4-11 #3 66. Please fill in the blank and answer the following questions:

has answers. must use formulas to calculate

*put in excel & look @ formulas.

See X Y Chart for data on 4-11

SUMMARY OUTPUT

Regression Statistics	
Multiple R	.52067 coefficient of correlation
R Square	.27109 (multiple R)^2 .52067^2
Adjusted R Square	0.125
Standard Error	____
Observations	7

ANOVA

	df	SS	MS	F	Significance F
Regression	1	0.828	____	____	____
Residual	____	2.227	____		
Total	6	____			

	Coefficients	Standard Error	t Stat	P-Value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	112.413	9.484	11.853	7.52722E-05	88.033	136.793	____	____
X	-19.137	--------	-1.364	0.231	-55.211	16.937	____	____

a. What is the coefficient of determination?	b. What is the coefficient of correlation?
formula on 4-7 need to calculate items prior on 4-4 to 4-6 calculator returns r = -.5195 shows work	get calc returns .5207

$$\boxed{Y = b_o + b_1X + \varepsilon}$$

Very important

Y = the dependent variable

X = the independent variable

b_o = the Y - intercept

b_1 = the slope of the line

ε = the random error

Here we are interested in using X (the independent variable) to predict the future value of Y (the dependent variable), therefore, b_o in this equation refers to the intercept term and b_1 is the slope.

$$b_1 = \frac{ChangeInY}{ChangeInX} = \frac{\Delta Y}{\Delta X}$$

The type of relationship which is often studied is the linear relationship. Two variables X and Y are linearly related if we can accurately predict one from the other. The general equation for any straight line can be written as $Y = b_o + b_1X + \varepsilon$, where b_o is the Y- intercept, b_1 is the slope of the line, and ε is the random error. We call this a deterministic relationship: if we know X, we can then determine Y.

Formulas:

$$b_1 = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2} \qquad b_o = \frac{\sum Y}{N} - b_1 \frac{\sum X}{N}$$

** **Note: The sign on b_1 is equal to sign on r which is Coefficient of Correlation.**

formula for intercept

Example 1: The Pizza Company is a chain of Italian-food restaurants located in a five state area. The most successful locations for The Pizza Company's have been near college campuses. The manager believes that quarterly sales for these restaurants (denoted by Y) are related positively to the size of the student population (denoted by X). The Pizza Company got 10 observations from those restaurants. The Pizza Company would like to know the slope and intercept of the regression equation. Moreover, if The Pizza Company wanted to predict sales for a restaurant to be located near a campus with 16 students, we would compute.

	X	Y	XY	X^2	Y^2	$\hat{Y}$	e	e^2
1.00	2.00	58.00	116.00	4.00	3,364.00	70.00	-12.00	144.00
2.00	6.00	105.00	630.00	36.00	11,025.00	90.00	15.00	225.00
3.00	8.00	88.00	704.00	64.00	7,744.00	100.00	-12.00	144.00
4.00	8.00	118.00	944.00	64.00	13,924.00	100.00	18.00	324.00
5.00	12.00	117.00	1,404.00	144.00	13,689.00	120.00	-3.00	9.00
6.00	16.00	137.00	2,192.00	256.00	18,769.00	140.00	-3.00	9.00
7.00	20.00	157.00	3,140.00	400.00	24,649.00	160.00	-3.00	9.00
8.00	20.00	169.00	3,380.00	400.00	28,561.00	160.00	9.00	81.00
9.00	22.00	149.00	3,278.00	484.00	22,201.00	170.00	-21.00	441.00
10.00	26.00	202.00	5,252.00	676.00	40,804.00	190.00	12.00	144.00
Total	**140.00**	**1,300.00**	**21,040.00**	**2,528.00**	**184,730.00**	**1,300.00**	**-**	**1,530.00**

$$b_1 = \frac{n*\left(\sum XY\right)-\left(\sum X\sum Y\right)}{n*\left(\sum X^2\right)-\left(\sum X\right)^2}$$
$$= \frac{10*(21040)-(140*1300)}{10*(2528)-(140)^2}$$
$$= \frac{210400-182000}{25280-19600}$$
$$= \frac{28400}{5680}$$
$$= \mathbf{5}$$

$$b_0 = \frac{\sum Y}{n} - b_1\frac{\sum X}{n}$$
$$= \frac{1300}{10} - 5*\left(\frac{140}{10}\right)$$
$$= 130-(5*14)$$
$$= 130 - 70$$
$$= \mathbf{60}$$

Thus, the estimated regression equation is: $\hat{Y} = 60 + 5X$

The slope of the regression equation is positive, implying that as student population increases, sales increase. In fact, we can conclude that an increase in the student population of 1 is associated with an increase of 5 in expected sales so sales are expected to increase by $5 per student.

If The Pizza Company wanted to predict sales for a restaurant to be located near a campus with 16 students.

$\hat{Y} = 60 + 5(16) = \mathbf{140}$

Coefficient of determination (r^2)

$$r^2 = \frac{b_1^2\left[\sum X^2 - \frac{\left(\sum X\right)^2}{n}\right]}{\sum Y^2 - \frac{\left(\sum Y\right)^2}{n}}$$

$$r^2 = \frac{5^2\left[2528 - \frac{(140)^2}{10}\right]}{184730 - \frac{(1300)^2}{10}}$$

$$r^2 = \frac{25(2528 - 1960)}{184730 - 169000}$$

$$r^2 = \frac{14200}{15730}$$

$r^2 = 0.9027$

Example 2:

	Y	X	XY	Y²	X²	Ŷ	e	e²
	10	1	10	100	1	9.714286	0.286	0.082
	8	2	16	64	4	8.028571	-0.029	0.001
	5	3	15	25	9	6.342857	-1.343	1.803
	6	4	24	36	16	4.657143	1.343	1.803
	3	5	15	9	25	2.971429	0.029	0.001
	1	6	6	1	36	1.285714	-0.286	0.082
Total	**33**	**21**	**86**	**235**	**91**	**33**	**-2.7E-15**	**3.771**

$$b_1 = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2} = \frac{6(86) - (21)(33)}{6(91) - (21)^2}$$

$$= \frac{516 - 693}{546 - 441} = \frac{-285}{105} = \mathbf{-1.6857}$$

$$b_0 = \frac{\sum Y}{n} - b_1 * \left(\frac{\sum X}{n}\right) = \frac{33}{6} - (-1.6857)\left(\frac{21}{6}\right)$$

$$= 5.5 + 5.89995 = \mathbf{11.3995}$$

Therefore, our equation of the line of best fit is:

$$\hat{Y} = 11.3995 - 1.6857b_1$$

Determine how much of Y is explained by X

Coefficient of determination

$$r^2 = \frac{b_1^2\left[\sum X^2 - \frac{(\sum X)^2}{N}\right]}{\sum Y^2 - \frac{(\sum Y)^2}{N}}$$

where $0 \le r^2 \le 1$

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.964108919
R Square	0.929506008
Adjusted R Square	0.91188251
Standard Error	0.971008312
Observations	6

ANOVA

	df	SS	MS	F	Significance F
Regression	1	49.728571	49.728571	52.742424	0.001909
Residual	4	3.771429	0.942857		
Total	5	53.5			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	11.4	0.903960	12.611184	0.000228	8.890206	13.909794	8.890206	13.909794
X	-1.68571428	0.232115	-7.262398	0.001909	-2.330170	-1.041259	-2.330170	-1.041259

Note: That the correct reading of the printout is r which is the coefficient of correlation is -0.96411 because b_1 is -1.6857.

R^2 is the coefficient of determination and it is usually between 0 and 1. R^2 of 0.9 implies the 90% of variation in the dependent variable is explained by the variable X.

8. Using the information below to answer the following ques

		X	Y
1	2013Q1	0.694575	98.39
2	2013Q2	0.676784	99.28
3	2013Q3	0.649949	99.88
4	2013Q4	0.692	98.92
5	2014Q1	0.695289	100.26
6	2014Q2	0.667741	99.33
7	2014Q3	0.652718	100.33

SUMMARY OUTPUT

Regression Statistics		
Multiple R	0.52067	Co-efficient Correlation
R Square	0.27109	
Adjusted R Square	0.12531	
Standard Error	0.66734	how much population mean differs from sample mean
Observations	7	

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.82819	0.82819	1.85963	0.23085
Residual	5	2.22677	0.44535		
Total	6	3.05497			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	112.41301	9.48408	11.85280	7.52722E-05	88.03339	136.79262	88.03339	136.79262
X	-19.13724	14.03347	-1.36368	0.23085	-55.21144	16.93696	-55.21144	16.93696

a. Find and interpret the meaning of the Y intercept b_0.
b. Find and interpret the meaning of the slope b_1.
c. Predict the average value of Y for X = 0.5.

9. Given that the sample size is seven. Use the information below to answer the following questions.

$\sum X = 20$ $\quad\quad$ $\sum X^2 = 58.42$ $\quad\quad$ $\sum Y = 21$ $\quad\quad$ $\sum XY = 62.40$

a. Find and interpret the meaning of the Y intercept b_0.
b. Find and interpret the meaning of the slope b_1.
c. Predict the average value of Y for X = 5.