G.T = 60

Probability & Statistics

Spring '25

Rubric $\left[\overset{①}{12} + \overset{\text{CLO } ②}{8} + \overset{③}{30} \right]$

CLO 1: (12)

Question 1:

Since the data is divided in data sheet 1 (given), so first combining the data points with respect to both required categories (No. of Atm transactions: ≤ 10 & > 10) and their Account balance.

No of Atms Transactions >10	Account Balance	No of Atms transaction ≤ 10	Account Balance
13	17 ✓	9	11 ✓
14	16 ✓	10	15 ✓
17	19 ✓	10	18 ✓
20	25 ✓	6	9 ✓
		10	15 ✓
		5	6 ✓

Data Display ②

To check the relative variation in account balances b/w users who make more than 10 ATM transactions in a month and those who make 10 or fewer transactions, we need to calculate mean and standard deviation of Account Balance of Both categories (>10, ≤ 10)

Taking Account Balance for >10 as 'A'

Taking Account Balance for ≤ 10 as 'B'

No. of ATM transactions >10 ①

Mean $= \bar{X}_A = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{17+16+19+25}{4} = \boxed{19.2500}$

②

var $= s_A^2 = \frac{\sum_{i=1}^{n} (x_i-\bar{x}_A)^2}{n-1} = \frac{48.7500}{4-1} = \boxed{16.2500}$

OR

SD $= s_A = \sqrt{16.2500} = \boxed{4.0311}$

No. of ATM transactions ≤ 10 ①

Mean $= \bar{X}_B = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{11+15+18+9+15+6}{6} = \boxed{12.3333}$

var $= s_B^2 = \frac{\sum_{i=1}^{n} (x_i-\bar{x}_B)^2}{n-1} = \frac{99.3333}{6-1} = \boxed{19.8667}$

OR ②

SD $= s_B = \sqrt{19.8667} = \boxed{4.4572}$

CV for Group A(Account Bal. with >10 ATM transactions)

$CV_1 = \frac{s_A}{\bar{X}_A} \times 100 = \frac{4.0311}{19.2500} \times 100 = \boxed{20.94\%}$ ①

CV for groupB (Account Bal. with ≤ 10 ATM transaction

$CV_2 = \frac{s_B}{\bar{X}_B} \times 100 = \frac{4.4572}{12.3333} \times 100 = \boxed{36.14\%}$ ①

COMMENT ② with Reason Case Conclusion

Customers who made more than 10 ATM transactions had a lower CV (20.94%) indicating more consistent and stable account balances compared to those who made 10 or fewer transactions, who had high CV

2- (36.14%). This suggests that frequent ATM users tend to maintain Accoud steadi

3-

CLO 2: ⑧

Question 2(a): 5

Average = $\lambda = 3$ (Per 10 seconds).

$P(X \geq 3) = ?$ for 5-second Interval

using Poisson distribution, the Pmf is as follows:

$$P(X=x) = \frac{e^{-\lambda t}\lambda t^{x}}{x!} \qquad \text{or } f(x) = \frac{e^{-\mu}\mu^{x}}{x!} \qquad x = 0,1,2\ldots,$$

Average for per 5-seconds $= 3 \times \frac{1}{2} = 1.5$

$$P(X \geq 3) = 1 - P(X < 3)$$

$$= 1 - \left[P(X=0) + P(X=1) + P(X=2)\right]$$

$$= 1 - \left[\frac{e^{-1.5}(1.5)^{0}}{0!} + \frac{e^{-1.5}(1.5)^{1}}{1!} + \frac{e^{-1.5}(1.5)^{2}}{2!}\right]$$

① ① ①

$$= 1 - (0.2231 + 0.3347 + 0.2510)$$

$$= \underline{1 - 0.8088} = \boxed{0.1912}$$

① ①

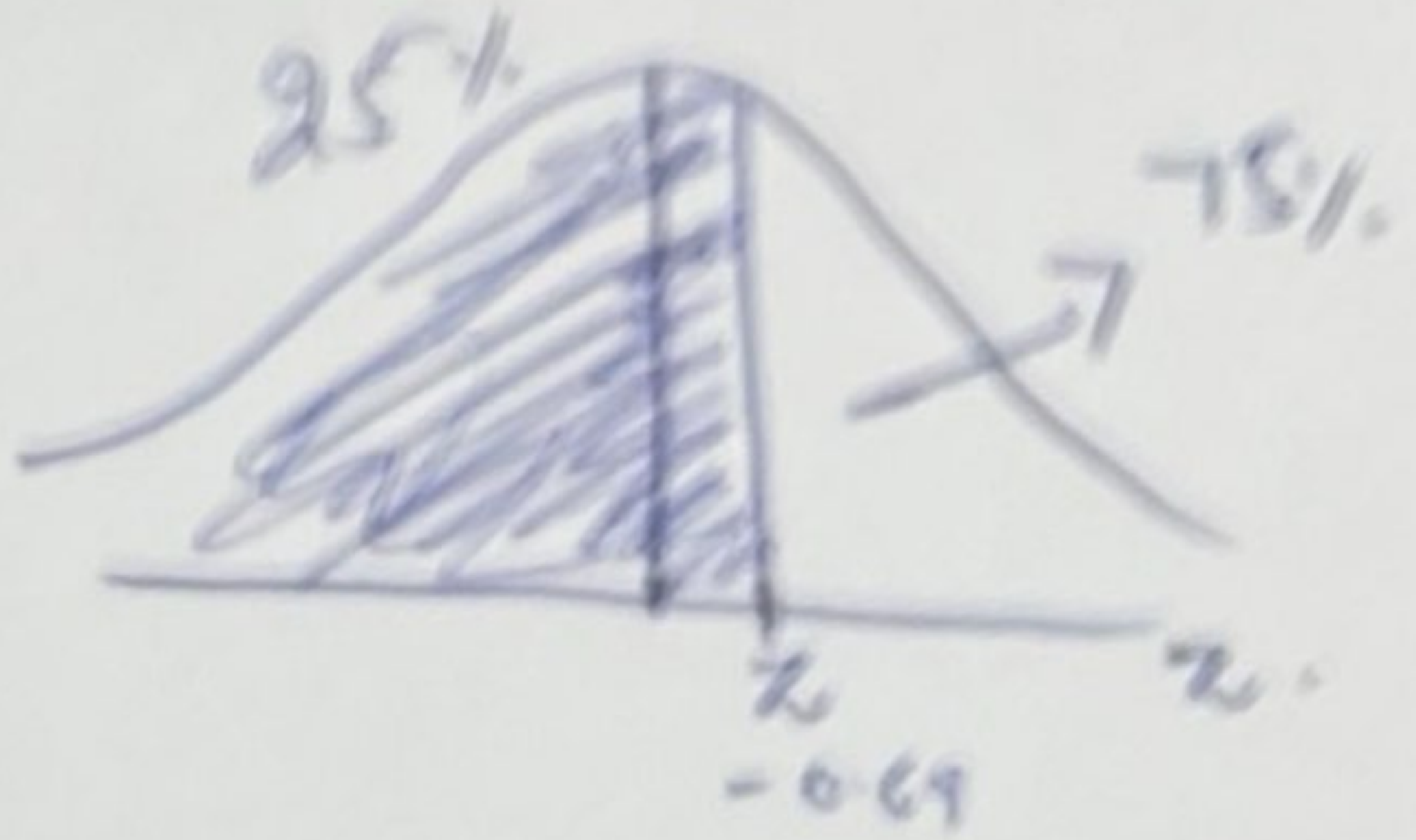
CLO2:

Question 2 (b): 3

mean = 120 (mins) , sd = 15 (mins)

$\Rightarrow P(X < x) = 0.25 = P(Z < z)$

$P(Z < -0.67) = 0.25$



To find out cutoff time:

$X = \mu + \sigma z$

$= 120 + 15(-0.67)$

$= 120 - 10.20$ ①

$= \boxed{109.95}$ mins. ①

① $z = -0.67$ most nearest table value.

CLO 3:

Question 3: 10

$n = 12$, $\bar{X} = 41.5$ mins, $s = 1.784$

1) $H_0: \mu \geq 43$ ①
   $H_1: \mu < 43$ ①

2) $\alpha = 5\% = 0.05$ ①

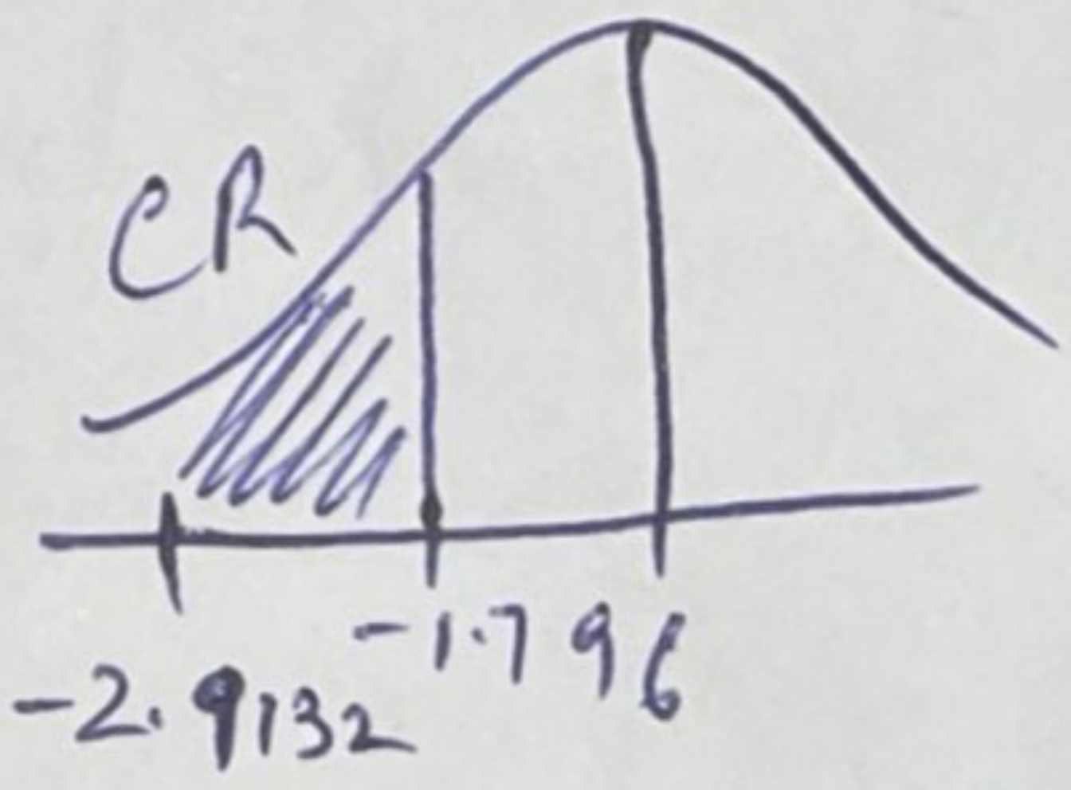
3) Test-Statistics: $t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$ ① with $n-1 = 11$ df.

4) Critical Region:

Reject Ho if $t \leq -t_{\alpha(v)}$ ①

$t \leq -t_{(0.05, 11)}$

$t \leq -1.796$ ①



5) Calculation:

$t = \frac{41.5 - 43}{1.784/\sqrt{12}}$ ① $= \frac{-1.5}{0.5149} = -2.9132$ ①

6- Conclusion:

Since calculated value of t lies in critical Region, so we reject H_0 and ② conclude that the new System has reduced the mean training time per epoch compared to the historical benchmark.

CLO 3: (30)

Question 4a 5

a- To calculate correlation b/w data size and Program efficiency, $\bar{x} = 9.3$ $\bar{y} = 34.1$

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left[n\sum x^2 - (\sum x)^2\right]\left[n\sum y^2 - (\sum y)^2\right]}}$$ (any relevant Formula)

② $$= \frac{10(2781) - (93)(341)}{\sqrt{\left[10(961) - (93)^2\right]\left[10(13707) - (341)^2\right]}} = \frac{-3903}{\sqrt{(961)(20789)}}$$

$$= \frac{-3903}{4469.70} = \boxed{-0.87}$$ ① $r = -0.8732$

② There is strong negative correlation between Data size and Program efficiency assessed through Processed requests. In other words, As data size increases, Processed requests decrease (InDirect relation).
Inverse

direction & strength

b- from the given information, the moded is as follows

$\hat{Y} = 71.87 - 4.06X$

To fill the anova table, we need to calculate Sum of squares for Total, Regression & Error.

$\Rightarrow SST = \Sigma(Y - \hat{Y})^2 = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n} = 13707 - 11628.1$

$= \boxed{2078.9}$

$\Rightarrow SSE = \Sigma(Y - \hat{Y})^2 = \Sigma y^2 - a\Sigma y - b\Sigma xy = 13707 - 71.87(341) + 4.06(2781)$

$= \boxed{493.994}$

$\Rightarrow SSR = SST - SSE = 2078.9 - 493.994 = \boxed{1588.71}$

Q4b 25

① $\hat{Y} = 71.8710 - 4.0614x$

1) **HYPOTHESES:**

H_0: *Overall regression model is insignificant* ①
H_1: *Overall regression model is significant* ①

2) **LEVEL OF SIGNIFICANCE: $\alpha = 0.05$** ①

3) **TEST STATISTIC:** $F = \frac{MSR}{MSE}$

4) **CALCULATION** ②

② $\Sigma(\hat{y}-\bar{y})^2$ ② $\Sigma(y-\bar{y})^2$ ② $\Sigma(y-\hat{y})^2$

	x	y	$\hat{y}$	SSR	SST	SSE
	6	40	47.50	179.63	34.81	56.29
	7	55	43.44	87.26	436.81	133.61
	7	50	43.44	87.26	252.81	43.02
	8	41	39.38	27.88	47.61	2.63
	10	17	31.26	8.08	292.41	203.26
	10	26	31.26	8.08	65.61	27.64
	15	16	10.95	535.92	327.61	25.50
	13	20	19.07	225.82	198.81	0.86
	12	24	23.13	120.25	102.01	0.75
	5	52	51.56	304.99	320.41	0.19
Sums	93	341		1585.16	2078.9	493.74

$\bar{X} = 9.3$

$\bar{Y} = 34.1$

①

Working may varry so mark any relevant working accordingly with Total = 10

SOV	*df*	*SS*	*MS*	*F-Ratio*	*p-value*	*Decision*
Regression	1	1585.16	1585.16	25.684	0.0009674	Reject Ho
Error	8	493.74	61.72			
Total	9	2078.9	-			

⑩ each entry one

5) **CRITICAL REGION**

REJECT H_0 IF $p \leq \alpha$ ①

6) **CONCLUSION**

As p value is less than α, so we reject Ho and conclude that overall regression model is significant. ①