SCHOOL OF COMPUTING
Faculty of Engineering

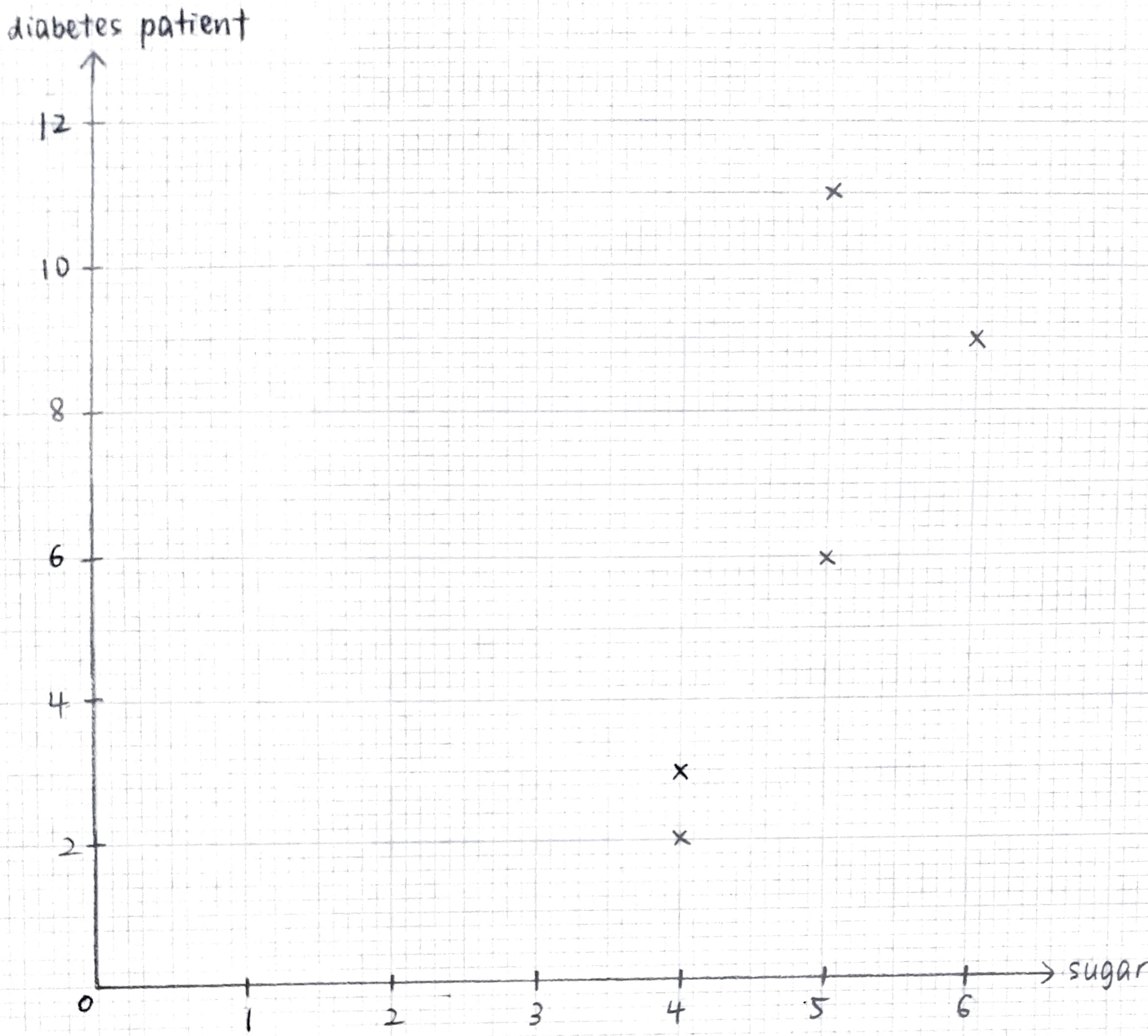
Assignment 4 (Group)
SECI2143-07
Probability & Statistical Data Analysis

SEMESTER II, SESSION 2021/2022

Lecturer: Dr. Nor Azizah Ali

Group Name: Little Group

Name	Student ID
TIA SIAW XUEN	A21EC0233
CHUA XIN LIN	A21EC0020
MA ZE JUN	A21EC4009
KEE LE WEI	A21EC0189



Comment: The scatterplot above shows a strong, positive linear association between sugar consumption and number of diabetes patient with a few potential outliers.

No: Date:

(b) Yes. There exists a linear relationship between the two variables. The relationship is positive and strong.

(c) Let x = sugar, y = number of diabetes patient.

	x	y	xy	x^2	y^2
	5	6	30	25	36
	6	9	54	36	81
	4	3	12	16	9
	4	2	8	16	4
	5	11	55	25	121
Σ	24	31	159	118	251

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

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

$$= \frac{10.2}{\sqrt{(2.8)(58.8)}}$$

$$= 0.7793$$

∴ Yes. The value r shows that both variable has strong and positive relationship.

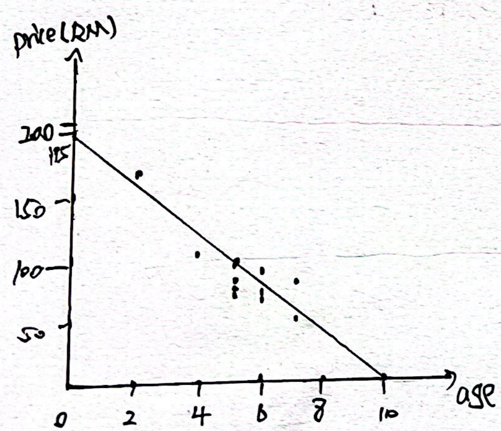
(d) Yes. It is reasonable to conclude that by sugar consumption against the number of diabetes patient. From the data, we can know that when the sugar consumption increase, the number of diabetes increase. Besides, the scatterplot and r value proves that there is a strong and positive relationship between the two variables.

Q2

b) $\hat{y} = b_0 + b_1 x$ $\bar{x} = 5.27$ $\bar{y} = 88.6$

$$b_1 = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = -20.26$$

$b_0 = \bar{y} - b_1 x$ for $x = 4$ $b_0 = 185.47$



(a)
(b) Q a. and b are in the figure

(c) the slope is negative. so is a negative relationship

(d) $y = 185.247 - 20.26x = 185.5 - 20.26x$

Each additional year will reduce 20.26 price

(e) for $x=3$. $y = 185.5 - 20.26 \times 3 = 124.72$

$x=4$ $y = 185.5 - 20.26 \times 4 = 114.46$

ASSIGNMENT 4

Question 3

a) $H_0: \mu_1 = \mu_2 = \mu_3$

H_1: At least one mean is different.

b) Product A:

$n = 5$

$$\bar{x} = \frac{210 + 240 + 270 + 270 + 300}{5}$$

$= 258$

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

$= 1170$

Product B:

$n = 5$

$$\bar{x} = \frac{210 + 240 + 240 + 270 + 270}{5}$$

$= 246$

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

$= 630$

Product C:

$n = 5$

$$\bar{x} = \frac{180 + 210 + 210 + 210 + 240}{5}$$

$= 210$

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

$= 450$

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

$= 238$

Subject: .... No: .... Date: 29/6/2022

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

$$= 624$$

$$ns_{\bar{x}}^2 = 5(624)$$

$$= 3120$$

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

$$= 750$$

Test statistics, $F = \frac{ns_{\bar{x}}^2}{s_p^2} = \frac{3120}{750} = 4.16$

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

Denominator $= k(n-1) = 3(5-1) = 3(4) = 12$

e) F-critical value $(\alpha = 0.05) = 3.89$

f) Since F test statistics $= 4.16 >$ F critical value $= 3.89$, hence we reject the null hypothesis. There is sufficient evidence to claim that the different product treatment have not the same mean for the size of product.