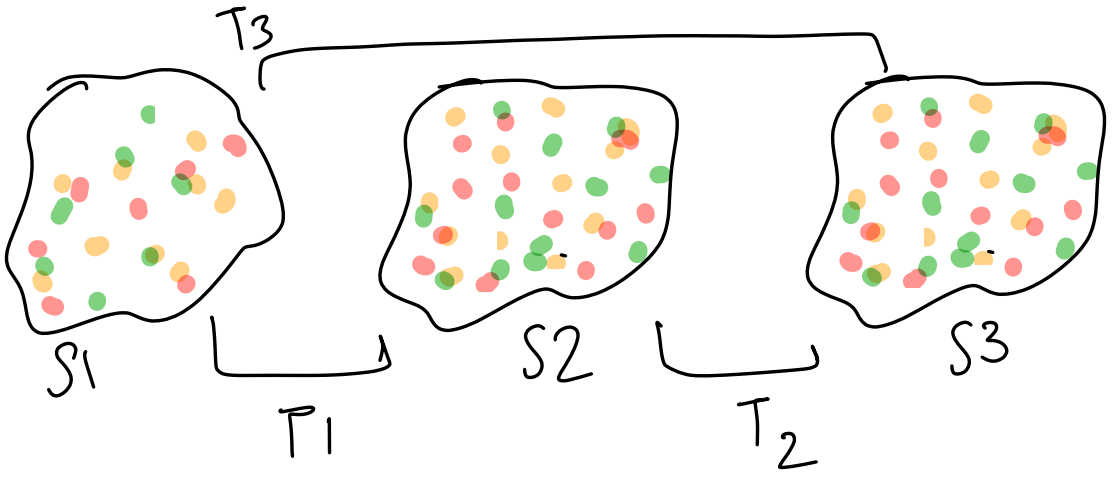
ANOVA

Analysis of Variance

1. Numerical v/s Categorical (2 categories) — TTest
2. Categorical v/s Categorical — Chi squared.
3. Numerical v/s Categorical (>2 categories) — ANOVA
4. Numerical v/s Numerical — Covariance → Correlation; pearson → spearman



3 Sample

$$3C_2 = \frac{3!}{2! \times 1!} = 3$$

10 Sample

$$10C_2 = \frac{10!}{2! \times 8!} = 45$$

(1) Too many Ttest

(2) Error Compounding

Var → low - within

I

1. American Basketball Player (80 inches)
2. Indian Cricket team players (70 inches)
3. Indonesian College Students (64 inches)

Var (high → b/w)

(1) Variance within group

(2) Variance b/w group.

II

Sorting alphabetically

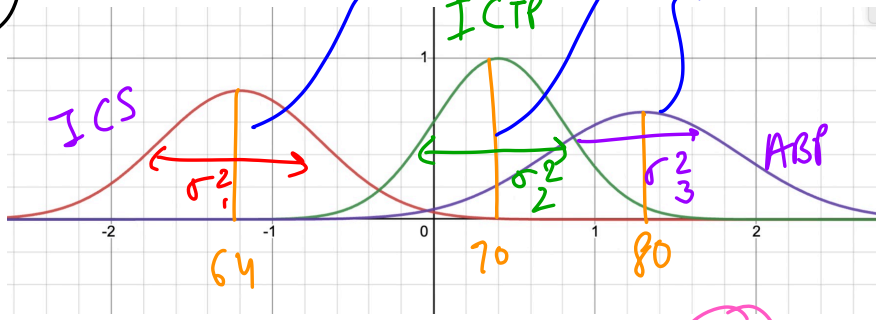
1. A - F
2. G - N
3. O - Z

Var within → high

Var b/w → low

$$F\text{ratio} = \frac{\text{Variance b/w grps}}{\text{Var. within grps.}}$$

(I)



= Variance within grp.

(5)

H_a

$H_0 \Rightarrow \mu_1 = \mu_2 = \mu_3$

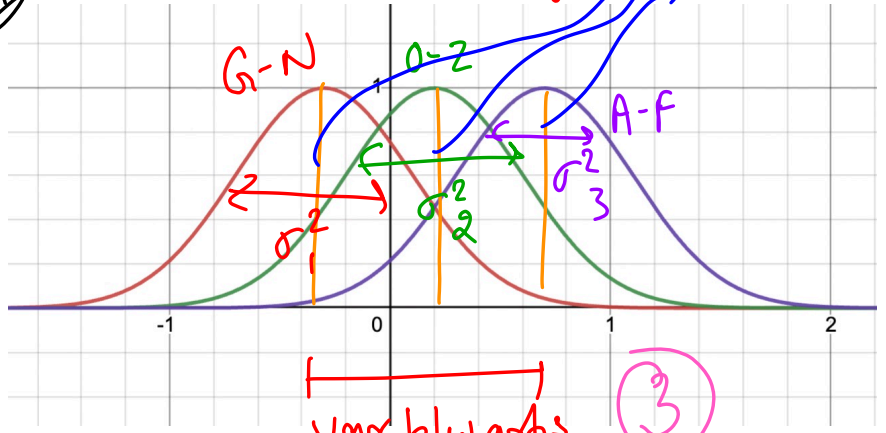
$H_a \Rightarrow \mu_1 \neq \mu_2 \neq \mu_3$

At least one of them is different

(4)

Var b/w grps

(II)



Var b/w grps

(3)

D variance within grp

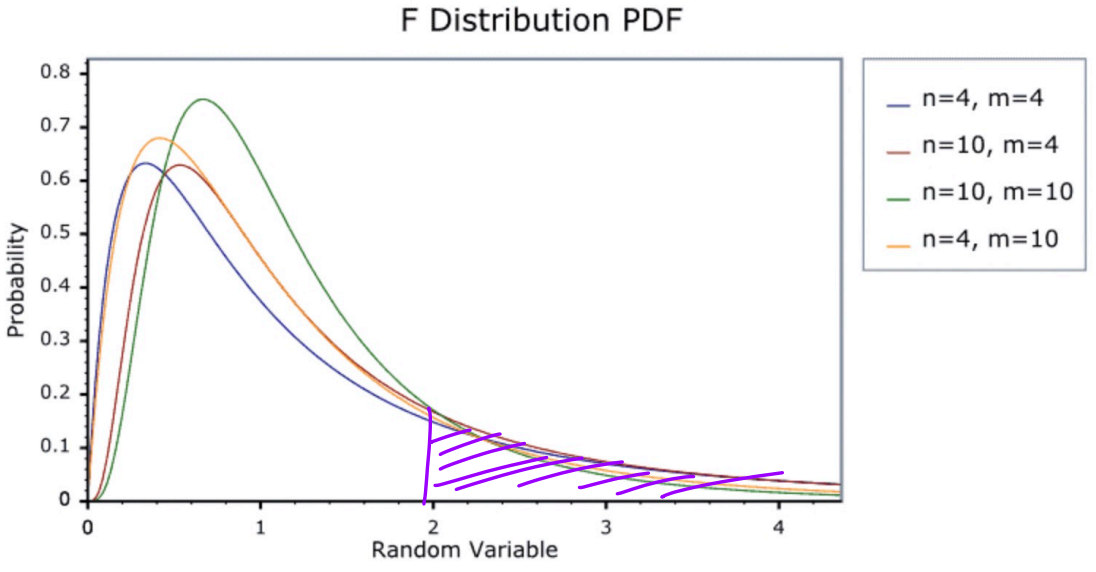
(10)

H_0

F_{ratio}

H_0 : 3/10

H_a : 4/5



Right tailed test

	M	T	W	T	F	S	(x1000)
a :	[25	25	27	30	23	20]	
b :	[30	30	21	24	26	28]	
c :	[18	30	29	29	24	26]	

$\mu_1 = 25$

$\mu_2 = 26.5$

$\mu_3 = 26$

H_0 : No difference

$\mu_1 = \mu_2 = \mu_3$

H_a : Atleast one is diff

$\mu_1 \neq \mu_2 \neq \mu_3$

① Step 1 : Compute Individual means

② Step 2. Compute mean of means of grps.

$$\bar{m} = \frac{\mu_1 + \mu_2 + \mu_3}{3} = 25.83$$

③ Step 3 : Calculate variance b/w grps.

SSB: Sum of Squares b/w

$$\sum n_i (\mu_i - \bar{m})^2$$

n_i → no. of sample

SSB: $6(25-25.83)^2 + 6(26.5-25.83)^2 + 6(26-25.83)^2$

SSB: 7.0002

$DOF_B = 2$

$\bar{m}, \mu_1, \mu_2, \mu_3$

Step 4: MSB → Mean Sum of Squares b/w groups.

$$MSB = \frac{SSB}{DOF_B} = \frac{7.0002}{2} = 3.5001$$

	M	T	W	T	F	S	(x1000)
a :	[25	25	27	30	23	20]	
b :	[30	30	21	24	26	28]	
c :	[18	30	29	29	24	26]	

$\mu_1 = 25$

$\mu_2 = 26.5$

$\mu_3 = 26$

Step 5. Compute Variance within groups.

SSW → Sum of Square within

$(x_i - \mu_i)^2$

$(25-25)^2 + (25-25)^2 + (27-25)^2 \ldots\ldots (20-25)^2 +$ } A

$(30-26.5)^2 + (30-26.5)^2 + (21-26.5)^2 \ldots (28-26.5)^2 +$ } B

$(18-26)^2 + (30-26)^2 + (29-26)^2 \ldots (26-26)^2$ } C

$$SSW = 223.5$$

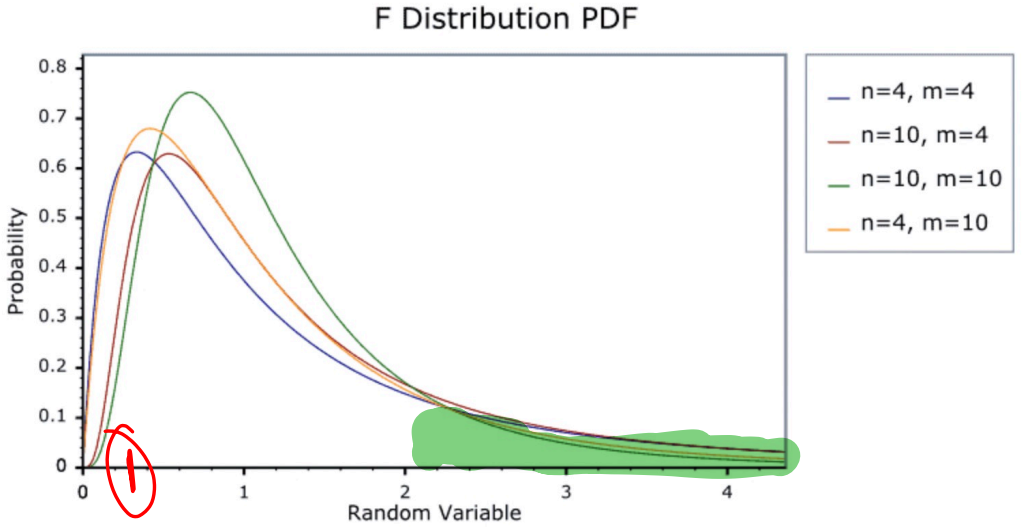
$DOF_W = 15$

$\frac{SSW}{DOF_W} = 14.9$

MSW = Mean Sum of squares within

$$F\text{ratio} = \frac{\text{Var b/w grps}}{\text{Var. within grps}} = \frac{3.5001}{14.9}$$

$F\text{ratio} = 0.2349$



Fratio = 0.2549

pvalue => $1 - f.cdf(Fratio, dfn, dfd)$

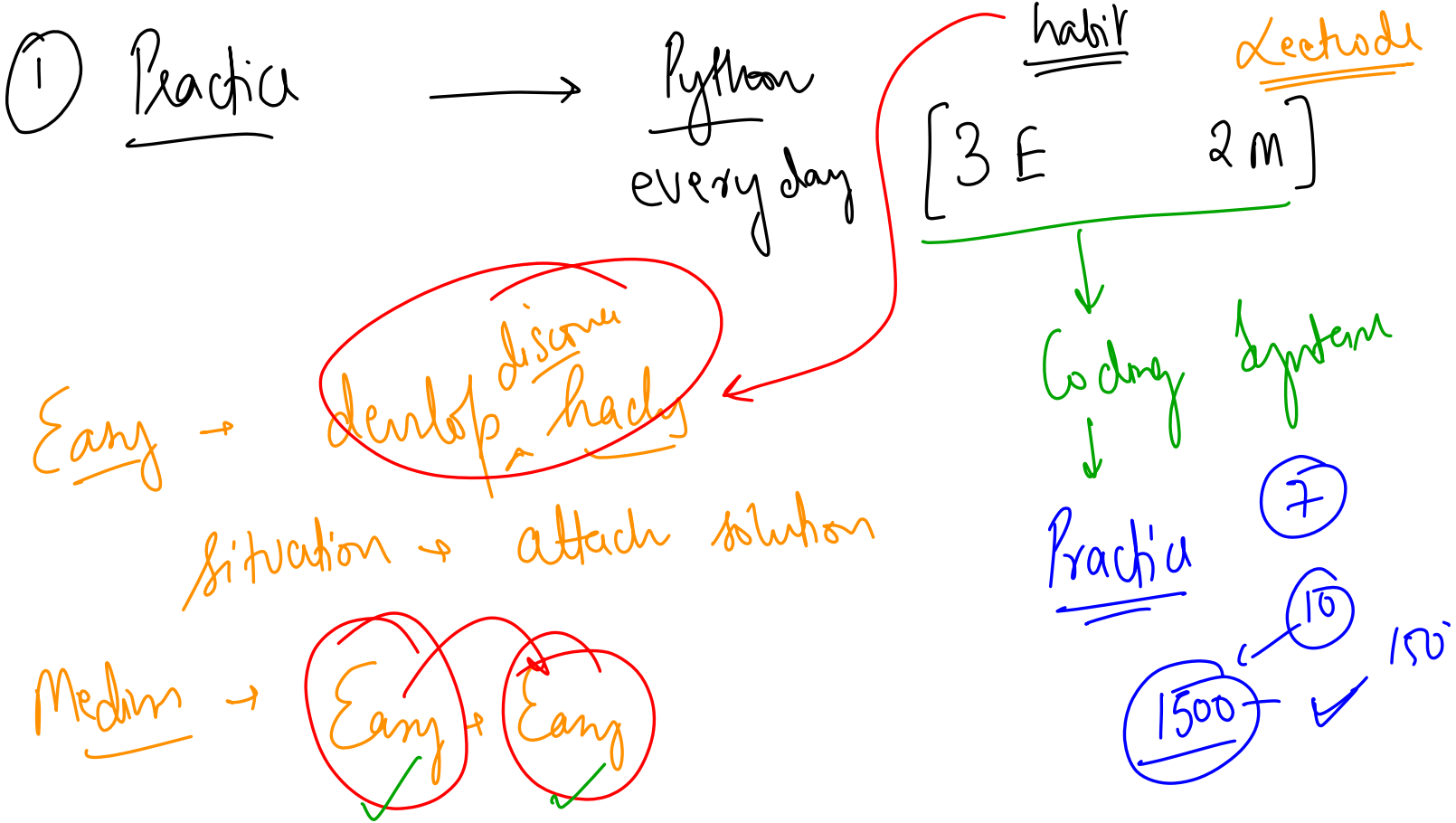
dfn: num
dfd: within

ASSUMPTIONS

1. Data should be Gaussian
   - QQ Plot Visualise
   - Wilkin Shapiro Test
   - KS Test
     Kolmogorov Smirnoff Test
2. Independant
3. Equal variances among diff/grps. → Levene Test

If these assumptions dont hold true

KRUSKAL WALLIS TEST


1 Practice
Python
every day
habit
Lectrodi
[3 E 2 M]
Coding
System
Practice
7
10
1500
150
Easy
develop
discuss
hacks
situation → attach solution
Medium
Easy
Easy