PROBABILITY ASSIGNMENT :

Q.1

a] $\Sigma xy = 182560$ $\Sigma x = 850$ $\Sigma x^2 = 9200$ $\Sigma y = 1737$

• $S_{xy} = \Sigma xy - n \cdot \bar{x}\bar{y} = \cancel{18} \; 34,915$

$S_{xx} = \Sigma x^2 - n(\bar{x})^2 = 20250$

$\therefore \hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{34915}{20250} = 1.724198.$

$\hat{\alpha} = -\hat{\beta}\bar{x} + \bar{y} \rightarrow \bar{y} = 173.7$; $\bar{x} = 85.$

$\boxed{\hat{\alpha} = 27.14317}$

$\rightarrow \because$ Eqn: $\hat{y} = 27.14317 + 1.724198x$

b.] Gradient of regression line refers to slope parameter
As $\hat{\beta}$ is significantly different from 0, there exists a positive relationship between X and Y.

Q.2 X — Share price of DELL   Y — Share price of IBM

$S_{xx} : \Sigma x^2 - n \cdot \bar{x}^2 = \cancel{875.16}$ 301.66

$S_{xy} \rightarrow \Sigma xy - n \cdot \bar{x}\bar{y} = 875.16$

$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = 2.90115$ | $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 3.74809.$

$\therefore \hat{y} = 3.74809 + 2.90115x.$

ii.) Assumption: Distribution of x and y are normal...
Explanatory variable values are independent of each other.

95% CI: $\hat{\beta} \pm t_{2.5\%, 10} \cdot \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$

$\hat{\sigma}^2 : \frac{\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}} \right)}{n-2}$ $\therefore$ $S_{yy} \rightarrow 6409.7125$

$\therefore \boxed{\hat{\sigma}^2 = 387.07447}$

95% CI: 2.90115 ± 2.52379 $\therefore$ CI: $(0.37736, 5.42494)$

iii) Mean IBM : μ_0 $\left| \frac{\hat{\mu}_0 - \mu}{se(\hat{\mu}_0)} \right. \sim t_{n-2}$ $\left| x_0 = 40 \right|$

$E(\mu_0) = \hat{\alpha} + \hat{\beta} x_0 = 119.78409$

$se(\hat{\mu}_0) : \left(\left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}} \right] \times \hat{\sigma}^2 \right)^{1/2} = 10.59897$

95% CI: $119.784 \pm 23.61443 \longrightarrow (96.17866, 143.408)$

Q.3 i.) Comments on Plot:

- Details in the centre of the plot differ for all plots.
- Diff. in mean time in peak & non peak hours is from. Highest (A) to lowest (D)
- Variation in data C is less than in data D.

Reason for underlying difference cannot be known.

ii.) Assumptions in ANOVA : • Populations to be normal.
• They have a common variance.
• Observations are independent.

iii) H_0 : Mean difference is same H_1 : Mean difference is not same.

SS_{TOT} : S_{YY} : $\Sigma y^2 - n.\bar{y}^2$ = 1116.11.

SS_{Reg} : $\frac{1}{7}\left[64^2 + 44^2 + 8^2 + (-13)^2\right] - \frac{103^2}{28}$ = 516.11.

SS_{Res} : $SS_{TOT} - SS_{Reg}$: 600

ANOVA Table:

	D.f.	SS..	MSS.
Regression	3	516.11	172.04
Residual	24	600	25
Total	27	1116.11	

→ Under F stat : $\frac{MSS_{Reg}}{MSS_{Res}}$ = 6.88 | T.S = 6.88.

At : $F_{5\%, 3, 24}$ → 3.009 as, $F_{cal} > F_{Tab}$ } Reject H_0

→ Thus, we conclude... there is difference between means.

iv.) Analysis of Mean Diff.

Q.4

a.) Mathematical Model of One Way ANOVA Table:

$$y_{ij} = \underbrace{\mu}_{\text{mean response}} + T_i + \underbrace{e_{ij}}_{\text{error term}} \; ; \; i = 1, 2, 3, \ldots k. \quad j = 1, 2, 3 \ldots n$$

where → e_{ij} is the error term $\sim N(0, \sigma^2)$

b.) H_0: Mean claim amts are equal

H_1: Mean claim amts are not equal.

$n = 33$: $\sum_{i=1}^{5} \sum_{j=1}^{n_i} y_{ij} = 1856$ | $\sum_{i=1}^{5} \sum_{j=1}^{5} y^2{ij} = 174,316$

CF: $\frac{174316}{n} = 104385.94$

SS_{TOT} : $174316 - CF = 69930.06$

SS_{Reg}: 22325.69 | SS_{Res}: $SS_{TOT} - SS_{Reg}$: 47604.37

Table:

	Df.	SS	MSS.
Reg	4	22326	5581.42
Res	28	47604	1700.16.
Total	32	69930	

F_{stat}: $\frac{MSS_{Reg}}{MSS_{Res}}$: 3.283 | $F_{5\%, 4, 28} \rightarrow 2.714$

As $F_{cal} > F_{Tab}$. We

Reject H_0.

c.) Observed Frequency (O_i)

Papers	Salary Per Annum 3 - 5	5 - 8	8-10	10 - 12	Total
0 - 3	45	20	6	5	76
4 - 7	7	20	9	6	42
7 - 9	5	8	15	12	40
Total.	57	48	30	23	158

Expected Frequency (E_i) :

Paper	Salary 3-5	5 - 8	8-10	10 - 12	Total.
0-3	27.41772	23.08	14.43	11.06	
4-7	15.151	12.76	7.97	6.11	
7-9	14.43	12.15	7.59	5.88.	
					158.

H_0 : Salary independent of paper cleared / H_1 : Salary dependent on paper cleared.

TS : $\frac{\Sigma(O_i - E_i)^2}{E_i} = \underline{\underline{49.81146}}$ Df : $(r-1)(c-1)$

$\rightarrow \chi^2_{6, 5\%} = 12.59$ | $\chi^2_{cal} > \chi^2_{tab} \rightarrow$ Reject H_0

$\rightarrow$ Salaries are dependend upon paper cleared...

Q.5 Assumptions of ANOVA: Populations are normal
Populations have common variance.
Observations are independent.

→ Mean at rate 1: $\sqrt{x}$: $\dfrac{\sqrt{29} + \sqrt{13} + \sqrt{21}}{3} = 4.52443.$

→ Vari at rate 2. $\log_e(x)$: $\dfrac{\sum(x' - \bar{x})^2}{n-1} = 0.12127$

iii) $\log_e$ transformation must be ~~done~~, before carrying out one way ANOVA. for transformation

iv.) a. $Y_{ij} = \mu + T_i + e_{ij}$ $i \rightarrow 1,2,3\ldots k$, $j = 1,2,3\ldots n$.
$e_{ij} \rightarrow$ error term $\sim N(0, \sigma^2)$

b.) $H_0: T_i = 0$ $H_1: T_i \neq 0$ $i = 1,2,3,4$

Rate	Mean	Var	Y_i^2
1	2.992	0.163	86.582
2	4.824	0.121	209.648.
3	5.716	0.186	294.05
4	6.545	0.148	389.06
Total:	20.110	0.618	973.373

SS_{Rate}: 21.153.

SS_{Res}: 1.236.

	Df.	SS	MSS.
Rate	3	21.153	7.051
Resid	8	1.236	0.155
Total	11	22.389	

TS: $\dfrac{MSS_{Rate}}{MSS_{Res}} = 45.638$

$\therefore F_{3,8,1\%} = 7.95 < F_{tab}$

→ Reject H_0.

c.) F_{cal} ~~will~~ is too high. P-value is very close to 0.
Thus strong evidence against Ho. Thus, means are not equal.

Q.6 : $n=10$, $r = \dfrac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$

$S_{xx} = \Sigma x^2 - n.\bar{x}^2 = \cancel{661.2}.549$

$S_{yy} = \Sigma y^2 - n.\bar{y}^2 = 8923.6$

$S_{xy} = \Sigma xy - n.\bar{x}\bar{y} = 661.2$

$\boxed{r = 0.94466}$

ii) Eqn line: $\hat{\beta} = \dfrac{S_{xy}}{S_{xx}} \longrightarrow 12.04371$

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 9.2295$ | $\hat{y} = 9.2295 + 12.04371x$

iii) $SS_{TOT} = SS_{yy} : 8923.6$

$SS_{Reg} = \dfrac{S_{xy}^2}{S_{xx}} = 7963.30492$

$SS_{Res} : 960.29508$

$R^2 = \dfrac{S_{xy}^2}{S_{xx} \cdot S_{yy}} = 89.2387\%$

★ Comment :

→ Model is a good fit as model explains 89% of total variance.
→ Better model with more R^2 can be used.

Q.7 i.) $S_{xx} = \Sigma x^2 - n.\bar{x}^2 = 4788.422$. $S_{xy} = \Sigma xy - n\bar{x}\bar{y}$

$S_{yy} = \Sigma y^2 - n.\bar{y}^2 = 1189.204$ $= 2176.84$

$\hat{\beta} = \dfrac{S_{xy}}{S_{xx}} : 0.45451$ $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 10.78056$.

$\therefore \hat{y} = 10.78056 + 0.45451x$

ii) $\beta = 0$ $H_1: \beta \neq 0$ $\hat{\sigma}^2 = \dfrac{\left(S_{yy} - \dfrac{S_{xy}^2}{S_{xx}}\right)}{n-2} = 22.20136$

TS = $\dfrac{\hat{\beta} - \beta}{\sqrt{\dfrac{\hat{\sigma}^2}{S_{xx}}}} \sim t_{n-2} = 6.67568$ | $t_{9, 2.5\%} = 2.262$

$t_{cal} > t_{tab} \rightarrow$ Reject H_0

$\rightarrow \beta \neq 0 \rightarrow$ Linear Relationship exists.

* Assumptions: Populations $\rightarrow$ Normal Distributions

↳ Common Variance

↳ Independent Observations

iii) $r = \dfrac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = 0.91213$

iv) Mean Response: $x_0 = 25$ $\mu_0 = \alpha + \hat{\beta} x = 22.15331$

$se(\hat{\mu}_0) = \sqrt{\left\{ 1/n + \dfrac{(\bar{x} - x_0)^2}{S_{xx}} \right\} \hat{\sigma}^2} = 1.55181$

$\therefore$ 95% CI: $\hat{\mu}_0 \pm t_{9, 2.5\%} \times s.e(\hat{\mu}_0)$

$= 22.15331 \pm 3.51018 \longrightarrow (18.643, 25.663)$

v) Individual Response:

$\hat{y}_0 = \hat{\alpha} + \hat{\beta} x_0 = 22.153$ | $s.e(\hat{y}_0) = \sqrt{\left(1 + 1/n + \dfrac{(\bar{x} - x_0)^2}{S_{xx}}\right) \times \hat{\sigma}^2}$

$= 4.860$

95% CI: 22.15331 ± 4.860

$= (10.932, 33.37462)$

vi) There is a relationship between residuals & explanatory variables

- Assumption of Normality is wrong. Obsv $\rightarrow$ Not independent $\rightarrow$ Not Good Fit

Q.8:

Q.9: $S_{xx} = \Sigma x^2 - n.\bar{x}^2 = 75$ | $S_{yy} = \Sigma y^2 - n.\bar{y}^2 = 2452.4$.

$S_{xy} = \Sigma xy - n.\bar{x}\bar{y} = -296$

$r = \dfrac{S_{xy}}{\sqrt{S_{xx}.S_{yy}}} = \underline{-0.686002}$

→ They have a weak -ve correlation

iii.) $H_0: \rho = 0$ $H_1: \rho \neq 0$.

~~to~~ $w = \tanh^{-1}(r) \sim N\left(\tanh^{-1}(\rho), \ 1/n-3\right)$

$\therefore$ TS: $\dfrac{\tanh^{-1}(r) - \tanh^{-1}(\rho)}{\sqrt{1/n-3}} \sim \cancel{t_{n-}}\ Z_{5\%}$.

$\therefore$ TS: $Z_{cal} = -2.07893$ $Z_{tab} = 1.644$.

As $|Z_{cal}| > Z_{tab}$ → We reject H_0, thus $\underline{\rho > 0}$.

iv.) $\hat{\beta} = \dfrac{S_{xy}}{S_{xx}} = -3.94667$ | $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

$= 82.16$.

$\therefore \hat{y} = 82.16 + 3.94667x$.

v.) $R^2 = \dfrac{S_{xy}^2}{S_{xx}.S_{yy}} = 47.059\%$. → Model explains only 47% of Total Variation.

→ Model not a good fit for data...

vi.) $x_0 = 1$, $\hat{\mu}_0 = \hat{\alpha} - \hat{\beta}x_0$ → expected change = $\underline{78.2133}$.

Q.10. x 0.27 0.36 0.3

y 0.05 0.1 0.08.

$S_{yy} = \sum y^2 - n.\bar{y}^2 = \cancel{0.0042}\ 0.00126$ | $S_{xy} = \sum xy - n.\bar{x}\bar{y} = 0.0022$

$S_{xx} : \sum x^2 - n.\bar{x}^2 = 0.0042$

$SS_{TOT} = S_{yy} = 0.00126.$ $SS_{Reg} = \frac{S_{xy}^2}{S_{xx}} = 0.00115.$

$SS_{Res} : 0.00011$

→ ANOVA TABLE :

Sources:	DF.	SS	MSS.
Reg	1	0.00115	0.00115
Res.	1	0.00011	0.00011
Total	2	0.00126	

$H_0 : \beta = 0$ $H_1 : \beta \neq 0$ → One sided. Test.

TS: $\frac{MSS_{Reg}}{MSS_{Res}} = 10.15455$

$F_{1,1,5\%} = 161.4$ ⟶ $F_{cal} < F_{tab}$:

We donot reject Ho, thus, no effect on residual rate...