Name - Sagar Powar [ISM Assignment - 2] (1)

Bits ID - 2025AA05421 [Section] - AIML Section - 6

Q.1

let, μ_1 = Mean nicotine content of sample A

a) [Hypotheses] - μ_2 = Mean nicotine content of sample B

(1) Null Hypotheses (H_0) : $\mu_1 = \mu_2$

There is no significant difference between avg. nicotine levels of sample A & B.

(2) Alternative Hypotheses (H_1) : $\mu_1 \neq \mu_2$

There is a significant difference between avg. nicotine levels of sample A & B

b) [Sample Means]

(1) sample A : $\bar{x}_A = \frac{24+27+26+21+25}{5} = \frac{123}{5} = 24.6$ mg

(2) Sample B : $\bar{x}_B = \frac{27+30+28+31+22+36}{6} = \frac{174}{6} = 29.0$ mg

conclusion : Based solely on avegrages, sample B appears to have a higher nicotine content.

c) [Variance & standard deviation]

(1) sample A : $s_A^2 = \frac{\Sigma(x-\bar{x}_A)^2}{n_A-1}$

$= \frac{(24-24.6)^2+(27-24.6)^2+(26-24.6)^2+(21-24.6)^2+(25-24.6)^2}{5-1}$

$= \frac{0.36+5.76+12.96+1.96+0.16}{4} = \frac{21.20}{4} = \boxed{5.30}$

$S_A = \sqrt{S_A^2} = \sqrt{5.30} = 2.30$

Q.1

c) Sample B : $S_B^2 = \frac{\sum (x - \bar{x}_B)^2}{n_B - 1}$

$= \frac{(27-29)^2 + (30-29)^2 + (28-29)^2 + (31-29)^2 + (22-29)^2 + (36-29)^2}{6-1}$

$= \frac{4+1+1+4+49+49}{5} = \frac{108}{5} = \boxed{21.60}$

$S_B = \sqrt{S_B^2} = \sqrt{21.60} = 4.65$

* Why variance is required?

→ (1) t-test compares the difference in means relative to variability

(2) variance measures spread in data (i.e. noise in data)

(3) large variance means more uncertainty, which means harder to detect true differences.

d) Hypothesis Testing

1) Given, $\alpha = 0.05$, $n_1 = 5$, $\bar{x}_1 = 24.6$, $s_1^2 = 5.30$

$n_2 = 6$, $\bar{x}_2 = 29.0$, $s_2^2 = 21.60$

2) $H_0 : \mu_1 = \mu_2$ & $H_1 : \mu_1 \neq \mu_2$

This is two-tailed test at $\alpha = 0.05$.

3) Using two-mean (small sample) t-formula,

$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ from H_0, $\mu_1 - \mu_2 = 0$

$t = \frac{(24.6 - 29.0) - (0)}{\sqrt{\frac{5.30}{5} + \frac{21.60}{6}}} = \frac{-4.4}{\sqrt{1.06 + 3.60}} = \frac{-4.4}{2.16}$

$\boxed{t = -2.04}$

Name- sagar powar
Bits ID - 2025AA05421

$$\boxed{t_{calculated} = |t| = 2.04}$$

(4) Find degree of freedom (df)

As sample size are ~~wrong~~ small & not equal we can use below formula to find df.

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} = \frac{(1.06 + 3.6)^2}{\frac{(1.06)^2}{4} + \frac{(3.6)^2}{5}}$$

$$= \frac{21.72}{0.281 + 2.592} = \frac{21.72}{2.873} = 7.56 \approx 8$$

(5) find critical value ($t_{critical}$)

from t-table, for ($\alpha = 0.05$), two-tailed test, df ≈ 8

$$\boxed{t_{critical} = \pm 2.306}$$

(6) Decision Rule: $|t| = 2.04$ & $t_{critical} = 2.306$

$$|t| < t_{critical}$$

$$\boxed{2.04 < 2.306}$$

$\boxed{\text{Decision}} \Rightarrow \boxed{\text{Fail to reject } H_0}$

i.e. There is insufficient stastistical evidence to conclude that mean nicotine content of Sample A & B differs.

Hence, the two samples A & B may be considered to come from populations with the same mean nicotine content.

Name - Sagar Powar

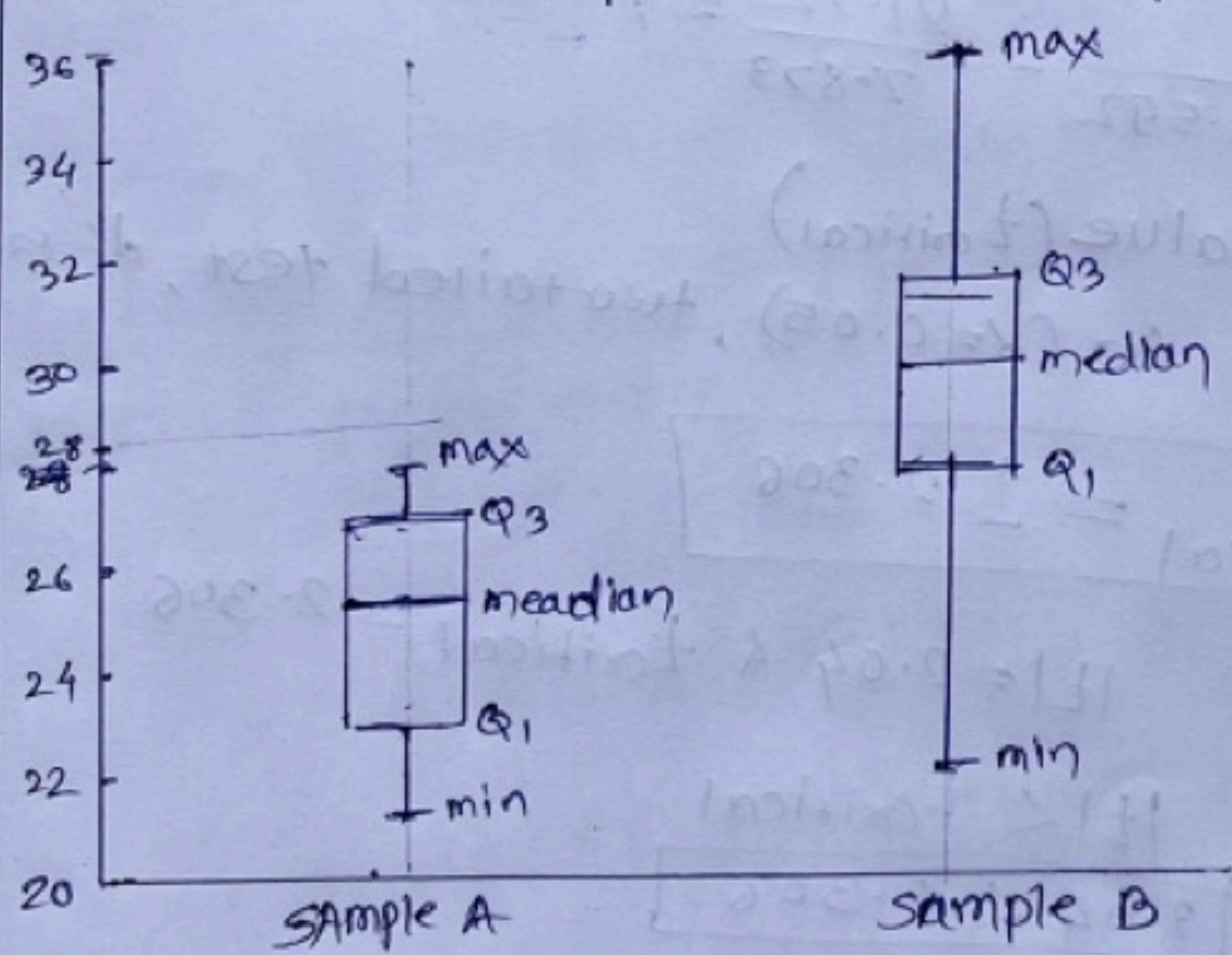
Bits ID - 2025AA05421 AIML Section - 6 (4)

Q.1

e) Side-by-side Box plots

Five-Number Summary

Measure	Sample A	Sample B
1) Min	21	22
2) Q_1	22.5	27
3) Median	25	29
4) Q_3	26.5	31
5) Max	27	36



Conclusion:

① Sample B has higher median (29) than sample A (25)

② Spread / Dispersion

a) Both samples have equal IQR (4)

b) Sample B has much larger range.

A range = 27 - 21 = 6

B range = 36 - 22 = 14

c) Sample B shows greater overall dispersion.

Name - sagar powar

Bits ID - 2025AA05421

$\boxed{\text{Section-6}}$

Q.2 Given, $k = 3$ groups, $N = 15$ (total observations)

a) $\boxed{\text{Hypothesis}}$

① H_0: All population means are equal

i.e. $\boxed{H_0 : \mu_A = \mu_B = \mu_C}$

② Alternative Hypothesis: H_1 : At least one group μ is different

$\boxed{H_1 : \text{At least one } \mu \text{ differs}}$

b) Degrees of freedom

i) Between Groups → $df_{bet} = k - 1 = 3 - 1 = 2$

ii) within groups → $df_{within} = N - k = 15 - 3 = 12$

c) calculate MSB, MSW, F

(I) Find group means.

$$\bar{X}_A = \frac{35+38+32+34+36}{5} = \frac{175}{5} = 35$$

$$\bar{X}_B = \frac{40+42+38+45+40}{5} = \frac{205}{5} = 41$$

$$\bar{X}_C = \frac{42+48+44+46+45}{5} = \frac{225}{5} = 45$$

(II) Grand mean

$$\bar{X}_{GM} = \frac{175+205+225}{15} = 40.33$$

(III) SSB (sum of squares Between)

$$SSB = \sum n_i (\bar{X}_i - \bar{X}_{GM})^2$$

(A) Group: $5(35-40.33)^2 = 5(28.44) = 142.22$

B Group: $5(41-40.33)^2 = 5(0.44) = 2.22$

C Group: $5(45-40.33)^2 = 5(21.78) = 108.89$

Name- Sagar Powar

Bits ID - 2025AA05421 $\boxed{\text{Section-6}}$

$$SSB = 142.22 + 2.22 + 108.89 = 253.33$$

(iv) $\underline{MSB} = \frac{SSB}{df_{bet.}} = \frac{253.33}{2} = 126.67$ — (1)

(v) Find SSW (Sum of squares within)

$$SSW = \sum (X_{ij} - \bar{X}_i)^2$$

(A) Group: $(35-35)^2 + (38-35)^2 + (32-35)^2 + (34-35)^2 + (36-35)^2$

$= 0 + 9 + 9 + 1 + 1 = 20$

(B) Group: $(40-41)^2 + (42-41)^2 + (38-41)^2 + (45-41)^2 + (40-41)^2$

$= 1 + 1 + 9 + 16 + 1 = 28$

(C) Group: $(42-45)^2 + (48-45)^2 + (44-45)^2 + (46-45)^2 + (45-45)^2$

$= 9 + 9 + 1 + 1 + 0 = 20$

Total $SSW = 20 + 28 + 20 = 68$

(vi) MSW (Mean square within) — (2)

$$MSW = \frac{SSW}{df_{within}} = \frac{68}{12} = 5.67$$

(vii) F-statistic

$$\boxed{F = \frac{MSB}{MSW} = \frac{126.67}{5.67}} = \boxed{22.34}$$

Name- Sagar Powar
Bits ID - 2025AA05421

Section-6

(9)

Q.2

d) Conclusion -

$F_{calculated} = 22.34 > F_{critical}$ (for 5% or 1%)

→ F value is far greater than critical F.

Decision ⇒ Reject H_0

i) Does this prove one method is Best?
⟶ NO. (ANOVA) cannot identify which is best method.

ii) It indicates that at least one study method differs significantly from the others.

e) Effect of increasing sample size to 60 per group.

The reliability of the F-test increases significantly due to below reasons.

① lower MSW (less random error)

② High statistical power - (N=180)

③ More precise mean estimates.

④ Greater ability to detect true differences.

Conclusion -

F-test becomes more reliable and more powerful