}{2}\right.$$

$$\left| \quad = 13.5\right.$$

odd :

$$\left(\frac{n+1}{2}\right)^{th} \qquad \frac{(n+1)^{th}}{2}$$

{11, 12, (13), 14, 15}

$n = 5$

$$\frac{(5+1)^{th}}{2} = \frac{6^{th}}{2} = 3 = 13$$

→ {11, 12, 13, 14, 15, 100}

$$\frac{\left(\frac{n}{2}\right)^{th} + \left(\left(\frac{n}{2}\right)^{th}+1\right)^{th}}{2}$$

$$\frac{\left(\frac{6}{2}\right)^{th} + \left(\left(\frac{6}{2}\right)^{th}+1\right)^{th}}{2}$$

$$\frac{3^{th} + 4^{th}}{2}$$

$$= \frac{13+14}{2}$$

$$= 13.5$$

{21, 23, 25, 29, 32, 100}

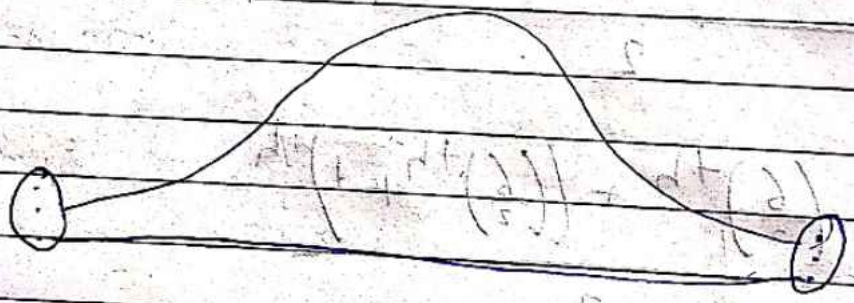
median works well with outliers

$$\frac{25+29}{2} = \frac{54}{2} = 27$$

outliers :-

a data point who doesn't follows pattern or trend of the dataset then it is considered as outlier

[are extreme point]



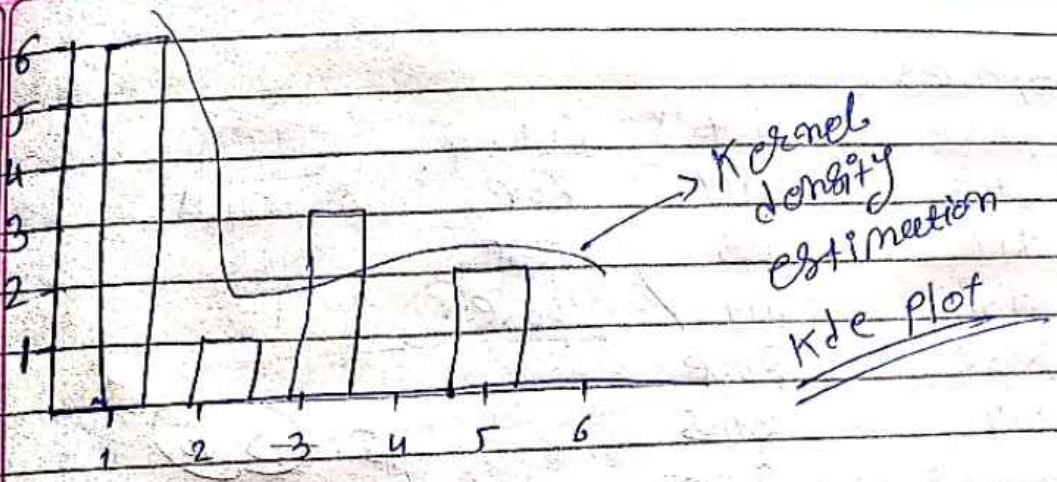
Mode

↳ most frequent value (repeated)

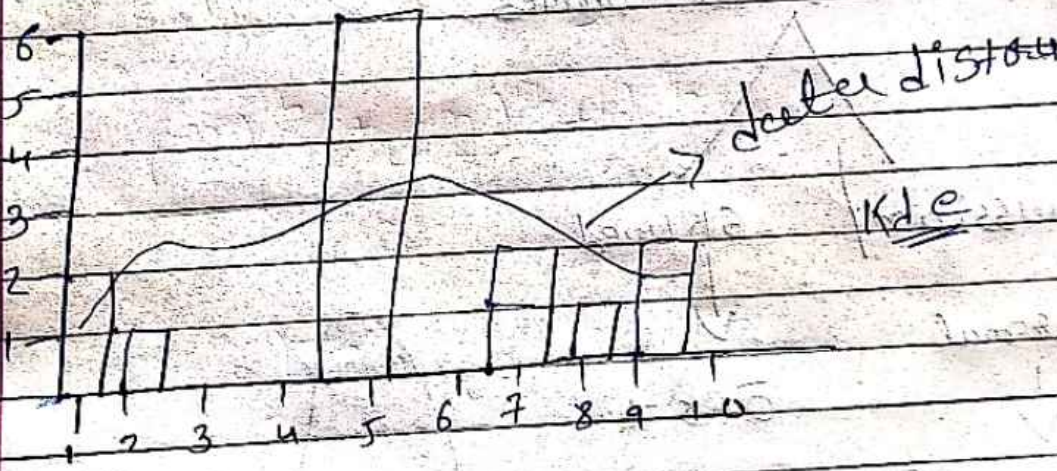
[1, 1, 2, 3, 5, 1, 1, 3, 1, 3, 5, 1]

1 = 6 3 = 3

2 = 1 5 = 2



{1, 1, 2, 5, 9, 5, 7, 8, 9, 5, 9, 5}



for categorical missing data

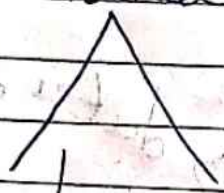
0-5% → mode

→ new category "missing"
"unknown"
or "random"

Spaces

Rose	lily	lily	
lotus	lily	Rose	null
lily	mob	Rose	
misi	lily	Rose	
Rose	Rose		
lily	lotus	15 → 10	
Rose	Rose		

For numerical values



Gaussian / Skiwed

Normal

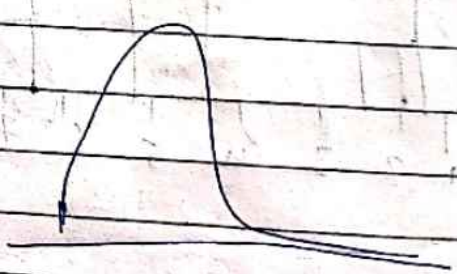
↓

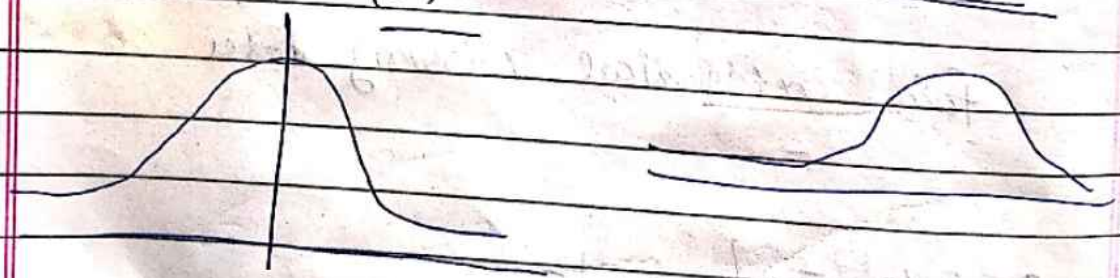
Mean

↓

0-5-10

10%





mean = median = mode

Measure of Dispersion

↳ Spread

[5, 1, 1, 1, 2] → $\frac{5+5}{5}$ = 2

[2, 2, 2, 2, 2] → 2

→ Variance

It measure how far the numbers in a dataset are from the mean (Avg).

(How each value differ from a change in mean)

High Variance → more spread (far from mean)

low Variance → closer to mean

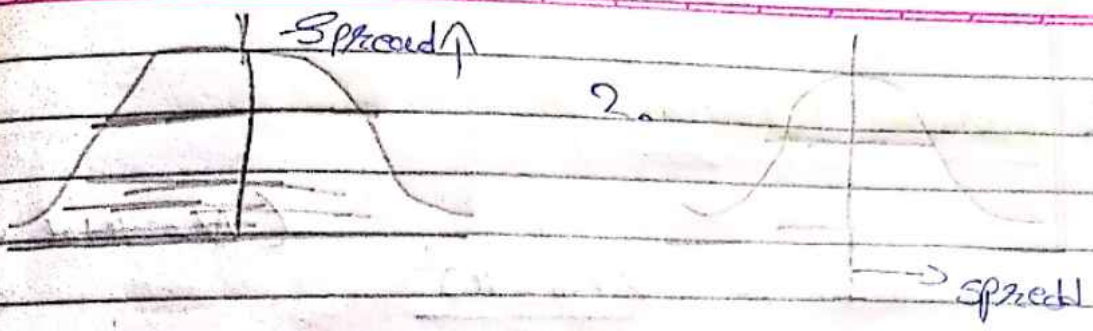
PoP

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

Sam

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

Bessel's correction



Standard Deviation

→ Just a squared root of Variance

→ it gives you a measure of spread that is in the same units as the original data, making it easier to interpret than variance.

[Same unit, easily comparable]

Population

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

Sample

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

Key point (Variance and SD)

- Variance gives you the averge squared distance from the mean.
- SD gives you a measure of squed in the same unit.

x_i = data point

N = total no. of population

n = total no. of sample

μ = mean of population

$\bar{x}$ = mean of sample

Q ⇒ Why $(n-1)$ is used in formula of Sample

x	$x-\mu$	$(x-\mu)^2$
1	-1.83	3.34
2	-0.83	0.68
3	-0.83	0.68
4	0.17	0.02
5	1.17	1.36
2.83	2.17	4.70
		10.78

$$\sigma^2 = \frac{10.78}{6} = 1.79$$

$$\sigma = 1.33$$

Representative of μ. This correction ensures that the sample variance is an unbiased estimator of the population variance.

outliers

* **percentage** ⟶ $\frac{\text{sum of nos of interest}}{\text{sum of all data pts}} \times 100$

1, 2, 3, 4, 5

% of the nos that are odd?

$\frac{3}{5} = 0.6 \rightarrow \times 100$

$= 60\%$

* **percentile**

-> A percentile is a value below which a certain percentage of observation lie.

[95, reading]

5.5 ⇒ 6 index

Data set = 2, 2, 3, 4, 5, 5, 5, 6, 7, 8, 8, 8, 8, 8, 9, 9

marks ⟶ 10, 11, 11, 13

15

$n = 20$ ceiling

* Variance formula

→ Population :- $\frac{\Sigma (x - \mu)^2}{N}$ (N = total no of data pts

When we have data from the entire population we use 'N' in the denominator. This gives us an exact measure how the data pts vary around the population mean (μ)

* Sample :- $\frac{\Sigma (x_i - \bar{x})^2}{n-1}$

→ When we're working with a sample, only have sample mean $\bar{x}$, which is an estimate of the population mean μ.

→ Using sample mean in calculations tend to make the variance slightly smaller than the true population variance.

→ To correct this bias (underestimating variance) we divide by 'n-1' instead of n. This makes variance estimate larger and accurate.

→ By subtracting 1 from n, We account from the fact that the '$\bar{x}$' is not perfectly.

Date
Page

Percentile rank = $\frac{\text{no. of values below } n}{n}$

$\frac{16}{20} = \frac{4}{5} \times 100 = 80$

$\frac{17}{20} \times 100 = 85$

Q) What value exists at percentile ranking of 2

$\text{Value} = \left(\frac{\text{percentile}}{100} \times n\right) + 1$

$= \left(\frac{25}{100} \times 20\right) + 1$

$= 5 + 1 = 6$

Index

75 percentile

$\text{Value} = \left(\frac{75}{100} \times 20\right) + 1$

$15 + 1 = 16$

Five Number Summary

1. Minimum
2. First Quartile (Q1)
3. median
4. Third Quartile (Q3)
5. Maximum

[to detect outliers]

[1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 6, 6, 7, 8, 8, 9, 15, 27]

[lower fence <——> higher fence]

Lower fence = Q1 − 1.5 (IQR)

higher fence = Q3 + 1.5 (IQR)

Q1 = 25 percentile (25%)

Q3 = 75 percentile (75%)

IQR = Inter Quartile Range

= Q3 − Q1

$$Q_1 = \left(\frac{25}{100} \times 20\right) + 1$$

$$= 6$$

$$Q1 = 3$$

$$Q_3 = \left(\frac{75}{100} \times 20\right) + 1$$

$$= 16$$

$$= 8$$

$$IQR = 8 - 3 = 5$$

Lower fence $= Q_1 - 1.5\,IQR$

$= 3 - 1.5(5)$

$= 3 - 7.5$

$= -4.5$

Upper fence $= Q_3 + 1.5\,IQR$

$= 8 + 1.5\,IQR$

$= 8 + 7.5$

$= 15.5$

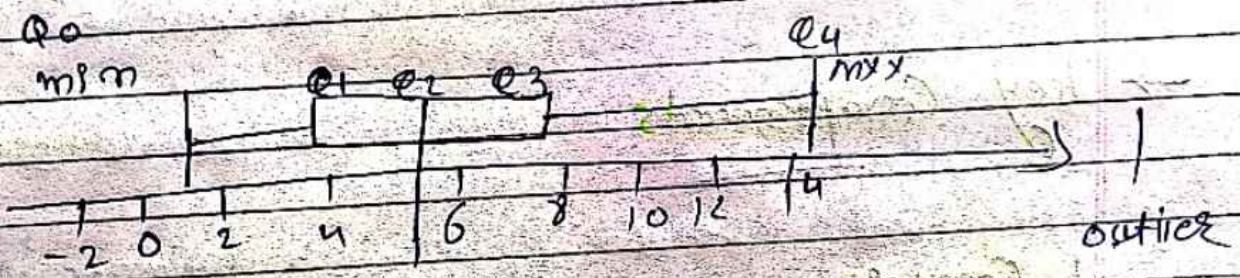
Examining data

{1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 6, 6, 7, 8, 8, 8, 9, 15}

5 No. Summary

min = 1
Q1 = 3
midian = 5
Q3 = 8
max = 15

} Boxplot



✱ Data distribution

→ A distribution refers to the way in which values or data pts are spread or arranged.

→ It shows how often different values occur in data set and describes the overall pattern of the data.

→ **Key Components**

1. **Center :-**
   where the middle data lies (mean, median)
2. **Spread :-**
   How wide or narrow the data is (range, variance, SD)
3. **Shape :-**
   The overall from of the distribution (symmetric, skewed, etc)
4. **outliers :-**
   Data points that are much higher or lower than the rest of the Data.

Types of Distribution

1. Gaussian / Normal Distribution

what?

-> A symmetric bell shape where most data pts cluster around the mean (center) and fewer values occur as you move away from the mean on both directions.

when?

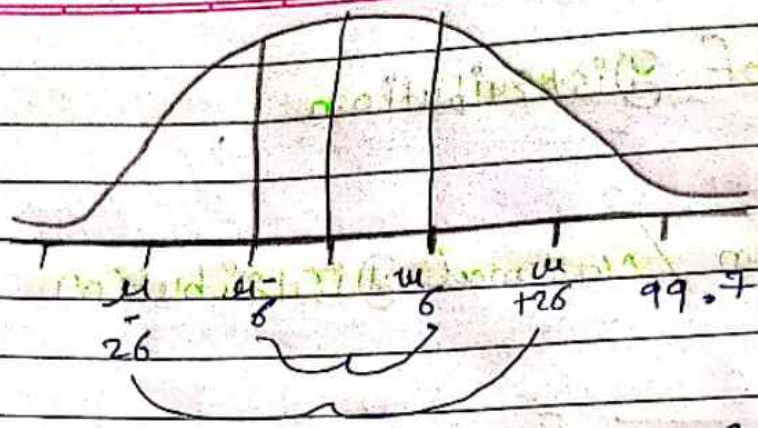
-> when data is evenly distributed around the mean and follows a natural pattern. (height, test scores etc)

why?

-> It's useful because many natural phenomenon follows this pattern and it's a foundation for many statistical methods

where?

-> Education, Biology and Quality Control



outlier

- Centered
- equally distributed
- Bell shape
- mean = mode = median
- Normal Dist
- Empirical formulas
- symmetric

$\mu - 1\sigma$ $\mu + 1\sigma$ ⇒ 68%.

$\mu - 2\sigma$ $\mu + 2\sigma$ ⇒ 95%

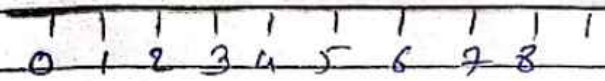
$\mu - 3\sigma$ $\mu + 3\sigma$ ⇒ 99.7%.

$\mu = 4$ $\sigma = 1$

4.5

4.75

6.95



Sample = $\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ Population

$\Rightarrow$ Z-Score

How?

$Z = \frac{x - \mu}{\sigma} = \frac{6.95 - 4}{1} = 2.95\sigma$

what?

The Z-score is a statistical method that indicates how many SD a specific data pts is from the mean.

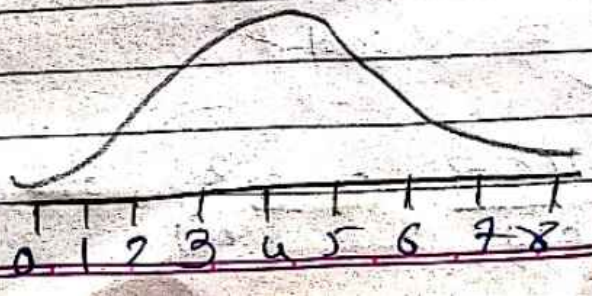
$a = [1, 2, 3, 4, 5] \Rightarrow \mu = 3$

$b = [2, 2, 4, 5, 6] \Rightarrow \mu = 3.8$

when?

Z-scores are used when comparing individual data pt to the average in normally distributed data set or when converting different types of data to a common one/scale.

$\mu = 4 \quad \sigma = 1$ [0-8]



Z-score

$$Z\text{-score} = \frac{x_i - \mu}{\sigma}$$

$\mu = 0 \quad \sigma = 1$

Standard Normal Distribution

$\mu = 0 \quad \sigma = 1$

Hight	Weight	$h_i - \mu_m$	
169	60	3.2	10.24
172	65	6.2	38.44
150	45	-15.8	249.64
168	70	2.2	4.84
170	71	4.2	17.64
829			320.8

$$\mu_m = \frac{829}{5} = 165.8$$

$$\sigma_m^2 = \frac{320.8}{5} = 64.16$$

$$\sigma_m = 8.009$$

$$z_1 = \frac{169 - 165.8}{8.009} = \frac{3.2}{8.009} = 0.3994$$

$$\approx 0.4$$

$z_2 = \frac{6.2}{8.009} = 0.7$

$z_3 = \frac{-15.8}{8.009} = -1.9$

$z_4 = \frac{2.2}{8.009} = 0.27$

$z_5 = \frac{4.2}{8.009} = 0.52$

$\mu = -0.01$

Normalization

$x_n = \frac{x - x_{min}}{x_{max} - x_{min}}$ → min max scaler

positively skewed distribution

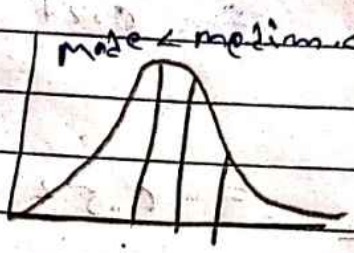
What? :-

In positively skewed distribution, most values are concentrated on the lower and with a long tail extending to the right. A few high values pull the average to the right of the median.

Skewness :-

A distortion or asymmetry that deviates from the symmetrical bell curve.

- also called right skewed or right tailed distribution



mean > median > mode

When? :-

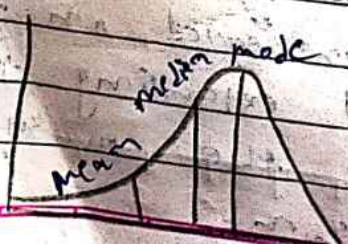
Useful for data with rare but significant high values such as income levels where few individuals earn much more than the rest.

- the mean is greater than the median due to outliers on the higher end.

Negatively skewed distribution

What? =

Most values are clustered at higher end with a few values creating a long tail to the left.



When?:-

used for datasets where values are typically higher, but a few lower values exist (retirement age)

- the mean is generally less than median due to leftward tail.

- Also called left skewed or left tailed

Why?:-

It helps detect cases where data points are generally higher but occasionally much lower.

Exponential Distribution (continuous) λ = constant rate

- It describes the time b/w events in a process where events occur independently and at a constant rate λ.

$$f(x) = \lambda e^{-\lambda x} \quad x \geq 0$$

Bernoulli distribution (Discrete)

- It models a single experiment with two possible outcomes.

success ($x=1$), failure ($x=0$)

Binomial distribution : Discrete

- Binomial dist extends Bernoulli to n independent trials.

$$P(X=K) = \binom{n}{K} p^K (1-p)^{n-K}$$

X = random variable (no. of sucess out of n trial)

n = total no of trial (100)

K = no of success $(0 \le K \le n)$

P = P(sucess in 1 trial) e.g. 0.5

$$\binom{n}{K} = \frac{n!}{(n-K)! \cdot K!}$$

$n=5 \quad K=3 \quad p=0.5$

$$P(3) = \frac{5!}{2! \cdot 3!} \times 0.5^{(3)} \times 0.5^{(2)}$$

Uniform distribution Continuous

$[a, b]$

$$F(x) = \frac{1}{b-a} \qquad a \le x \le b$$

The probability of any value within the range $[a, b]$ is same.

Uniform Distribution (Discrete)

all outcomes are equally likely

$$P(x) = \frac{1}{n} \longrightarrow \text{total no.}$$

Confidence Interval

$\bar{x} = \mu \longrightarrow$ point estimate $\rightarrow \mu$

40-60 → 25K

2 lakh

5 - 50 + 5

45 - 55 → 50K

40 - 50 60 ± 10 - P0

±5 ±10 margin of error

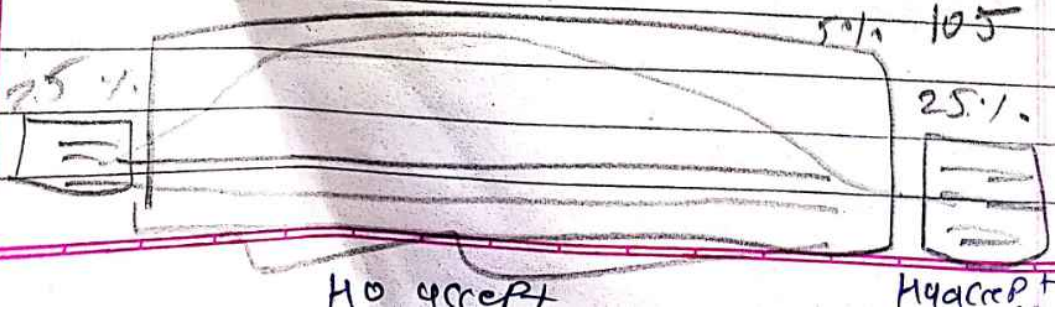
50 → point estimation

45 - 55
40 - 60 Confidence interval

→ It is a range of values within which we expect a particular population parameter to fall

Confidence Interval = Point estimate ± margin of error

Confidence level = 95%



Hypothesis Testing

left → loss coin

10 1000

- A statistical hypothesis test is a method of statistical inference used to decide whether the data at hand is sufficient to support a particular hypothesis.

Hypothesis testing allows us to make probability - stic statements about population parameters.

Null Hypothesis H0

- The null hypothesis assumes that there is no significant relationship or effect b/w two variables. [in simpler terms → it says "nothing new is happening"]

It serves as a starting point for HT and represents a cautious or the assumption of effect until proven otherwise.

- The purpose of HT is to gather evidence to reject or fact Null hypothesis in favour of alternate hypothesis, which claims there is significant effect or relationship.

Alternate hypothesis — Ha or H1

- It is a statement, that contradicts the NH and claims there is significance effect or relation.

Rejection Region method

1. H0 and Ha
2. α → value — 5% [0.05] — 95% Loss of significance
   - significance level → 95%.
3. assumptions — ND → $n \geq 30$
4. decide test — 2-test, t-test
5. value
6. Test Conclusion

7. Rejection of H_0

8. $n \geq 50$

2. Avg - 50g 40g → $\bar{x} - \mu$

6 - 4g

1. $H_0 : \mu = 50g$ $H_1 : \mu \neq 50g$

2. $\alpha = 0.05$

3. $n \geq 30$ Z - test

4. Z - test

5. $= \frac{49 - 50}{4} \times \sqrt{40}$

$= \frac{-1}{4} \times \sqrt{40}$

$= \frac{-\sqrt{40}}{4}$

$Z = 1.58$

7. Reject / Accept

8. State Result

1. To → units per day

$\sigma = 5$

Training → (45-55)

↓

30 emps → 53 units per day

1. $\mu = 50 \rightarrow H_0$ $\quad H_a : \mu > 50$ (one tailed test)

2. $\alpha = 0.05$ 5%

3. Data normal, & Random
$n = 36$

4. Z-test

5. $Z\text{-score} = \frac{\bar{x} - \mu}{\sigma}$ - PDF

$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{53 - 50}{5/\sqrt{30}}$

$= \frac{3}{5} = \sqrt{36} = \frac{9}{20} \times 30$ &

6.

±1.96

-P=96 1.96 001 20/2
20/2 0025
0025 97.5

0.975

7. Null hypo Accept

8. $\mu = 50g$

11 Errors Type 1 Type 2

Ho true Ho false

reject Ho Type 1 correct

accept Ho correct Type 2

Type 1 False +ve

↳ Ho reject → Ho true (correct)

rejecting Ho when Ho is actually correct.

Type-2 → False-ve

accept H_0 when H_0 is actually incorrect

$P \geq \frac{0.05}{\alpha}$ → Null accept

$P < 0.05$ → Null reject

$P < 0.001$ → Strong evidence

$0.01 \leq P < 0.05$ → moderate evidence

$0.05 \leq P < 0.1$ → weak evidence

$P \geq 01$ → No evidence

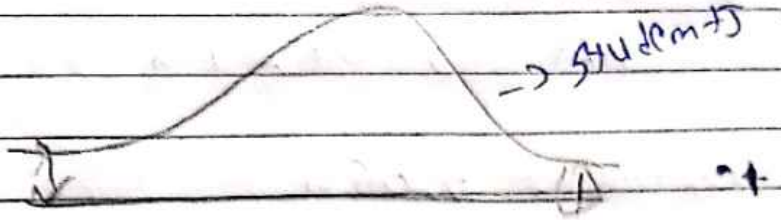
$P < \alpha$ → H_0 reject

$P \geq \alpha$ → H_0 accept

P-value :-

It is a measure of the strength of the evidence against the null Hypothesis.

T-test



3 types

1. one Sample t-test

→ Compares the mean of a single sample to a known μ

$\mu = 50g \quad n = 30 \quad \bar{x} = 49.7g$

$S = 1.2$

(when we $\sigma \rightarrow$ z-test (pop → standard dev)

$\sigma \not\rightarrow$ t-test] (Sample - std)

Q $\mu = 50g \quad n = 25 \quad Std = 1.2g$

$\bar{x} = 49.7g$

$H_0 : \mu = 50 \qquad H_1 : \mu \neq 50$

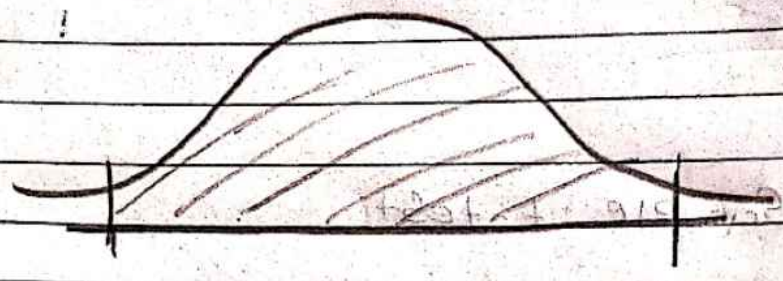
$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{49.7 - 50}{1.2/\sqrt{25}} = \frac{0.3}{1.2} \times 5$$

$$= \frac{1.5}{1.2} = -1.25$$

df = degree of freedom

$n-1 \rightarrow df = 24$

tcritical = 2.064



-2.064 2.064

H0 accept → 50%

2. Independent Two Sample t-test

$$t = \bar{x}_1 - \bar{x}_2$$

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \rightarrow \text{standard error}$$

3. paired t-test

$$t = \frac{d}{s/\sqrt{n}}$$

$\bar{d}$ = 2 ⇒ mean diff

sd = std → diff

Chi-Square test

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

O → observed frequency

E → Expected frequency

$$E = \frac{\text{Row} \times \text{Column total}}{\text{Grand total}}$$

$$df = (r-1) \times (c-1)$$

r = no. of rows

c = no. of cols

Q

	Satisfied	Not satisfied	Total
High school	50	70	120
college	90	60	150
PG	20	16	30
	160	140	300

EF	S	NT	Total
HS	$\frac{120 \times 160}{300} = 64$	$\frac{120 \times 140}{300} = 56$	
C	$\frac{150 \times 160}{300} = 80$	$\frac{150 \times 140}{300} = 70$	
PG	$\frac{160 \times 30}{300} = 16$	$\frac{140 \times 30}{300} = 14$	
S	64	56	120
C	80	70	150
PG	16	14	30

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$S.S = \frac{(50-64)^2}{64} = \frac{196}{64} = 3.06$$

$$S.NS = \frac{(70-56)^2}{56} = \frac{196}{56} = 3.5$$

$$C.S = \frac{(90-80)^2}{80} = \frac{100}{80} = 1.25$$

$$C.NS = \frac{(60-70)^2}{70} = \frac{100}{70} = 1.42$$

$$P.H.S = \frac{(20-16)^2}{16} = \frac{16}{16} = 1$$

$$P.H.NS \quad \frac{(10-14)^2}{14} = \frac{16}{14} = 1.14$$

$$\chi^2 = 3.06 + 3.5 + 1.25 + 1.42 + 1 + 1.14$$

$$= 11.37$$

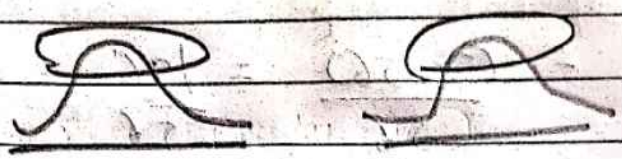
$$DF = (3-1) \times (2-1)$$

$$= 2$$

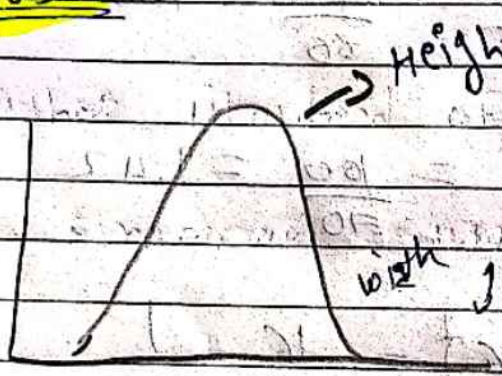
X^2 critical = 5.991

$X^2 = 11.37$

reject H_0



Kurtosis



- kurtosis measures "tailedness" of a distribution or how extreme the outliers are.

1. Mesokurtic

- Tails are similar to ND
- distribution with kurtosis ≈ 3

Eg. Standard normal distribution

2. Leptokurtic

- A distribution with kurtosis > 3
- Heavy tails (with more extreme outliers)

eg - t-distribution with very small df

3. Platykurtic

- distribution with kurtosis < 3
- Light tails (fewer extreme outliers)

eg: Uniform dist

Excess kurtosis → K - 3

K = 0 → Leptokurtic

K < 0 → Platykurtic

$$K = \frac{n \cdot \sum (x_i - \bar{x})^4}{\left(\sum (x_i - \bar{x})^2\right)^2} \cdot \frac{1}{n}$$

n = no of observation

x_i = each data point

$\bar{x}$ = mean

=> High kurtosis (Leptokurtic)

- more extreme outliers
- Higher likelihood of rare, extreme values
- eg. financial returns during market crisis/ crashes.

Lower kurtosis (Platykurtic)

- Fewer extreme outliers
- Data is evenly spread.

Kurtosis near 3 (Mesokurtic)

- Similar to normal distribution.

Leptokurtic distribution

Normal distribution

Platykurtic distribution