Assignment-3

Q1 Region A: [10, 15, 12, 8, 14]

Region B: [18, 20, 16, 22, 25]

mean of Region A = $\frac{10+18+12+18+14}{5} = \frac{69}{5} \simeq 14$

mean of Region B = $\frac{18+20+16+22+25}{5} = \frac{101}{5} = 20.2$

Q2 Satisfaction level [4, 5, 2, 3, 5, 4, 3, 2, 4, 5]

Occurred

mode → 4 : 3 → 4 & 5 are modes.

5 : 3

2 : 2

3 : 2

Q3 Median of the Salary

(i) sort the arrays

mean ⇒ 5750 ⇒ median

Department A: [5000, 5500, 6000, 7000]

Department B: [4500, 5200, 5500, 5800, 6000]

middle term :- median

Q4 range would be the lowest & largest values of the list

hence :- Range ⇒ (24.8, 26.1)

Q5 T-test

Group A: [85, 90, 92, 88, 91]

Group B: [82, 88, 90, 86, 87]

$H_0 \Rightarrow \mu_1 = \mu_2$

$H_1 \Rightarrow \mu_1 \neq \mu_2$

T test $\Rightarrow \dfrac{\mu_1 - \mu_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

$\mu_1 = 89.2 \qquad \mu_2 = 86.6$

$n_1 = n_2 = 5$

$\sigma_1 = \dfrac{\Sigma(x_i - \bar{x})^2}{n-1} \Rightarrow \dfrac{(4.2)^2 + (0.8)^2 + (1.2)^2 + (1.8)^2}{4} = 5.74$

$\sigma_2 = \dfrac{(4.6)^2 + (1.4)^2 + (3.4)^2 + (0.6)^2 + (0.4)^2}{4} = 8.8$

$T\text{-test} = \dfrac{89.2 - 86.6}{\sqrt{\dfrac{5.74}{5} + \dfrac{8.8}{5}}} \Rightarrow 1.54$

By table 0.741 < T calculated

So, reject null hypothesis.

$H_1 \rightarrow$ Yes there is a significant difference in the mean scores b/w two groups.

Q6

Advertising Expenditure (in thousands) : [10, 15, 12, 8, 14]

Sales (in thousands) : [25, 30, 28, 20, 26]

Correlation Coefficient $= \dfrac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}$

$n = 5$, $\Sigma x = 59$, $\Sigma y = 129$, $\Sigma xy = 1236$, $\Sigma x^2 = 729$

$\Sigma y^2 = 3385$

Correlation coeff $= -0.03$

Scanned with CamScanner

Q7

Height [160, 170, 165, 185, 175, 180, 170]

~~After~~ → ~~SD of~~

mean = 167.86

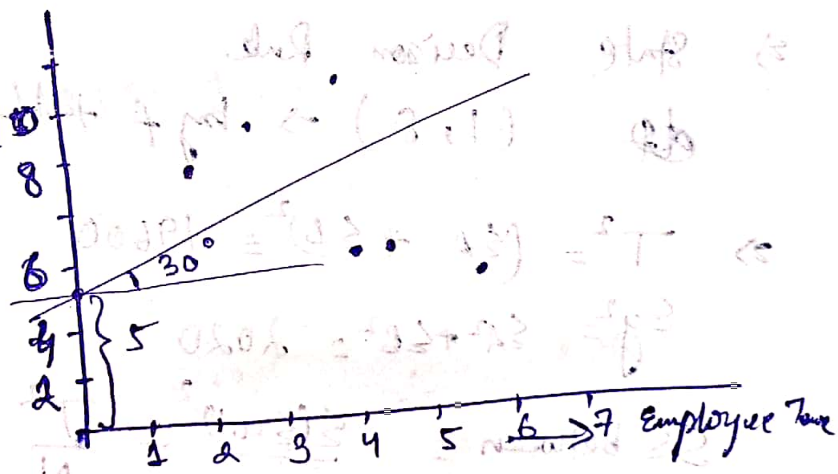
$$SD = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{(7.86)^2 + (2.14)^2 + (2.86)^2 + (12.86)^2 + (7.14)^2 + (12.14)^2 + (2.14)^2}{6}}$$

$\boxed{SD = 8.59}$

Q8 Employee Tenure (in years) [2, 3, 5, 4, 6, 2, 4]

Job satisfaction :- (on 10 scale) [7, 8, 6, 9, 5, 7, 6]

On a rough graph



$y = mx + c$

$m = \tan 30°$

$c = 5$

$y = \frac{x}{2} + 5$

$2y = x + 10$

$\boxed{x - 2y + 10 = 0}$

where x is Employee Tenure & y is Job Satisfaction.

Q5 ANOVA

Medication A: [10, 12, 14, 11, 13]
Medication B: [15, 17, 16, 14, 18].

A	B
10	15
12	17
14	16
11	14
13	18

$H_0 = \overset{med}{A} = \overset{med}{B}$

$\alpha = 0.05$ CI = 95%

$N = 10$, $n = 5$, $a = 2$

$\Sigma A = 60$

$\Sigma B = 80$

⇒ Degree of freedom

df between = $a-1$ = $2-1$ = 1

df within = $N-a$ = $10-2$ = 8

Total = $N-1$ = $10-1$ = 9

⇒ State Decision Rule.

df (1, 8) → by f table $\boxed{5.318}$

⇒ $T^2 = (\Sigma A + \Sigma B)^2 = 19600$, $\Sigma(\Sigma a_i)^2 = (60)^2 + (80)^2 = 10000$

$\Sigma y^2 = \Sigma A^2 + \Sigma B^2 = 2020$

$SS \text{ between} = \frac{\Sigma(\Sigma a_i)^2}{n} - \frac{T^2}{N} = \frac{10000}{5} - \frac{19600}{10}$

SS between = 40

$SS \text{ within} = \Sigma y^2 - \frac{\Sigma(\Sigma a_i)^2}{n} = 2020 - \frac{10000}{5}$

SS within = 20

Calculate f test

	SS	df	MS (SS/df)	F
Between	40	1	40	$\frac{40}{2.5} = 16 > 5.318$
Within	20	8	2.5	
Total	60	9		

∴ Reject H_0

medication A ≠ medication B

Q 10

ratings :- [8, 9, 7, 6, 8, 10, 9, 8, 7, 8]

Value

Sort ratings → [6, 7, 7, 8, 8, 8, 8, 9, 9, 10]

$$\text{Value} = \frac{\text{Percentile}(n+1)}{100}$$

$$= \frac{75}{100} \times 11 \rightarrow 8.25 \text{ Index} \approx 8$$

at 75 percentile (9) rating exists.

Q 11 weights [10.2, 9.8, 10.0, 10.5, 10.3, 10.1]

H_0 $\mu_0 = 10gm$

H_1 $\mu_0 \neq 10gm$

$\bar{x} = 10.15$

$$S = \sqrt{\frac{\Sigma(\bar{x}-x_i)^2}{n-1}} = 0.243$$

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

$$t = \frac{10.15 - 10}{\frac{0.243}{\sqrt{6}}} \Rightarrow ~~1.55~~\ 1.51$$

at $\alpha = 0.05$, df = 5

$t = 2.015$

t table > t calculate ∴ accept H_0 ∴ reject H_0

Here :- The mean weight is equal to 10 gm.

Q_{12} Design A : [100 , 120, 110, 90, 95]
Design B : [80, 85, 90, 95, 100]

$H_0 \Rightarrow A = B$
$H_1 = A \neq B$

Chi square test

$$x^2 = \frac{\Sigma(f_o - f_e)^2}{f_e} = \frac{(100-80)^2}{80} + \frac{(120-85)^2}{85} + \frac{(110-90)^2}{90} + \frac{(90-95)^2}{95} + \frac{(95-100)^2}{100}$$

$x^2_{Cal} = 21.036$

$df = n-1 = 4 \quad \alpha = 0.05$

$x^2_{Table} = 9.488$

$x^2_{Cal} > x^2_{Table}$ ∴ Reject Null hypothesis.
accept H_1

Hence there is a slight difference in the click through rates between the two designs.

Q_{13} [7, 9, 6, 8, 10, 7, 8, 9, 7, 8]

at 95% CI the mean satisfaction score.

from table at 95% CI $\alpha = 0.05$ & $df = 9$

t-score · 1.833

$$t\,score = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = 1.833$$

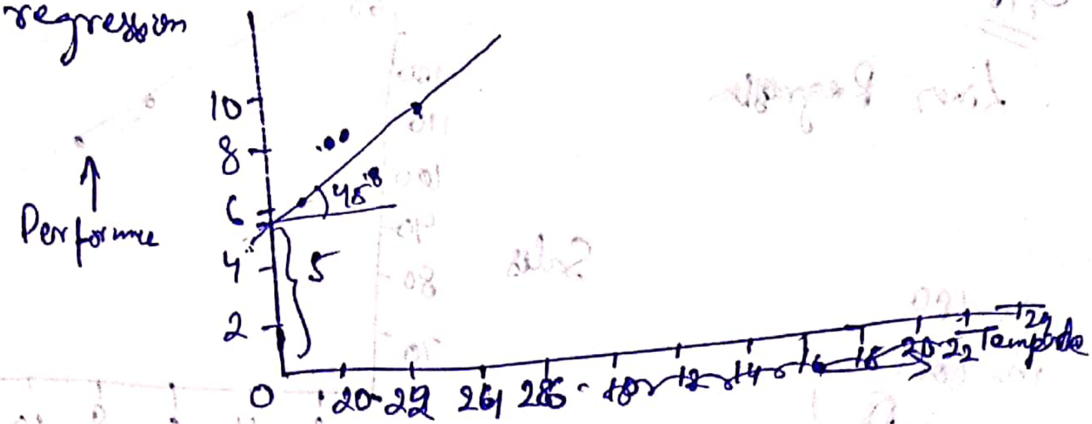
$\bar{x} = 7.9$
$S = 1.12$

$$\Rightarrow \frac{7.9 - \mu}{\frac{1.12}{\sqrt{10}}} = 1.833$$

$\boxed{\mu = 7.25}$ → at 95% CI

Q14

Linear regression



$y = mx + c$

$m = \tan 45° = \frac{1}{\sqrt{2}}$

$c = 5$

$y = \frac{x}{\sqrt{2}} + 5$

$\sqrt{2}y = x + 5\sqrt{2}$

$\boxed{x - y\sqrt{2} + 5\sqrt{2} = 0}$

Q15

[25, 30, 35, 40, 45, 50, 55, 60, 65, 70]

Calculate IQR

$Q_1 = \frac{25}{100} \times (n+1) = \frac{25}{100} \times 11 = 2.75 \simeq 3^{rd}$ Index $= 35$

Q_3 $\frac{75}{100} \times 11 = 8.25 =$ 8 index $= 60$

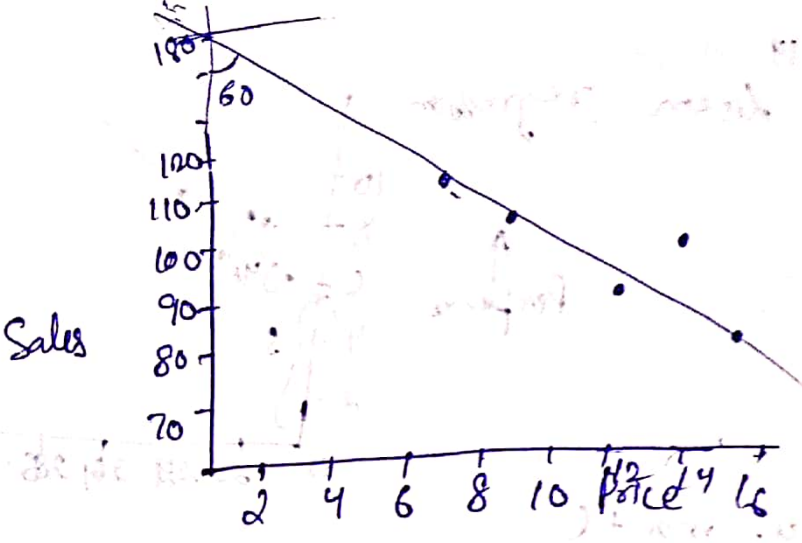
$\boxed{IQR = 60 - 35 = 25}$

Lower fence $= Q_1 - 1.5(IQR) = 35 - 1.5(25) = -2.5$

Higher fence $= Q_3 + 1.5(IQR) = 97.5$

Q18

Linear Regression



$C = 180$

$m = \cos 60$

$= \frac{1}{2}$

$y = mx + c = \frac{1}{2}x + 180$

$x \Rightarrow$ Price

$y =$ Sales.

~~$2x + 4y + 360$~~ $\boxed{x - 2y + 360 = 0}$

Q19 $[7, 8, 9, 6, 8, 7, 9, 7, 8, 7)$

mean: $\bar{x} = 7.6$

Standard error $= \sqrt{\frac{\Sigma(\bar{x} - x_i)^2}{n-1}} = 0.966 \simeq 0.97.$

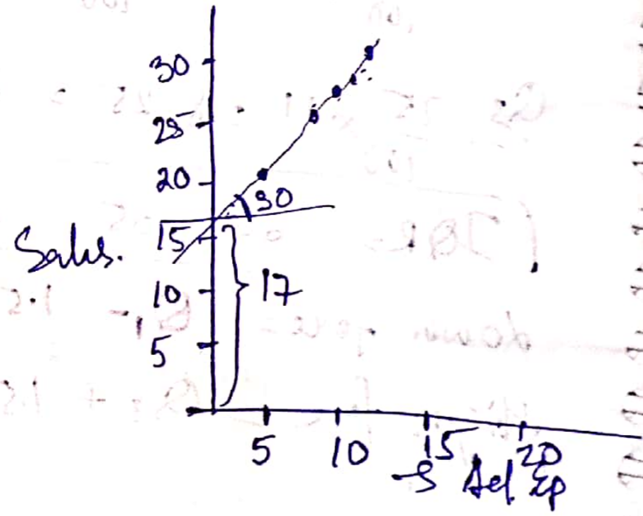
Q20 Advertising Expenditure $= [10, 15, 12, 8, 14]$

$x \leftarrow$ Advertising

$y \leftarrow$ Sales : $[25, 30, 28, 20, 26]$

regression

Predict Sales : y



$c = 17$

$m = \sin 30 = \frac{1}{2}$

$y = \frac{x}{2} + 17$

$$\boxed{\text{Sales} = \frac{(\text{Adv. Exp}) + 34}{2}}$$