Page No.
Date

ANOVA

- f_d - distribution -
  - ↳ continuous prob. distribution
  - ↳ It has 2 parameters df_1 & df_2
  - ↳ positively skewed
  - ↳ used for testing equality of variance

Suppose -

for $\chi_1^2 \rightarrow df_1$

$\chi_2^2 \rightarrow df_2$

then

$\frac{\chi_1^2/df_1}{\chi_2^2/df_2}$ → follows f-distribution

⇒ one way ANOVA (Analysis of variance)
- ↳ a statistical method used to compare means of 3 or more independent grps
- ↳ It is extension of t-test

one-way ⇒ only 1 independent variable with multiple levels

- Steps - (with example)
  - i) Null hypothesis - all groups mean are equal
  - Alternate " - at least one grp mean is significantly diff

A	B	C
3	1	8
6	8	6
3	9	10

(ii) Calculate grand mean

$\bar{\bar{x}} = \frac{3+6+3+1+8+9+8+6+10}{9}$

$= 6$

iii) Calculate mean of individual-

$\bar{x}_A = 4, \bar{x}_B = 6, \bar{x}_C = 8$

iv) • Calculate SST (Sum of Square total)

$(6-3)^2 + (6-6)^2 + (6-8)^2 + (6-7)^2 + (6-8)^2 + (6-9)^2 + (6-8)^2 + (6-6)^2 + (6-10)^2$

$= 76$

$\rightarrow df = n-1 = 9-1 = 8$

• Calculate SSW (Sum of Square within)

$(4-3)^2 + (4-6)^2 + (4-8)^2 + (6-7)^2 + (6-8)^2 + (6-9)^2 + (8-8)^2 + (8-6)^2 + (8-10)^2$

K = # categ

$SSW = 52$

$df = n-k = 9-3 = 6$

• Calculate SSB

$3 \times (6-4)^2 + 3 \times (6-6)^2 + 3 \times (6-8)^2$

3 → # of elements in group A

(6 − 4) → grand mean

4 → mean of A

$= 24$

$df = k-1 = 2$

$SST = SSW + SSB$

v) Calculate test statistics (f-ratio)

F-ratio = $\frac{SSB}{dfssB} \Big/ \frac{SSW}{dfssw}$ = 1.4

→ p-value ⇒ 0.31

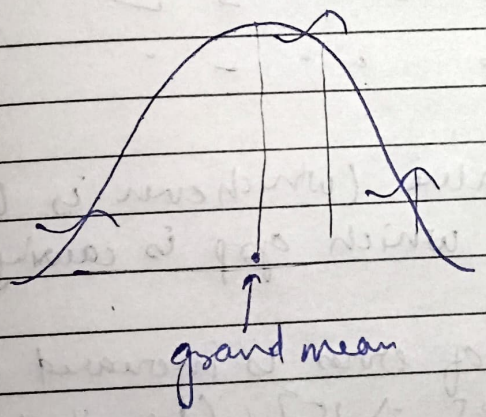
0.31 > 0.05 so can't reject Null hypothesis

. Geometric intution -

$$F = \frac{\frac{SSB}{dfssB} \longrightarrow \chi^2 \text{ dist}}{\frac{SSW}{dfssw} \longrightarrow \chi^2 \text{ dist}}$$

so F follows f- distribution

→ H0 : $\mu_A = \mu_B = \mu_C$ ⇒ all the three groups comes from same populat^n



SSB is dist. b/w grandmean & individual mean

→ F↑ then SSB↑ → Pvalues is small

→ SSW tells about variance of each individual

SSW↑ → F↓

(P-value is area to right in graph)

Assumptions-

- Observations should be independent
- The data within each group should be approx normally distributed
- Homogeneity of variance
  - var of "populn" from each sample should be equal

- Post-hoc-Test =>
  - used in context of ANOVA
  - when we get significant diff in group means then we use it and tell because of which group problem is happening

i) Bonferroni correctn-

A-B → do t-test → Pval ✓

B-C → " → ✓

C-A → " " → "

By seeing p-value (whichever is low) we can tell which grp is causing problem

Problem is Prob of error is increased to 5+5+5 => 15%. (family wise Error rate FWER)

So we divide it by # groups i.e 3

=> 15/3 = 5

2) Tukey's HSD (Honestly Significant Difference) Test

Page No.
Date

- Why t-test is not used for more than 3 categories?
  - Increased Type-I error
  - Difficulty in interpreting result
  - Inefficient