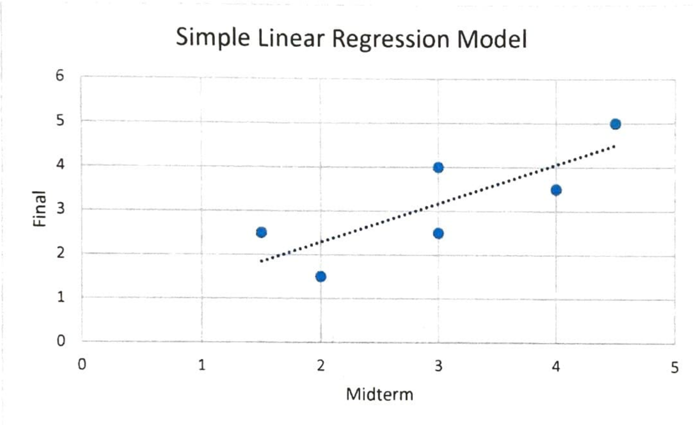
ASSIGNMENT 2

Q1.1:

i	midterm, xi	final, yi	xi^2	yi^2	xiyi	yi hat	ei^2
1	1.5	2.5	2.25	6.25	3.75	1.83974359	0.43593853
2	4	3.5	16	12.25	14	4.05128205	0.3039119
3	3	4	9	16	12	3.16666667	0.69444444
4	3	2.5	9	6.25	7.5	3.16666667	0.44444444
5	4.5	5	20.25	25	22.5	4.49358974	0.25645135
6	2	1.5	4	2.25	3	2.28205128	0.61160421
sum	18	19	60.5	68	62.75	19	2.74679487
mean	3	3.16666667	10.0833333	11.3333333	10.4583333	3.16666667	0.45779915

Q1.2



$$\hat{\beta}_1 = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2}$$

$$\hat{\beta}_1 = \frac{6 \times 62.75 - (18 \times 19)}{6 \times 60.5 - (18)^2}$$

$$= \frac{34.5}{39}$$

$$\hat{\beta}_1 = 0.8846$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
$$= 3.16667 - (0.8846 \times 3)$$
$$\hat{\beta}_0 = 0.5129$$

$$\therefore \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$
$$\hat{y} = 0.5129 + 0.8846x$$

Interpretation:

$\hat{\beta}_1 = 0.8846$: As the midterm grade increases 1%, the average final exam grade increases by 0.8846%.

$\hat{\beta}_0 = 0.5129$: When the midterm grade is 0%, the average final exam grade is 0.5129%.

Q1.3 Calculate the error sum of squares & the residual mean square

$$SS_{Res} = \sum_{i=1}^{n} e_i^2 = 2.7468$$

$$\hat{\sigma}^2 = MS_{Res} = \frac{SS_{Res}}{n-2} = \frac{2.7468}{4} = 0.6867$$

Q1.4 Calculate the 90% confidence intervals for β_0, β_1. And interpret the confidence interval of the slope

i	x_i	(xi − xbar)	$(x_i - \text{xbar})^2$
1	1.5	−1.5	2.25
2	4	1	1
3	3	0	0
4	3	0	0
5	4.5	1.5	2.25
6	2	1	1
sum			6.5
mean	3		

90% CI of β_0:

$$\hat{\beta}_0 \pm t_{\alpha/2}^{(n-2)} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

$\alpha = 0.1$

$\beta_0 = 0.5129$

$$\hat{\sigma} = \sqrt{MS_{Res}} = \sqrt{0.6867} = 0.8287$$

$$0.5129 \pm t_{0.05}^{4} \times 0.8287 \times \sqrt{\frac{1}{6} + \frac{3^2}{6.5}}$$

$= 0.5129 \pm 2.1318 \times 0.8287 \times 1.2455$

$= 0.5129 \pm 2.200$

$= (-1.6874, 2.7129)$

$\therefore$ Let $\alpha = 0.1$, we are 90% confident that the confidence interval constructed contains the true β_0.

90% CI of $\hat{\beta}_1$

$$\hat{\beta}_1 \pm t_{\alpha/2}^{(n-2)} \hat{\sigma} \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

$\hat{\beta}_1 = 0.8846$

$\alpha = 0.1$

$\hat{\sigma} = 0.8287$

$\sum_{i=1}^{n}(x_i-\bar{x})^2 = 6.5$

$$0.8846 \pm t_{0.05}^{4} \times 0.8287 \times \sqrt{\frac{1}{6.5}}$$

$= 0.8846 \pm 2.1318 \times 0.8287 \times 0.3922$

$= 0.8846 \pm 0.6929$

$= (0.1917, 1.5775)$

① We are 95% confident that if the midterm mark increased by 1%, the mean final grade will increase by at least 0.1917% and at most 1.5775%.

② $0 \notin$ CI, hence we have evidence that X (the midterm mark) & y (final exam grade) has a linear relation

Q1.5 Apply a t-test to test $H_0: \beta_0 = 0$, $H_a: \beta_0 \neq 0$. Solve using reject point.

$H_0: \beta_0 = 0$
$H_a: \beta_0 \neq 0$

$$t = \frac{\hat{\beta}_0 - \beta_{00}}{\hat{\sigma}\sqrt{c_{00}}} \sim t(n-2)$$

$$t = \frac{0.5129 - 0}{0.8287 \times 1.2455} = \frac{0.5129}{1.0321} = 0.4969$$

RR:

$$|t| > t_{\alpha/2}^{(n-2)}$$

$$RR\ \pm t_{\alpha/2}^{(n-2)} = t_{0.05}^{4} = \pm 2.132$$

$RR.\ (-\infty, -2.132) \cup (2.132, \infty)$

$TS = 0.4969$

TS $t = 0.4969 \notin RR \longrightarrow$ do not reject H_0

∴ Since the test statistic is not in reject ion region, we do not reject the null hypothesis.
∴ $\beta_0 = 0$ and the regression is insignificant.

Q1.6 Apply a t-test to $H_0: \beta_1 = 0$, $H_a: \beta_1 \neq 0$. Solve using P-value

$H_0: \beta_1 = 0$
$H_a: \beta_1 \neq 0$

$$t = \frac{\hat{\beta}_1 - 0}{\hat{\sigma}\sqrt{c_{11}}}$$

$$= \frac{0.8846 - 0}{0.8287 \times 0.3922} = \frac{0.8846}{0.3250}$$

$= 2.7217$

Reject region:

$2P(T > |t|) < \alpha$

$2P(T > 2.7217) < \alpha$

$2 \times P(T > 2.7217)$

$2 \times (1 - pt(2.7217, 4))$

$= 2 \times 0.026447 = 0.05289$

$0.05289 < 0.1$

↳ P-value $< \alpha$ - reject H_0

∴ We reject the null hypothesis because the p-value is less than alpha. $\beta_1 \neq 0$ - the regression is significant

Q1.7 Calculate the prediction value of y, given x=3

$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0$

$X_0 = 3$

$\hat{y}_0 = 0.5124 + 0.8846 \times 3$
$\hat{y}_0 = 3.1667$

Q1.8 Calculate the 90% CI for the expected value of y, given x=3

$\alpha = 0.1$

$\hat{y}_0 \pm t_{\alpha/2}^{(n-2)} \hat{\sigma} \sqrt{h_{00}}$

$= \hat{y}_0 \pm t_{\alpha/2}^{(n-2)} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$

$= 3.1667 \pm t_{0.05}^{4} \times 0.8287 \times \sqrt{\frac{1}{6} + \cancel{\frac{(3-3)^2}{6.5}}^{\to 0}}$

$= 3.1667 \pm 2.132 \times 0.8287 \times 0.40825$
$= 3.1667 \pm 0.7213$
$= (2.4454, 3.888)$

∴ A 90% CI for the expected value of y given x=3 is (2.4454, 3.888)

Q1.9 calculate the 90% prediction interval for y, given x=3

$\hat{y}_0 \pm t_{\alpha/2}^{(n-2)} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$

$= 3.1667 \pm 2.132 \times 0.8287 \times \sqrt{1+\frac{1}{6}} =$

$= 3.1667 \pm 1.9083$
$= (1.2584, 5.075)$

∴ A 90% prediction interval for y, given x=3 is (1.2854, 5.075)

9.10. Calculate sample coefficient of determination & sample correlation coefficient based on the regression model. Interpret the relationship b/w midterm & final grade

$$R^2 = \frac{SSR}{SS_T} = 1 - \frac{SS_{Res}}{SS_T}$$

$$SST = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$

$$SST = 68 - 6 \times (3.16667)^2$$
$$SST = 7.8332$$

$$R^2 = 1 - \frac{2.7468}{7.8332}$$

$$R^2 = 0.6493$$

$\therefore$ 0.6493% of the total variation is explained by the regression model

Since $\hat{\beta}_1 = 0.884670$, $r = +\sqrt{R^2}$

$$r = +\sqrt{0.6493}$$
$$r = 0.8058$$

Since $r \longrightarrow 1$ & is very close to 1, midterm & final exam grade are highly positively correlated

Q1.11 Apply a F-test to test $H_0: \beta_1 = 0$, $H_a: \beta_1 \neq 0$. Note: $F_{0.1}^{(1,4)} = 4.5448$

$H_0: \beta_1 = 0$

$H_a: \beta_1 \neq 0$

$$F(\text{model}) = \frac{MSR}{MSRes} = \frac{SSR/1}{SSRes/(n-2)} \sim F(1,4)$$

$$F(\text{model}) = \frac{MSR}{MSRes} = \frac{5.0865/1}{2.7468/4} = 7.4072$$

$$\frac{F(\text{model})}{7.4072} > \frac{F_{0.1}^{(1,4)}}{4.5448}$$

$\rightarrow$ reject H_0

$\therefore$ We reject H_0 since $F(\text{model}) > F_{\alpha}^{(1,n-2)}$, which means we accept the null hypothesis is $\exists\ \beta_1 \neq 0$.