Mean → Find average of number.
→ sum of all points divided by no of data pts.

$$\text{mean} = \frac{\text{sum of data}}{\#\text{ of data points}} = \frac{\Sigma x_i}{n}$$

e.g. 1, 2, 4, 5 find mean

$$\frac{1+2+4+5}{4} = \frac{12}{4} = 3$$

Median :→ median is the middle point in a data set
→ half of data pts are smaller than the median and half are larger than median

To find median
→ arrang data points in asscending order
→ if no of pts are odd the median is middle data pt.
→ if no of pts are even then median is avg of two middle points.

$$\text{median} = \frac{n}{2}\text{th term (if } n \text{ is odd)}$$

$$= \left(\frac{n}{2}\text{th term}\right) + \left(\frac{n}{2}+1\text{ th term}\right) \text{ if } n \text{ is even}$$

eg 1, 4, 2, 5, 0
0, 1, 2, 4, 5

10, 40, 20, 50
10, 20, 40, 50
(20+40)/2 = 30

Mode :→ mode is most commonly occurring data pt. in a dataset
→ mode is useful when there are a lot of repeated value in a dataset.
→ There can be no mode, one mode or multiple mode in a dataset

e.g $0, 0, \underbrace{1, 1, 1, 1, 1, 1}, 2, 2, 2, 3, 5$

mode is 1

(2) $0, 0, 0, \underbrace{1, 1, 1, 1}, \underbrace{2, 2, 2, 2}, 4$

tie bet two

modes are 1, 2

⇒ [These are known as measure of central tendency, represent all the values of data.]

Grouped Frequency Table ⟷

que

Seconds	freq^n
51-55	2
56-60	7
61-65	8
66-70	4

Estimated mean from group data.

find midpoint of each class

(1)

midpoint	freq^n
53	2
58	7
63	8
68	4

(2) Data is like that now

53, 53, 58, 58, 58, 58, 58, 63, 63, 63,
63, 63, 63, 63, 63, 68, 68, 68, 68

(x) midpoint	(f) freq^n	fx
53	2	106
58	7	406
63	8	504
68	4	278
	21	1288

Estimated mean $= \frac{1288}{21}$

$= 61.337$

Estimated median is

median is middle value ie 11^{th} which is in class 61-65

SD → SD is a measure of how spread out no. are. denoted by by σ.
formula is is sqrt(var) ie $\sqrt{var}$.

Variance → Avg of squared diff from the mean

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

$$\sigma = \sqrt{var} \quad \text{i.e} \quad \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2}$$

μ = mean
x_i = sample
N = Total no of pts

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

The sample SD $s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2}$

Discrete Data → Can take only certain values
e.g. Result of rolling two dice
2, 3, 4, 5, 6, 7, 8, 9, 10, 11 & 12

Continuous Data → Continuous data can take any value (within a range).
e.g A person's height (Could be any value
- Time in a race
- weight

Data
→ Qualitative Data → is descriptive info. (means describe someth-
e.g "Weather is Cool"
→ Quantitative Data → is numerical info (number

Discrete — 5
Continuous — 3.265 - -

e.g → Abt Dog

Qualitative
→ He is brown & black
→ He has long Hair
→ He has lot of energy

Quantitative
Discrete
→ 4 legs
→ 2 brothers

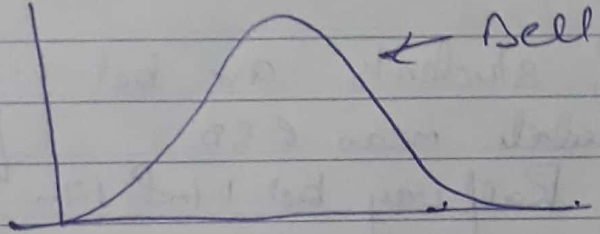
Continuous
→ weight 25.5 kg
→ 565 mm tall

Data Can be Collected in many ways →
→ simple is direct observation
→ by survey

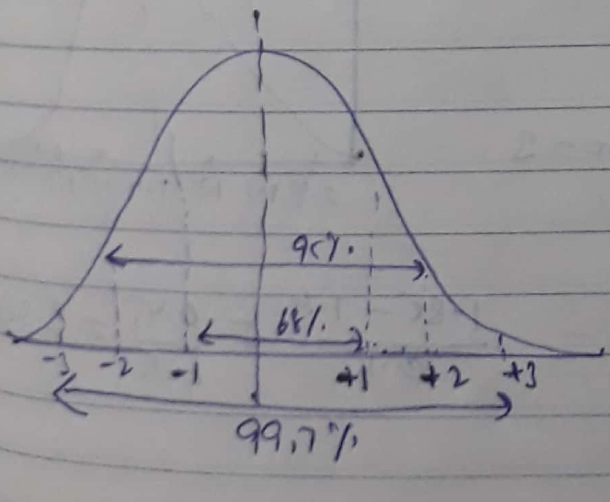
Normal Distribution → A ND, called the "bell Curve" that occurs is symmetrical means half of data will be fall to the left of mean, half will fall right to the mean.

e.g Height of people
measurement of error
Blood pressure
IQ Scores
Salaries etc.

It look like



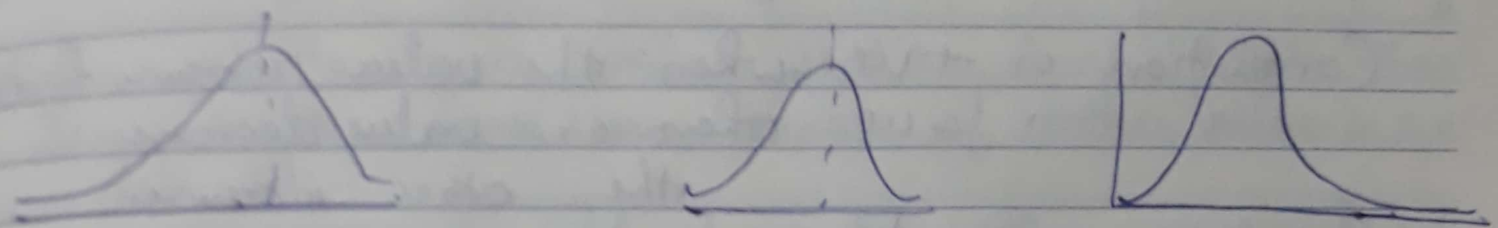
Properties of ND are :-
→ mean, mode and median are equal
→ Curve is symmetrical about center
→ 50% of values are less than mean and 50% greater than mean
→ Total area under the curve is 1.



68% of values are within 1 SD of the mean
95% of values are within 2 SD of the mean
99.7 of values are within 3 SD of the mean

1 SD = 68 out of 100
2 SD = 95 out of 100
3 SD = 997 out of 1000

Skewed Data → mean It tends to have a long tail on one side or another



-ve skewed
or Skewed to left
It has long tail on -ve side of peak
mean is on the left of peak value

No Skew

+ve skewed
or right skewed.
→ mean is on the right of peak value

Skew can be calculate using SKEW() func^n in excel.

SUNDAY 16

Condition for ANOVA
→ Data are randomly sampled
→ var. of each sample are assumed equal
→ Residual are Normally distributed

21 FRIDAY

SHIVALIK GROUP OF INSTITUTIONS Ambala (Haryana)

ANOVA → Analysis of variance is a collection of statistical models and their associated estimation procedure used to analyze the differences among group means in a sample

Steps are

1. Calculate the mean
   → Calculate mean, SD and variance of each gp.
   → Calculat overall mean (Grand mean) $\bar{\bar{x}}$

$$\bar{\bar{x}} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \dots + n_k \bar{x}_k}{n_1 + n_2 + \dots n_k}$$

n_i → no of in gp i
$\bar{x}_i$ → mean of gp i

2. Setup the null and alternate hypothesis
   → null hypothesis is that means are all equal
   $H_0 = \mu_1 = \mu_2 = \dots \mu_i$
   → Alternate hypothesis is that atleast one of the mean is different

3. Calculat the Sum of Squares

$$SS(B) = \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{\bar{x}})^2$$

$$SS(W) = \sum_{i=1}^{k} df_i \, s_i^2$$

df = degree of freedom
s_i = SD of i^{th} gp
or
s_i^2 = variance of i^{th} gp

$SS_{total} = SS_B + SS_W$

4. Calculate the Degree of freedom.

df(total) = n-1

df within = ~~[illegible]~~ n-K df(W)

df between = ~~[illegible]~~ K-1 df(B)

E-mail: shivalik.ambala@gmail.com

Scanned by CamScanner

(5) Calculate the mean square

$MS(B) = SS(B) / df(B)$

$MS(W) = SS(W) / df(W)$

$MS(T) = SS(T) / df(T)$

(6) Calculate the F statistic

$F = MS(B) / MS(W)$

- F test is a right tail test
- F test statistic has an F distribution with df(B) and df(W)

→ The P-value is the area to the right of the test statistic

(7) Look up statistical Table and state your Conclusion

Look up the tabulated value of F (critical value) from the statistical table and compare it with the value you calculated (abs value)

if abs value (calculated value) > critical value then reject the null hypothesis and conclude that there is significant diff bet. the means of population.

otherwise accept null hypothesis or fail to reject null hypothesis

Significant level $\alpha = 0.05$

Source	SS	df	MS	F	P
B					
W					
T					

MONDAY

23 24 25 26 27 28

Null hypotheses → if we compare method A with method B abt its superiority and if we proceed on the assumption that both method are equally good than this assumption is termed as Null Hypotheses.

represented by H_0

As Against this the method A is superior or the method B is inferior, we are than stating is termed as alternative hypothesis rep. by H_a.

eg μ = population mean

μH_0 = Hypothesis mean

Simple Hyp $H_a : \mu \neq \mu H_0$ population mean is $\neq$ to Hypothesis mean

Composite Hyp
$H_a : \mu > \mu H_0$ > - -
$H_a : \mu < \mu H_0$ < - -

level of Significance → Significance level is the max. value of the probability of rejecting H_0 when it is true and is usually determined in advance before testing the hypotheses