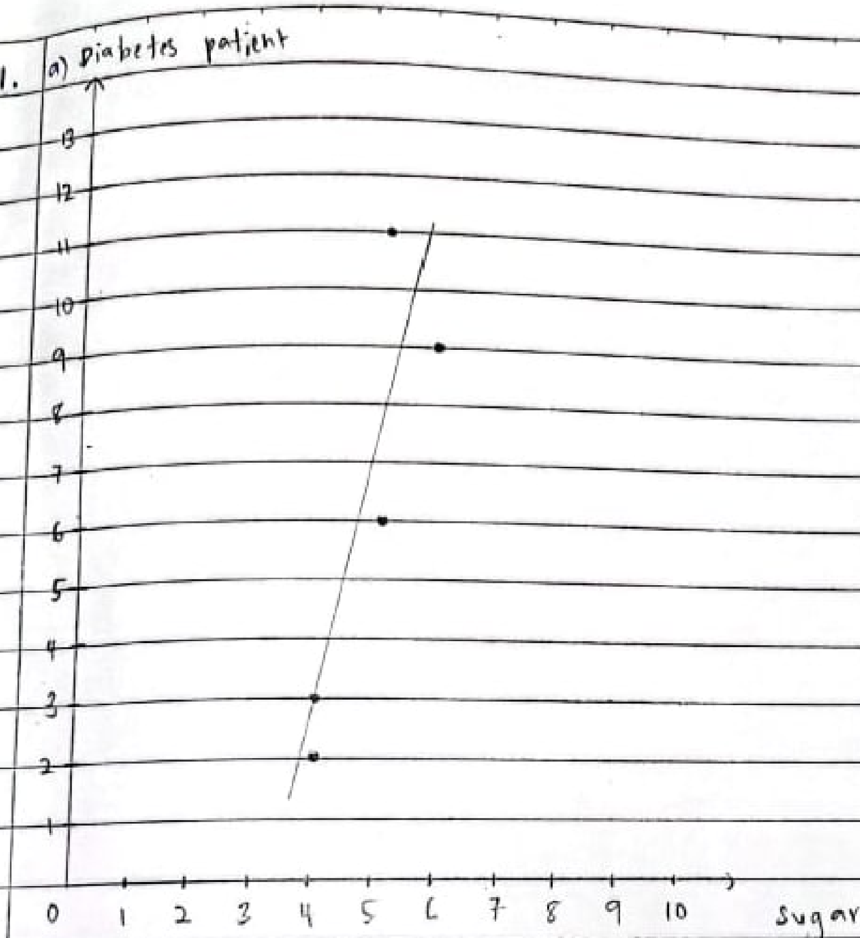
SECI 2143-04 PROBABILITY & STATISTICAL DATA ANALYSIS – SEMESTER 2

Assignment 4

GROUP MEMBERS	MATRIC NUMBER
AFIF BIN GENARI@AZHARI	A21EC0003
MUHAMMAD ADAM HAIKAL BIN MOHD SHUHADZLEE	A21EC0062
MUHAMMAD FITRI BIN ISMAIL	A21EC0076
MUHAMMAD SYAIF ALFARIZ BIN ILYAS SUSANTO	A21EC0094

Deadline of submission: 30 June 2022

QUESTION 1

1. a) 

The feature of the scatter plot is it's a positive linear relationship.

b) Yes its appear that the two variables are linear relationship. It is positive and strong relationship due to the plots that near to the line. It can be seen that the diabetes patient increases as the sugar increases. Thus, this indicate that there is positive relationship between the two variables.

No. Date

c)

Sugar (x)	diabetes patient (y)	xy	y^2	x^2
5	6	30	36	25
6	9	54	81	36
4	3	12	9	16
4	2	8	4	16
5	11	55	121	25
$\Sigma x = 24$	$\Sigma y = 31$	$\Sigma xy = 159$	$\Sigma y^2 = 251$	$\Sigma x^2 = 118$

$$r = \frac{\Sigma xy - (\Sigma x \Sigma y)/n}{\sqrt{[(\Sigma x^2) - (\Sigma x)^2/n][(\Sigma y^2) - (\Sigma y)^2/n]}} \qquad n = 5$$

$$= \frac{159 - (24 \times 31)/5}{\sqrt{(118 - 24^2/5)(251 - 31^2/5)}}$$

$$= \frac{10.2}{12.8312}$$

$= 0.7949$

$= 0.8$, strong relationship

$\therefore$ Yes, the value of r is consistent with my answer which it is strong relationship

d) Yes, because based on the data given the diabetes patients are increases as the sugar consumption increases which logically diabetes is type of disease that happened due to excessive amounts of sugar in the body.

QUESTION 2

No.: Date:

Question 2

x	y	xy	x^2
5	85	425	25
4	103	412	16
6	70	420	36
5	82	410	25
5	89	445	25
5	98	490	25
6	66	396	36
6	95	570	36
2	169	338	4
7	70	490	49
7	48	336	49
$\Sigma x = 58$	$\Sigma y = 975$	$\Sigma xy = 4732$	$\Sigma x^2 = 326$

$$b_1 = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} = \frac{4732 - \frac{(58)(975)}{11}}{326 - \frac{(58)^2}{11}}$$

$$= -20.26126126$$

$$b_0 = \bar{y} - b_1 \bar{x} \qquad \bar{y} = \frac{975}{11} \qquad \bar{x} = \frac{58}{11}$$

$$= 88.64$$

$$b_0 = 88.64 - (-20.26126126)(58/11)$$

$$= 195.4684685$$

$$\hat{y} = 195.47 - 20.261x$$

Nama: .... Tarikh: ....

$\hat{y} = -20.261x + 195.47$

when $x = 0$ (First coordinate)

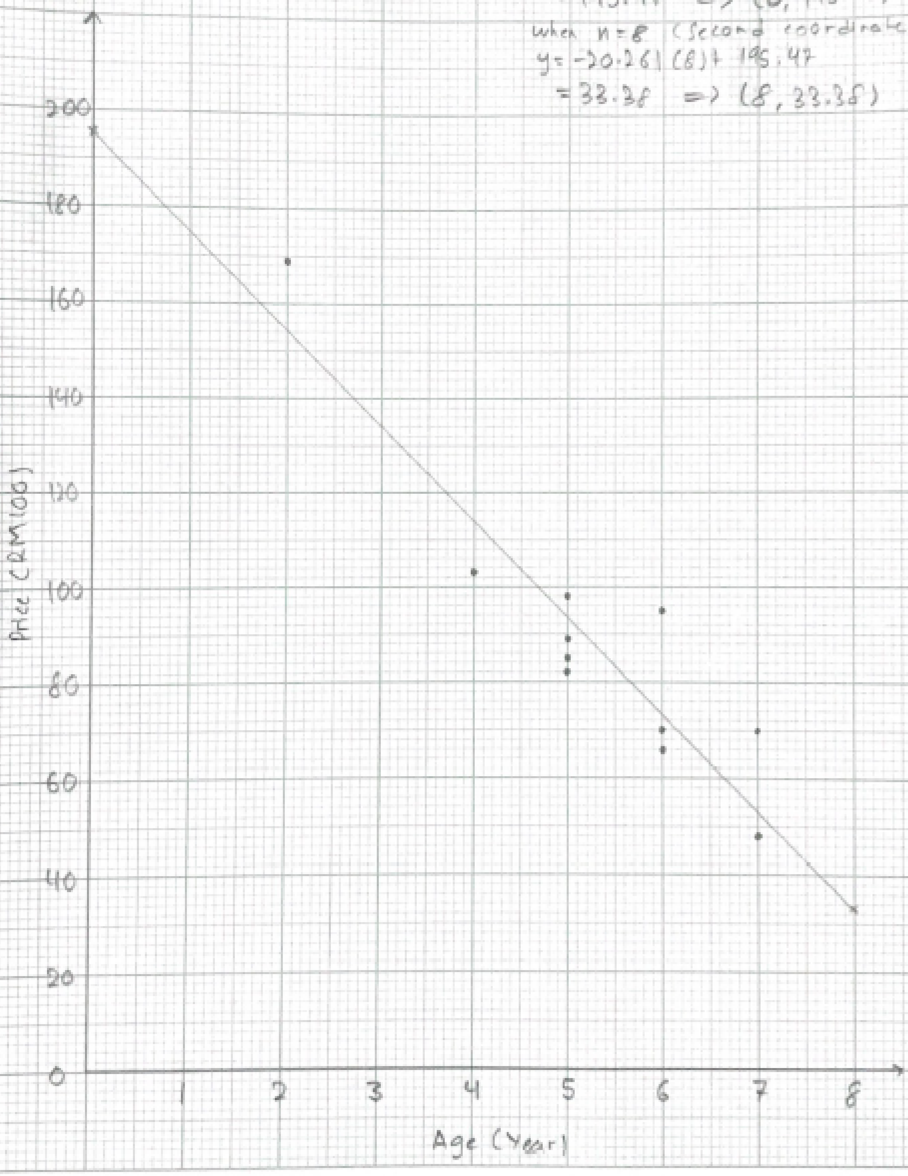
$y = -20.261(0) + 195.47$

$= 195.47 \Rightarrow (0, 195.47)$

when $x = 8$ (Second coordinate)

$y = -20.261(8) + 195.47$

$= 33.38 \Rightarrow (8, 33.38)$



Price (RM100) Y	Age (year) x	$\hat{y}$	$(\hat{y}-\bar{y})^2$	$(y_i-\bar{y})^2$
85	5	94.17	30.62	13.22
103	4	114.43	665.31	206.31
70	6	73.90	217.16	347.31
82	5	94.17	30.62	44.04
89	5	94.17	30.62	0.13
98	5	94.17	30.62	87.68
66	6	73.90	217.16	512.40
95	6	73.90	217.16	40.50
169	2	154.95	4397.50	6458.31
70	7	53.64	1224.75	347.31
48	7	53.64	1224.75	1651.31
			SSR = 8286.27	SST = 9708.52

$$R^2 = \frac{SSR}{SST} = \frac{8286.27}{9708.52} = 0.85$$

85% of the variation in price for second-hand car is explained by variation in age.

c. Based on the value of $R^2 = 0.85$ we can say that the relationship between age (independent variable) and price (dependent variable) is strong linear relationship. Some but not all of the variation in Y is explained by variation in x.

d. $\hat{Y} = -20.261x + 195.47$

Here b_1 which is the slope coefficient $b_1 = -20.261$ tells us that the average value for second-hand car decreases by 20.261 (RM 100) = RM 2026.10, on average, for each additional age or we can say that when the age of the second-hand car increases the price will be decreased by RM 2026.10.

e. 3 year second-hand car

$\hat{Y} = -20.261x + 195.47$
$= -20.261(3) + 195.47$
$=$ RM 11442.60

4 year second-hand car

$\hat{Y} = -20.261x + 195.47$
$= -20.261(4) + 195.47$
$=$ RM 13468.70

QUESTION 3

Question 3.

Product A	Product B	Product C
210	210	180
240	240	210
270	240	210
270	270	210
300	270	240

(a) H_0 : Group of products receive different treatment has no effect on the size of product.

H_1 : Group of products receive different treatment has affect on the size of product.

(b) Product A,

mean, $\bar{x} = \dfrac{210+240+270+270+300}{5} = 258$

Variance, $s^2 = \dfrac{(210-258)^2+(240-258)^2+(270-258)^2+(270-258)^2+(300-258)^2}{5-1}$

$= 1170$

Product B,

mean, $\bar{x} = \dfrac{210+240+240+270+270}{5} = 246$

variance, $s^2 = \dfrac{(210-246)^2+(240-246)^2+(240-246)^2+(270-246)^2+(270-246)^2}{5-1}$

$= 630$

Product C,

mean, $\bar{x} = \dfrac{180+210+210+210+240}{5} = 210$

variance, $s^2 = \dfrac{(180-210)^2+(210-210)^2+(210-210)^2+(210-210)^2+(240-210)^2}{5-1}$

$= 450$

(c) Test statistic,

$$F = \frac{\text{variance between samples}}{\text{variance within samples}} = \frac{ns_{\bar{x}}^2}{s_p^2}$$

mean between sample,

$$\bar{\bar{x}} = \frac{258 + 246 + 210}{3} = 238$$

standard deviation between samples,

$$s_{\bar{x}} = \sqrt{\frac{(258-238)^2 + (246-238)^2 + (210-238)^2}{3-1}}$$

$$= 24.98$$

variance between sample,

$$ns_{\bar{x}}^2 = 5(24.98)^2 = 3120.002$$

variance within sample,

$$s_p^2 = \frac{(\sqrt{1170})^2 + (\sqrt{630})^2 + (\sqrt{450})^2}{3}$$

$$= 750$$

$$F = \frac{3120.002}{750} = 4.16$$

(d) - Numerator = $k-1 = 3-1 = 2$

- Denominator = $k(n-1) = 3(5-1) = 12$

(e) Critical value, F with $\alpha = 0.05$

- F critical value = 3.89

(f) Test the claim and state the conclusion.

- since $F_{\text{test statistic}}$ (4.16) > $F_{\text{critical value}}$ (3.89), thus we reject null hypothesis at $\alpha = 0.05$.

- There is sufficient evidence to claim that group of products receive different treatment has affect on the size of product