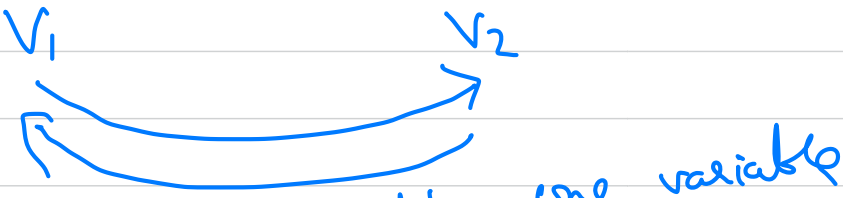
Statistics

Statistics

- Descriptive ⇒ Summarizing
  - Central tendency
    - ↳ Mean
    - ↳ Median
    - ↳ mode
  - Dispersion
    - ↳ standard deviation
    - ↳ Range
    - ↳ Co variance
    - ↳ Quartile (Boxplot)
      - ↳ Q3 − Q1 = Inter Quartile Range.
- Inferential

Co - variance

So Do & variance apply to only one variable whereas Co-variance apply to two variable.



effect b/w one variable on the other.

Co - relation

bounded Co-variance.

It has a range of

-1	0	+1

Inferential Statistics :

Not only understand but also Infer from it [True / False]

Assumption → Hypothesis.

Hypothesis Testing → Will work or not → Two statistics yes / No.

Infer from the Sample about Population

Popular belief / status quo } => Null hypothesis → H_0 (Null)

Ex : For vaccine

H_0 : vaccine has (no) effect on disease.

H_a : vaccine is effective

Popular belief / status quo / No

Whatever you are claiming or testing becomes Alternate hypothesis H_a

D - C - D Approach

↓ → Calculate → Decide

Define hypothesis — Test Statistics — Decide whether to accept or reject H_0

H_0 ⇒ Statement about the population parameters (mean, standard deviation, proportion) which one decide to prove (or) disprove.

Ex of hypothesis

H_0 ⇒ Average height of Adult male is 6 feet
H_a ⇒ Average height of Adult male greater than 6 feet

Three ways of defining hypothesis

Null (H_0)	Alternate (H_a)
$H_0 = M$	$H_a \neq M$
$H_0 \leq M$	$H_a > M$
$H_0 \geq M$	$H_a < M$

Null will always contain equal to (=) sign.

Mostly this followed for practical purpose.

One tile test

Lower tail

Upper tail

One tail test

Two tail test

Level of Significance
Type I & Type II Error.

Judgement hypothesis :

H_0 : Innocent
H_a : Guilty

If Sample is not Representing the population then Statistics will give a wrong Result.

Conclusion	Population Condition: H_0 True	Population Condition: H_0 False
Accept (H_0)	Correct Decision	Type II Error (β)
Reject (H_0)	Type I Error (α)	Correct Decision

Type I Error (α) → False positive : Rejecting H_0 when it is True
Type II Error (β) → False negative : Accepting H_0 when it is false.

More Detrimental

α = Level of Significance = Type I error.
β = Type II error

when σ is known ⇐ ⇒ when σ is not known.

Z test VS T-Test

> Practically T-test is used as if we know σ then we will also know mean.

$\sigma \equiv$ S.D of population

So in T-~~test~~ when we do not know σ we use S i.e. standard deviation of sample is used

> $n = 30$ is considered as a large sample

T-Test : One & Two Sample.

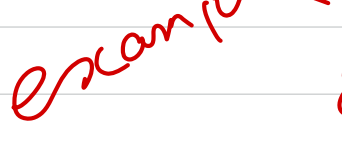
degrees of freedom → Amount of information one has to take the decision

$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

↓

$\leftarrow n - 1$

example



→ five person needs to sit we can find the exact by only 4 people

Hypothesis testing.

Alpha

P value → PLANR

1

P value Less than → Reject Null hypothesis

One Sample → When there is only one variable in hypothesis.

i.e. $H_0 : M = 2.5$

$H_a : m \neq 2.5$

two Sample → When there is two variables in hypothesis

i.e. $H_0 : M_a = M_b \,/\, M_a - M_b = 2.5$ etc

$H_a : m_a \neq M_{b_0} \,/\, m_a - m_b \neq 2.5$ etc.

Independent → (Mismatch possible) Dependent (Matched sample means)

Independent → Male, Female

Dependent → Type, Weight reduction program, After

Analysis of variance

ANOVA :

When there are more than 2 variable in a hypothesis then we will use Anova.

Eg: Mileage of 3 Brands.
M_1 M_2 M_3

Two tail test → $M_1 = M_2$
$M_2 = M_3$
$M_3 = M_1$

Anova → hypothesis testing → more than 2 variables (means in hypothesis)

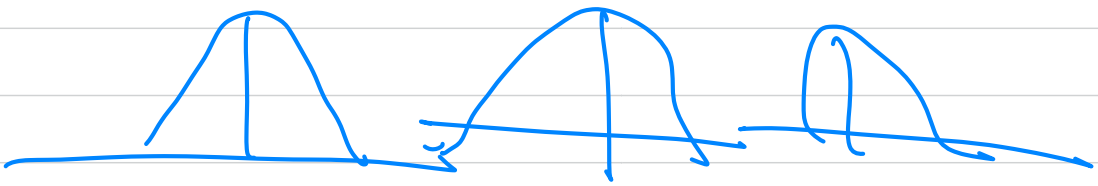
H_0 : $M_A = M_B = M_C$ (Mileage of 3 brands)

H_a : Not all mean are equal { atleast one of the mean is different }

(i.e.

Why ANOVA is called Analysis of variance not Analysis of Mean.

Sample 1 2 3



⇒ within variance
⇒ Between variance

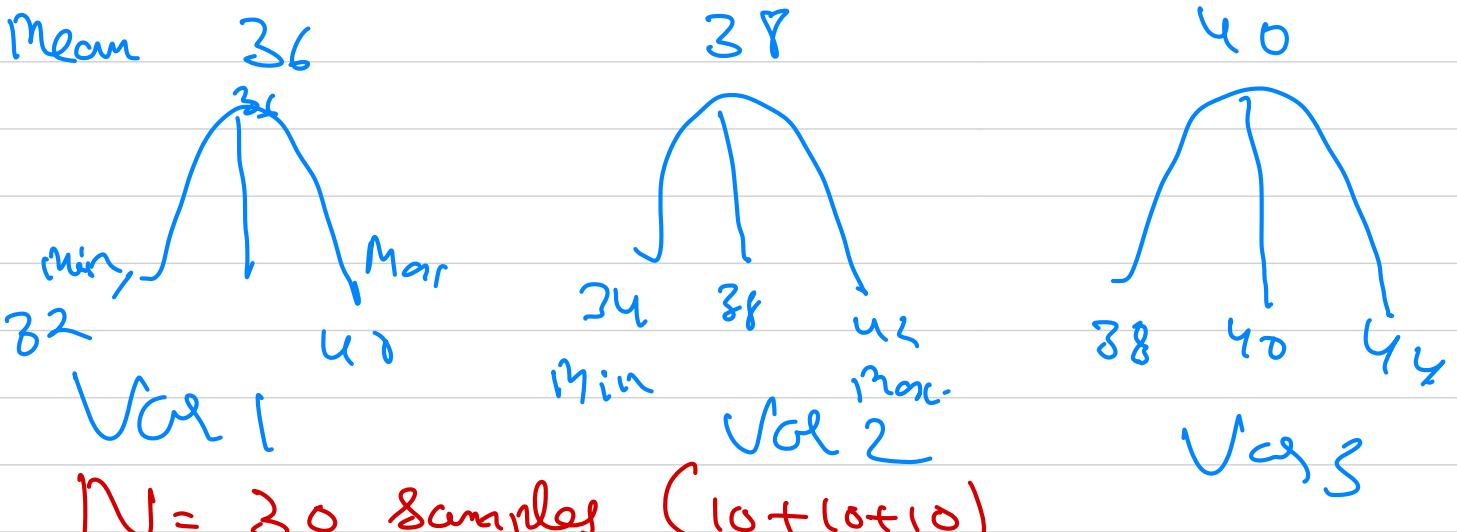
- Calculating the Grand Mean.
- Calculate between variance.

we do F-test for this.

Example :

Mileage of 3 brands.

	A	B	C
Mean	36	38	40
index			
1	32	1	1
2	36	1	1
3	42	1	1
4	38	1	1
5	32	1	1
6			
10			

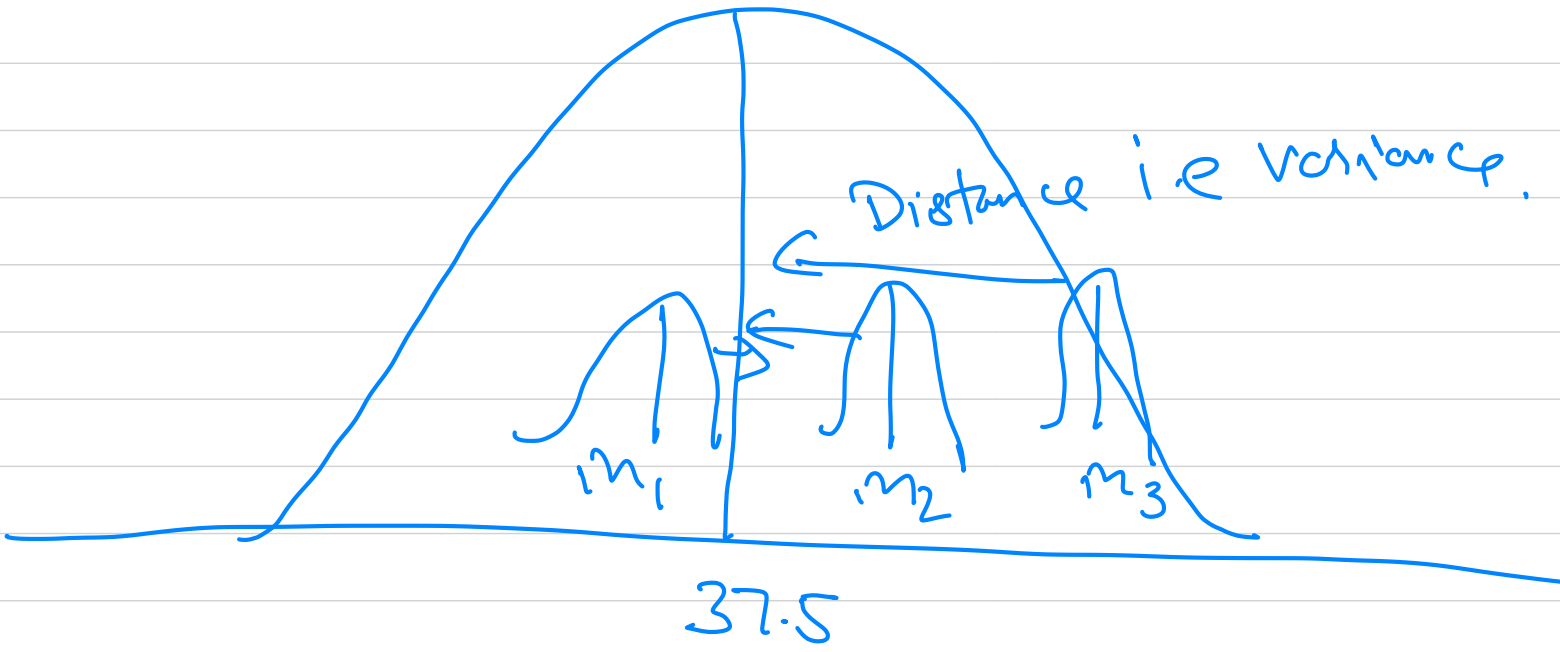


N = 30 samples (10+10+10)

Calculate Grand Mean i.e. of all 30 samples

Grand Mean = 37.5

Calculate grand variation.



So we are calculating the variance

$$F\ test = \frac{\text{Between}}{\text{within}}$$

Between → Grand variance calculation.

within → Individual variance