Name: Greendee Roper B. Panogalon **BSIT 3B**

Linear Regression Analysis

1. The sales of a company (in million dollars) for each year are shown in the table below.

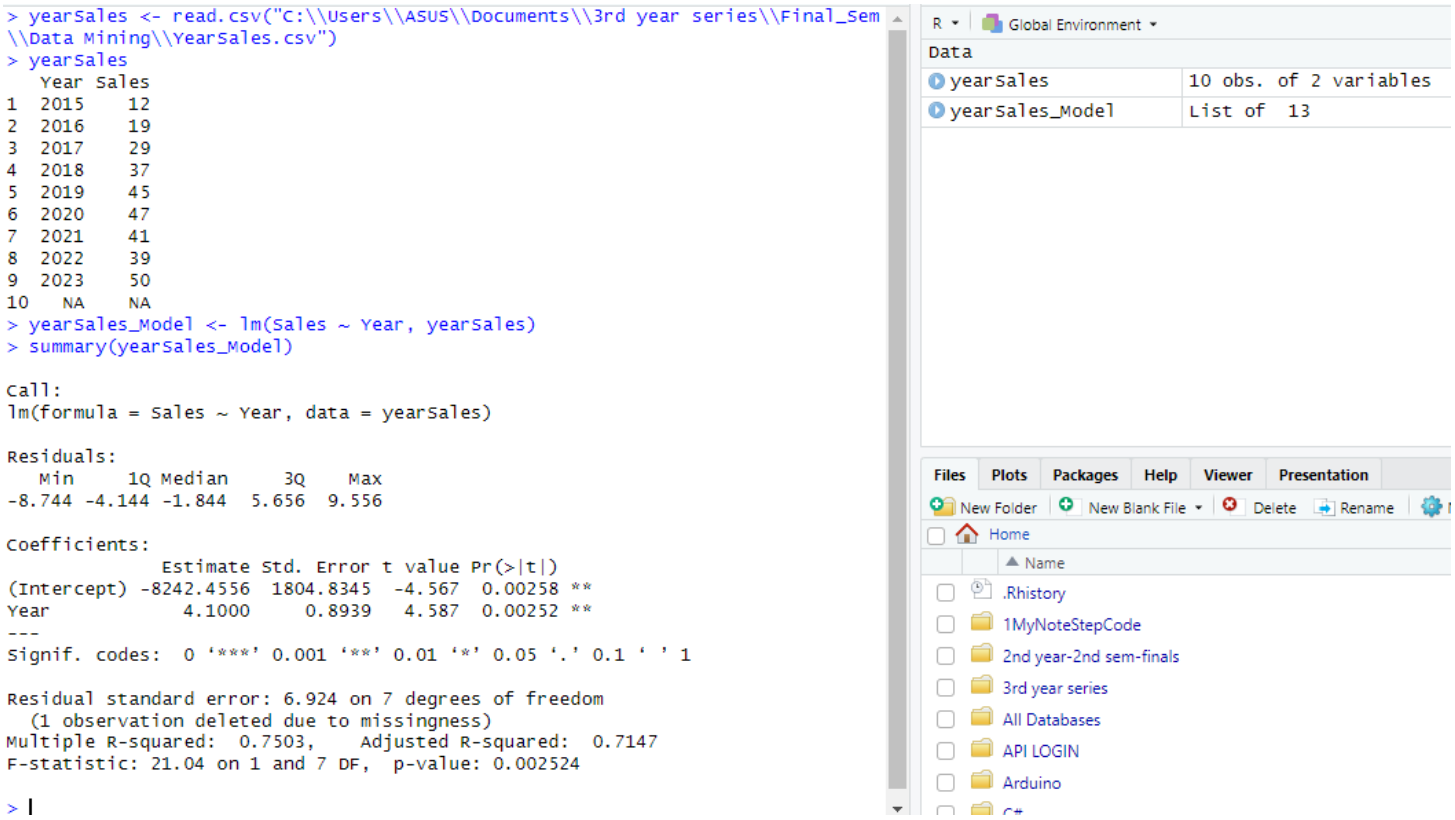
X (Year)	2015	2016	2017	2018	2019	2020	2021	2022	2023
Y (Sales	12	19	29	37	45	47	41	39	50

a) Find the least square regression line y = a x + b.
b) Use the least squares regression line as a model to estimate the sales of the company in 2024-2028.

Excel:

1									
2		Year (X)	Sales (Y)	XY	X^2				
3		2015	12	24180	4060225				
4		2016	19	38304	4064256				
5		2017	29	58493	4068289				
6		2018	37	74666	4072324				
7		2019	45	90855	4076361				
8		2020	47	94940	4080400				
9		2021	41	82861	4084441				
10		2022	39	78858	4088484				
11		2023	50	101150	4092529				
12	Summation	**18171**	**319**	**644307**	**36687309**				
13	n = 9								
14		X^1 = summation(X)/n			2019				
15		Y^1 = summation(EY)/n			35.44444444				
16									
17									
20									
21		b=EXY-nX'Y'			246				
22		EX2-nX'2			60				
23									
24			b=		4.1				
25									
26									
27			a=		-8242.455556				
28									
29			Y = a + bx	Y = -8242.46 + 4.1x					
30									
31		a)	Y = 4.1x -8242.46						
32									
33		b)	Estimated sales for year 2024 - 2028						**Sales**
34			2024	4.1	-8242.46	8298.4			**55.94**
35			2025	4.1	-8242.46	8302.5			**60.04**
36			2026	4.1	-8242.46	8306.6			**64.14**
37			2027	4.1	-8242.46	8310.7			**68.24**
38			2028	4.1	-8242.46	8314.8			**72.34**

R studio:



```
> yearSales <- read.csv("C:\\Users\\ASUS\\Documents\\3rd year series\\Final_Sem
\\Data Mining\\YearSales.csv")
> yearSales
   Year Sales
1  2015    12
2  2016    19
3  2017    29
4  2018    37
5  2019    45
6  2020    47
7  2021    41
8  2022    39
9  2023    50
10   NA    NA
> yearSales_Model <- lm(Sales ~ Year, yearSales)
> summary(yearSales_Model)

Call:
lm(formula = Sales ~ Year, data = yearSales)

Residuals:
   Min     1Q Median     3Q    Max
-8.744 -4.144 -1.844  5.656  9.556

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -8242.4556  1804.8345  -4.567  0.00258 **
Year            4.1000     0.8939   4.587  0.00252 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.924 on 7 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.7503,    Adjusted R-squared:  0.7147
F-statistic: 21.04 on 1 and 7 DF,  p-value: 0.002524

>
```


2. You have to examine the relationship between the age and price for used cars sold in the last year by a car dealership company.

   What is the estimated cost value of car ages 2, 3, 15, and 16?

Excel:

	Car Age (in years) X	Price (in dollars) Y	XY	X^2
	4	6300	25200	16
	4	5800	23200	16
	5	5700	28500	25
	5	4500	22500	25
	7	4500	31500	49
	7	4200	29400	49
	8	4100	32800	64
	9	3100	27900	81
	10	2100	21000	100
	11	2500	27500	121
	12	2200	26400	144
	12	2000	24000	144
	12	1900	22800	144
	13	1800	23400	169
	13	2100	27300	169
	13	1700	22100	169
	14	1500	21000	196
	14	1400	19600	196
	14	1600	22400	196
Summation	**187**	**59000**	**478500**	**2073**

n = 19						
	X^1 = summation(X)/n			9.842105263		
	Y^1 = summation(EY)/n			3105.263158		
	b=EXY-nX'Y'			-102184.211		
	EX2-nX'2			232.5263158		
		b=		-439.452241		
		a=		7430.39837		
		Y = a + bx	Y = 7430.39837 - 439.452241x			
	a)	**Y =-439.452241x + 7430.39837**				
	b)	Estimated cost value of car age 2, 3, 15, and 16				**Sales**
		2	-439.452241	7430.39837	-878.904	**6551.494**
		3	-439.452241	7430.39837	-1318.36	**6112.042**
		15	-439.452241	7430.39837	-6591.78	**838.6148**
		16	-439.452241	7430.39837	-7031.24	**399.1625**

R studio: