Correlation and Simple Linear Regression

Week 9

Learning Objectives

LO13-1 Explain the purpose of correlation analysis

LO13-2 Calculate a correlation coefficient to test and interpret the relationship between two variables

LO13-3 Apply regression analysis to estimate the linear relationship between two variables

LO13-4 Evaluate the significance of the slope of the regression equation

LO13-5 Evaluate a regression equation's ability to predict using the standard estimate of the error and the coefficient of determination

LO13-6 Calculate and interpret confidence and prediction intervals

LO13-7 Use a log function to transform a nonlinear relationship

What is Correlation Analysis?

- Used to report the relationship between two variables

> CORRELATION ANALYSIS A group of techniques to measure the relationship between two variables.

- In addition to graphing techniques, we'll develop numerical measures to describe the relationships
- Examples
- Does the amount Healthtex spends per month on training its sales force affect its monthly sales
- Does the number of hours that students study for an exam influence the exam score

Scatter Diagram

- A scatter diagram is a graphic tool used to portray the relationship between two variables

> INDEPENDENT VARIABLE A variable that provides the basis for estimation.
> DEPENDENT VARIABLE The variable that is being predicted or estimated.

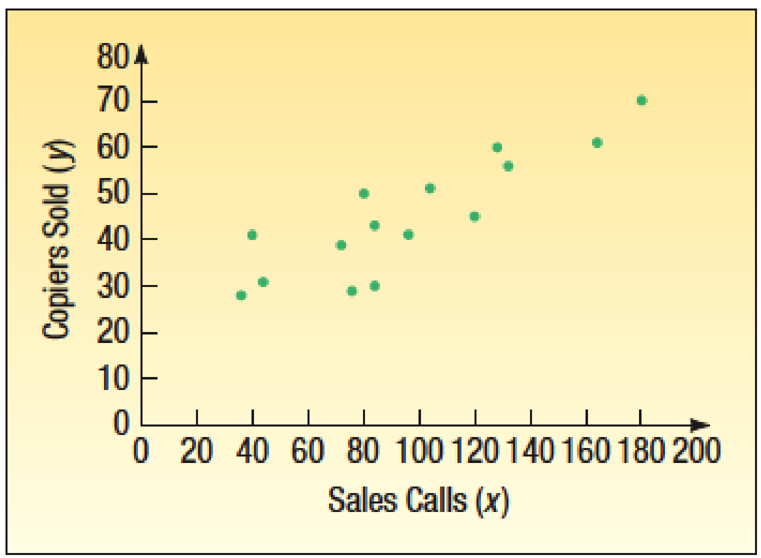
- The independent variable is scaled on the X-axis
- The dependent variable is scaled on the Y-axis

Sales Representative	Sales Calls	Copiers Sold
Brian Virost	96	41
Carlos Ramirez	40	41
Carol Saia	104	51
Greg Fish	128	60
Jeff Hall	164	61
Mark Reynolds	76	29
Meryl Rumsey	72	39
Mike Kiel	80	50
Ray Snarsky	36	28
Rich Niles	84	43
Ron Broderick	180	70
Sal Spina	132	56
Soni Jones	120	45
Susan Welch	44	31
Tom Keller	84	30

Graphing the data in a scatter diagram will make the relationship between sales calls and copiers sales easier to see.

Scatter Diagram Example

North American Copier Sales sells copiers to businesses of all sizes throughout the United States and Canada. The new national sales manager is preparing for an upcoming sales meeting and would like to impress upon the sales representatives the importance of making an extra sales call each day. She takes a random sample of 15 sales representatives and gathers information on the number of sales calls made last month and the number of copiers sold. Develop a scatter diagram of the data.

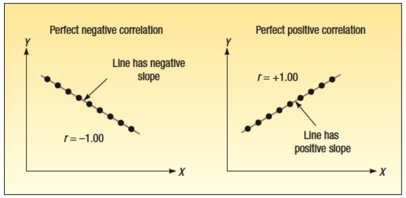


Sales reps who make more calls tend to sell more copiers!

Correlation Coefficient

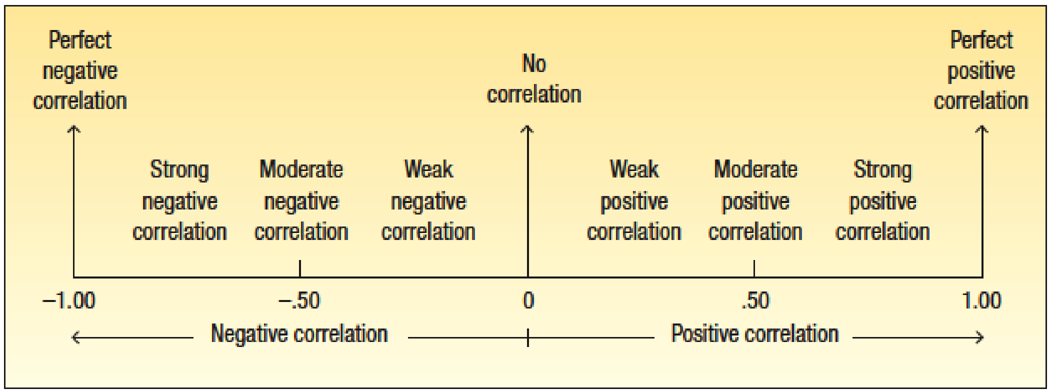
CORRELATION COEFFICIENT A measure of the strength of the linear relationship between two variables.

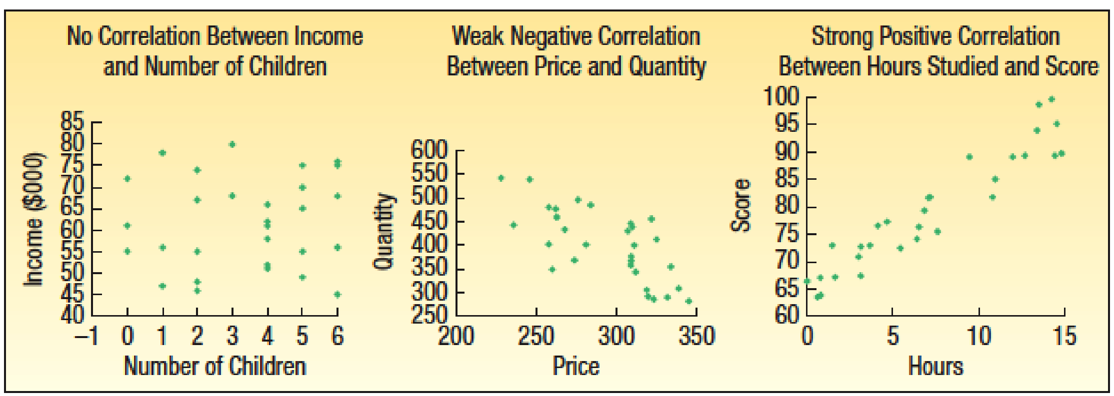
- Characteristics of the correlation coefficient are
  - The sample correlation coefficient is identified as r
  - It shows the direction and strength of the linear relationship between two interval- or ratio-scale variables
  - It ranges from -1.00 to 1.00
  - If it's 0, there is no association
  - A value near 1.00 indicates a direct or positive correlation
  - A value near -1.00 indicates a negative correlation



Correlation Coefficient

- The following graphs summarize the strength and direction of the correlation coefficient

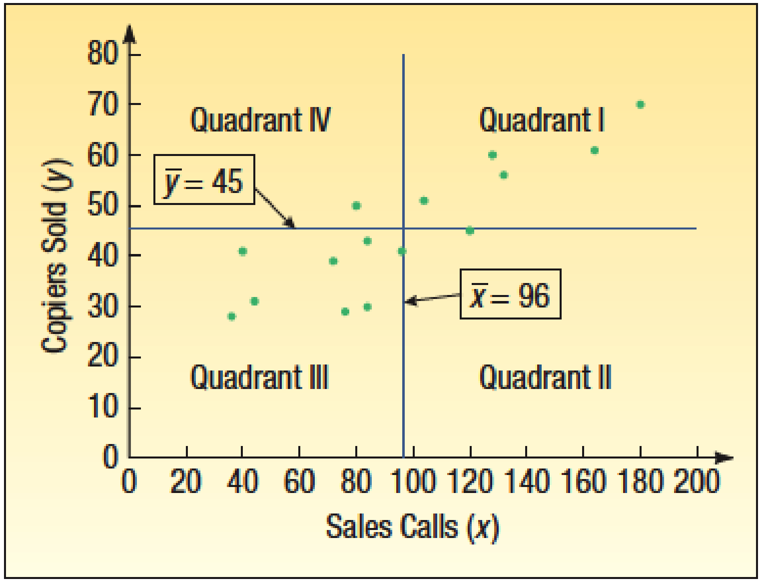




Correlation Coefficient, r

How is the correlation coefficient determined? We'll use the North American Copier Sales as an example. We begin with a scatter diagram, but this time we'll draw a vertical line at the mean of the x-values (96 sales calls) and a horizontal line at the mean of the y-values (45 copiers).

Sales Representative	Sales Calls	Copiers Sold
Brian Virost	96	41
Carlos Ramirez	40	41
Carol Saia	104	51
Greg Fish	128	60
Jeff Hall	164	61
Mark Reynolds	76	29
Meryl Rumsey	72	39
Mike Kiel	80	50
Ray Snarsky	36	28
Rich Niles	84	43
Ron Broderick	180	70
Sal Spina	132	56
Soni Jones	120	45
Susan Welch	44	31
Tom Keller	84	30



Correlation Coefficient, r, Continued

How is the correlation coefficient determined? Now we find the deviations from the mean number of sales calls and the mean number of copiers sold; then multiply the them. The sum of their product is 6,672 and will be used in formula 13-1 to find r. We also need the standard deviations. The result, r=.865 indicates a strong, positive relationship.

CORRELATION COEFFICIENT

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{(n - 1)s_x s_y} \qquad (13\text{–}1)$$

Sales Representative	Sales Calls (x)	Copiers Sold (y)	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$
Brian Virost	96	41	0	−4	0
Carlos Ramirez	40	41	−56	−4	224
Carol Saia	104	51	8	6	48
Greg Fish	128	60	32	15	480
Jeff Hall	164	61	68	16	1,088
Mark Reynolds	76	29	−20	−16	320
Meryl Rumsey	72	39	−24	−6	144
Mike Kiel	80	50	−16	5	−80
Ray Snarsky	36	28	−60	−17	1,020
Rich Niles	84	43	−12	−2	24
Ron Broderick	180	70	84	25	2,100
Sal Spina	132	56	36	11	396
Soni Jones	120	45	24	0	0
Susan Welch	44	31	−52	−14	728
Tom Keller	84	30	−12	−15	180
Totals	1440	675	0	0	6,672

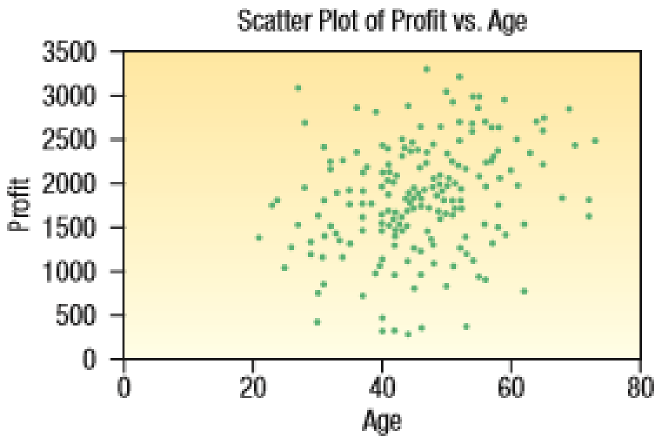
	A	B	C	D
1		Sales Representative	Sales Calls (x)	Copiers Sold (y)
2		Brian Virost	96	41
3		Carlos Ramirez	40	41
4		Carol Saia	104	51
5		Greg Fish	128	60
6		Jeff Hall	164	61
7		Mark Reynolds	76	29
8		Meryl Rumsey	72	39
9		Mike Kiel	80	50
10		Ray Snarsky	36	28
11		Rich Niles	84	43
12		Ron Broderick	180	70
13		Sal Spina	132	56
14		Soni Jones	120	45
15		Susan Welch	44	31
16		Tom Keller	84	30
17		Total	1440	675

F	G: Sales Calls (x)	H: Copiers Sold (y)
Mean	96.00	45.00
Standard Error	11.04	3.33
Median	84.00	43.00
Mode	84.00	41.00
Standard Deviation	42.76	12.89
Sample Variance	1828.57	166.14
Kurtosis	-0.32	-0.73
Skewness	0.46	0.36
Range	144.00	42.00
Minimum	36.00	28.00
Maximum	180.00	70.00
Sum	1440.00	675.00
Count	15.00	15.00

$$r = \frac{6672}{(15-1)(42.76)(12.89)} = 0.865$$

Correlation Coefficient Example

The Applewood Auto Group's marketing department believes younger buyers purchase vehicles on which lower profits are earned and older buyers purchase vehicles on which higher profits are earned. They would like to use this information as part of an upcoming advertising campaign to try to attract older buyers. Develop a scatter diagram and then determine the correlation coefficient. Would this be a useful advertising feature?



	Age	Profit
Age	1	
Profit	0.262	1

The scatter diagram suggests that a positive relationship does exist between age and profit. But it does not appear to be a strong relationship.

Next, calculate r, it is 0.262. The relationship is positive but weak. The data does not support a business decision to create an advertising campaign to attract older buyers!

Testing the Significance of r

- Recall that the sales manager from North American Copier Sales found an r of 0.865
- Could the result be due to sampling error? Remember only 15 salespeople were sampled
- We ask the question, could there be zero correlation in the population from which the sample was selected?
- We'll let ρ represent the correlation in the population and conduct a hypothesis test to find out

Testing the Significance of r Example

Step 1: State the null and the alternate hypothesis

H_0: $\rho = 0$ The correlation in the population is zero

H_1: $\rho \neq 0$ The correlation in the population is different from zero

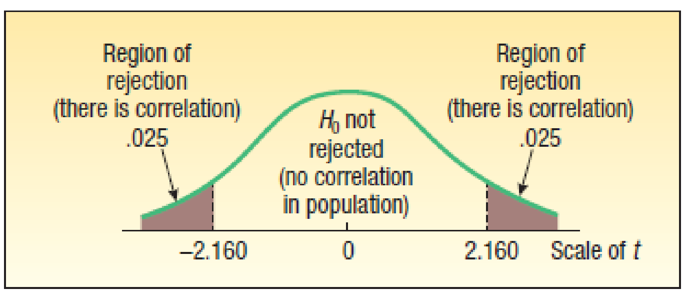
Step 2: Select the level of significance, we'll use .05

Step 3: Select the test statistic, we use t

Step 4: Formulate the decision rule, reject H_0 if $t < -2.160$ or > 2.160

Step 5: Make decision, reject H_0, $t=6.216$

Step 6: Interpret, there is correlation with respect to the number of sales calls made and the number of copiers sold in the population of salespeople.



Decision Rule for Test of Hypothesis at .05 Significance Level and 13 *df*

***t* TEST FOR THE CORRELATION COEFFICIENT**

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \text{ with } n-2 \text{ degrees of freedom} \tag{13–2}$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{.865\sqrt{15-2}}{\sqrt{1-.865^2}} = 6.216$$

Testing the Significance of the Correlation Coefficient

In the Applewood Auto Group example, we found an r=0.262 which is positive, but rather weak. We test our conclusion by conducting a hypothesis test that the correlation is greater than 0.

Step 1: State the null and the alternate hypothesis

H_0: $\rho \le 0$ The correlation in the population is negative or zero

H_1: $\rho > 0$ The correlation in the population is positive

Step 2: Select the level of significance, we'll use .05

Step 3: Select the test statistic, we use t

Step 4: Formulate the decision rule, reject H_0 if t > 1.653

Step 5: Make decision, reject H_0, t=3.622

Step 6: Interpret, there is correlation with respect to profits and age of the buyer

***t* TEST FOR THE CORRELATION COEFFICIENT**

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \text{ with } n-2 \text{ degrees of freedom} \qquad (13\text{–}2)$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.262\sqrt{180-2}}{\sqrt{1-0.262^2}} = 3.622$$

Exercise

- The following sample of observations was randomly selected

x	4	5	3	6	10
y	4	6	5	7	7

- Determine the correlation coefficient and interpret the relationship between x and y.

Solution

X	Y	$(X-\overline{X})$	$(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$	$(X-\overline{X})(Y-\overline{Y})$
4	4	-1.6	-1.8	2.56	3.24	2.88
5	6	-0.6	0.2	0.36	0.04	-0.12
3	5	-2.6	-0.8	6.76	0.64	2.08
6	7	0.4	1.2	0.16	1.44	0.48
10	7	4.4	1.2	19.36	1.44	5.28
28	29			29.2	6.8	10.6

$$\overline{X} = \frac{28}{5} = 5.6 \qquad \overline{Y} = \frac{29}{5} = 5.8 \qquad s_x = \sqrt{\frac{29.2}{4}} = 2.7$$

$$s_y = \sqrt{\frac{6.8}{4}} = 1.3 \qquad r = \frac{10.6}{(5-1)(2.7)(1.3)} = 0.75$$

The 0.75 coefficient indicates a rather strong positive correlation between X and Y.

Exercise

- The following hypotheses are given

$$H_0: \rho = 0$$
$$H_1: \rho > 0$$

A random sample of 12 paired observations indicated that a correlation of 0.32. Can we conclude that the correlation in the population is greater than zero ? Use the 0.05 significance level.

Exercise

- The following hypotheses are given

$$H_0: \rho = 0$$
$$H_1: \rho < 0$$

A random sample of 15 paired observations indicated that a correlation of -0.46. Can we conclude that the correlation in the population is less than zero ? Use the 0.05 significance level.

Answer

Reject H_o if $t > 1.812$	$t = \frac{0.32\sqrt{12-2}}{\sqrt{1-(0.32)^2}} = 1.07$	Do not reject H_o. **(LO13-2)**
Reject H_o if $t < -1.771$	$t = \frac{-0.46\sqrt{15-2}}{\sqrt{1-(-0.46)^2}} = -1.868$	Reject H_o. **(LO13-2)**

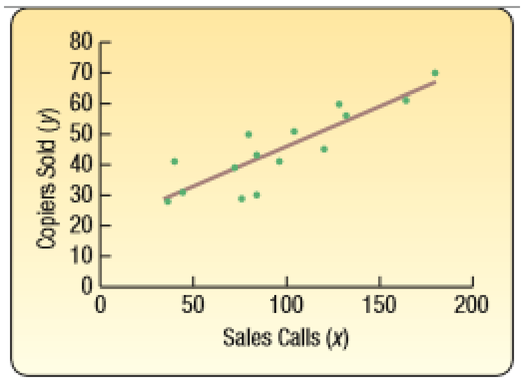
Regression Analysis

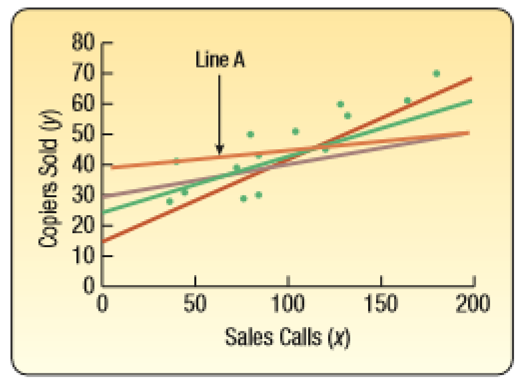
- In regression analysis, we estimate one variable based on another variable
- The variable being estimated is the dependent variable
- The variable used to make the estimate or predict the value is the independent variable
- The relationship between the variables is linear
- Both the independent and the dependent variables must be interval or ratio scale

REGRESSION EQUATION An equation that expresses the linear relationship between two variables.

Least Squares Principle

- In regression analysis, our objective is to use the data to position a line that best represents the relationship between two variables
- The first approach is to use a scatter diagram to visually position the line





- But this depends on judgement, we would prefer a method that results in a single, best regression line

Least Squares Regression Line

LEAST SQUARES PRINCIPLE A mathematical procedure that uses the data to position a line with the objective of minimizing the sum of the squares of the vertical distances between the actual y values and the predicted values of y.

- To illustrate, the same data are plotted in the three charts below

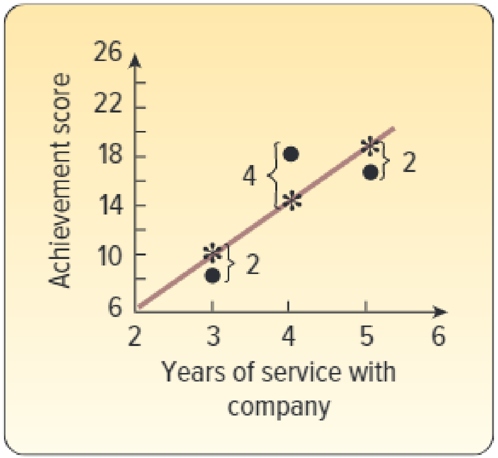


CHART 13–9 The Least Squares Line

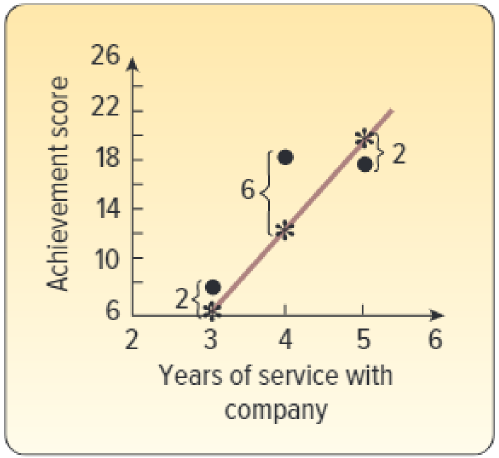


CHART 13–10 Line Drawn with a Straight Edge

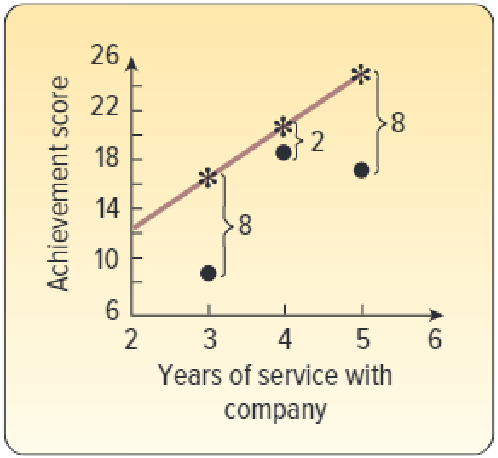


CHART 13–11 Different Line Drawn with a Straight Edge

Least Squares Regression Line

- This is the equation of a line

GENERAL FORM OF LINEAR REGRESSION EQUATION	$\hat{y} = a + bx$	(13–3)

  - $\hat{y}$ is the estimated value of y for a selected value of x
  - a is the constant or intercept
  - b is the slope of the fitted line
  - x is the value of the independent variable
- The formulas for a and b are

SLOPE OF THE REGRESSION LINE	$b = r\left(\frac{s_y}{s_x}\right)$	(13–4)

Y-INTERCEPT	$a = \bar{y} - b\bar{x}$	(13–5)

Least Squares Regression Line Example

Recall the example of North American Copier Sales. The sales manager gathered information on the number of sales calls made and the number of copiers sold. Use the least squares method to determine a linear equation to express the relationship between the two variables.

The first step is to find the slope of the least squares regression line, b

$$b = r\left(\frac{s_y}{s_x}\right) = .865\left(\frac{12.89}{42.76}\right) = 0.2608$$

Next, find a

$$a = \bar{y} - b\bar{x} = 45 - .2608(96) = 19.9632$$

Then determine the regression line

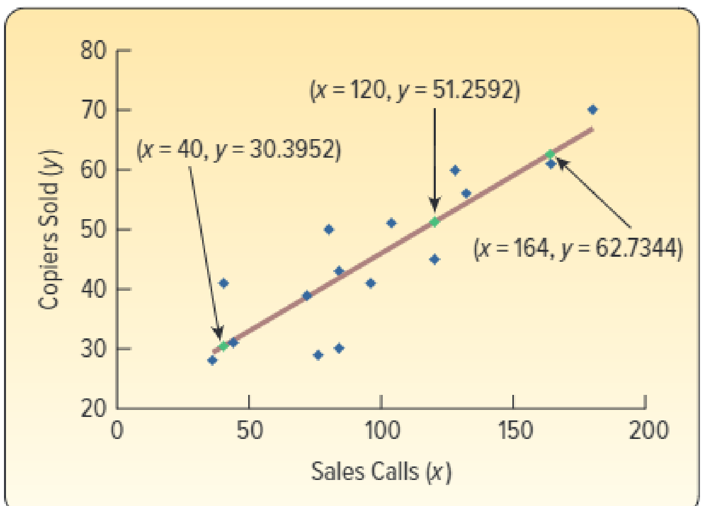
$$\hat{y} = 19.9632 + 0.2608x.$$

So if a salesperson makes 100 calls, he or she can expect to sell 46.0432 copiers

$$\hat{y} = 19.9632 + 0.2608x = 19.9632 + 0.2608(100) = 46.0432$$

Drawing the Regression Line

The least squares equation can be drawn on the scatter diagram. For example, the fifth sales representative is Jeff Hall. He made 164 calls. His estimated number of copiers sold is 62.7344. The plot x = 164 and $\hat{y}$ = 62.7344 is located by moving to 164 on the x-axis and then going vertically to 62.7344. The other points on the regression equation can be determined by substituting a particular value of x into the regression equation and calculating $\hat{y}$.



Sales Representative	Sales Calls (x)	Copiers Sold (y)	Estimated Sales ($\hat{y}$)
Brian Virost	96	41	45.0000
Carlos Ramirez	40	41	30.3952
Carol Saia	104	51	47.0864
Greg Fish	128	60	53.3456
Jeff Hall	164	61	62.7344
Mark Reynolds	76	29	39.7840
Meryl Rumsey	72	39	38.7408
Mike Kiel	80	50	40.8272
Ray Snarsky	36	28	29.3520
Rich Niles	84	43	41.8704
Ron Broderick	180	70	66.9072
Sal Spina	132	56	54.3888
Soni Jones	120	45	51.2592
Susan Welch	44	31	31.4384
Tom Keller	84	30	41.8704

The least squares equation can be drawn on the scatter diagram. For example, the fifth sales representative is Jeff Hall. He made 164 calls. His estimated number of copiers sold is 62.7344. The plot x = 164 and $\hat{y}$ = 62.7344 is located by moving to 164 on the x-axis and then going vertically to 62.7344. The other points on the regression equation can be determined by substituting a particular value of x into the regression equation and calculating $\hat{y}$.

Exercise

- The following sample of observations was randomly selected.

x	4	5	3	6	10
y	4	6	5	7	7

- (a) Determine the regression equation.
- (b) Determine the value of y when x is 7.

Answer

a.

X	Y	$(X-\overline{X})$	$(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$	$(X-\overline{X})(Y-\overline{Y})$
4	4	-1.6	-1.8	2.56	3.24	2.88
5	6	-0.6	0.2	0.36	0.04	-0.12
3	5	-2.6	-0.8	6.76	0.64	2.08
6	7	0.4	1.2	0.16	1.44	0.48
10	7	4.4	1.2	19.36	1.44	5.28
28	29			29.2	6.8	10.6

$$\overline{X} = \frac{28}{5} = 5.6 \qquad \overline{Y} = \frac{29}{5} = 5.8 \qquad s_x = \sqrt{\frac{29.2}{4}} = 2.702$$

$$s_y = \sqrt{\frac{6.8}{4}} = 1.304 \qquad r = \frac{10.6}{(5-1)(2.702)(1.304)} = 0.752$$

$$b = \frac{(0.752)(1.304)}{2.702} = .363 \qquad a = 5.8 - (.363)(5.6) = 3.767$$

$$\hat{Y} = 3.767 + 0.363X$$

b. 6.3081, found by $\hat{Y} = 3.7671 + 0.3630(7)$ **(LO13-3)**

Regression Equation Slope Test

- For a regression equation, the slope is tested for significance
- We test the hypothesis that the slope of the line in the population is 0
- If we do not reject the null hypothesis, we conclude there is no relationship between the two variables
- When testing the null hypothesis about the slope, the test statistic is with n – 2 degrees of freedom
- We begin with the hypothesis statements

$$H_0: \ \beta = 0$$
$$H_1: \ \beta \neq 0$$

Regression Equation Slope Test Example

Recall the North American Copier Sales example. We identified the slope as b and it is our estimate of the slope of the population, β. We conduct a hypothesis test.

Step 1: State the null and alternate hypothesis

H_0: $\beta \leq 0$

H_1: $\beta > 0$

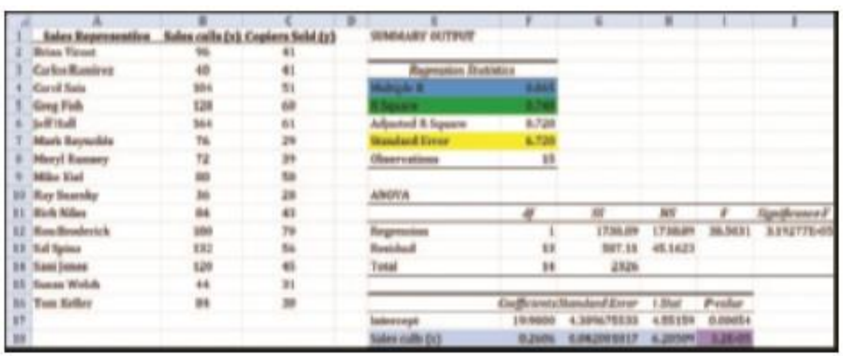
Step 2: Select the level of significance, we use .05

Step 3: Select the test statistic, t

Step 4: Formulate the decision rule, reject H_0 if t > 1.771

Step 5: Make decision, reject H_0, t = 6.205

Step 6: Interpret, the number of sales calls is useful in estimating copier sales



Highlighted, b is .2606; the standard error is .0420

TEST FOR THE SLOPE $t = \frac{b - 0}{s_b}$ with $n - 2$ degrees of freedom **(13–6)**

$$t = \frac{b - 0}{s_b} = \frac{0.2606 - 0}{0.042} = 6.205$$

Exercise

- The regression equation is $y=29.29 - 0.96x$, the sample size is 8, and the standard error for the slope is 0.22. Use the 0.05 significance level. Can we conclude that the slope of the regression line is less than zero ?

Answer

H_o: $\beta \geq 0$ H_1: $\beta < 0$ df = n – 2 = 8 – 2 = 6 Reject H_o if $t < -1.943$
$t = -0.96/0.22 = -4.364$ Reject H_o and conclude the slope is less than zero.

Evaluating a Regression Equation's Ability to Predict

- Perfect prediction is practically impossible in almost all disciplines, including economics and business
- The North American Copier Sales example showed a significant relationship between sales calls and copier sales, the equation is

$$\text{Number of copiers sold} = 19.9632 + .2608(\text{Number of sales calls})$$

- What if the number of sales calls is 84, we calculate the number of copiers sold is 41.8704—we did have two employees with 84 sales calls, they sold just 30 and 24
- So, is the regression equation a good predictor?
- We need a measure that will tell how inaccurate the estimate might be

The Standard Error of Estimate

- The standard error of estimate measures the variation around the regression line

STANDARD ERROR OF ESTIMATE A measure of the dispersion, or scatter, of the observed values around the line of regression for a given value of x.

- It is in the same units as the dependent variable
- It is based on squared deviations from the regression line
- Small values indicate that the points cluster closely about the regression line
- It is computed using the following formula

STANDARD ERROR OF ESTIMATE

$$s_{y \cdot x} = \sqrt{\frac{\Sigma(y - \hat{y})^2}{n - 2}} \qquad (13\text{–}7)$$

The Standard Error of Estimate Example

We calculate the standard error of estimate in this example. We need the sum of the squared differences between each observed value of y and the predicted value of y, which is $\hat{y}$. We use a spreadsheet to help with the calculations.

	A	B	C	D	E	F
1	**Sales Rep**	**Sales Calls (x)**	**Copiers Sold (y)**	**Estimated Sales**	**$(y-\hat{y})$**	**$(y-\hat{y})^2$**
2	Brian Virost	96	41	45.0000	–4.0000	16.0000
3	Carlos Ramirez	40	41	30.3952	10.6048	112.4618
4	Carol Saia	104	51	47.0864	3.9136	15.3163
5	Greg Fish	128	60	53.3456	6.6544	44.2810
6	Jeff Hall	164	61	62.7344	–1.7344	3.0081
7	Mark Reynolds	76	29	39.7840	–10.7840	116.2947
8	Meryl Rumsey	72	39	38.7408	0.2592	0.0672
9	Mike Kiel	80	50	40.8272	9.1728	84.1403
10	Ray Snarsky	36	28	29.3520	–1.3520	1.8279
11	Rich Niles	84	43	41.8704	1.1296	1.2760
12	Ron Broderick	180	70	66.9072	3.0928	9.5654
13	Sal Spina	132	56	54.3888	1.6112	2.5960
14	Soni Jones	120	45	51.2592	–6.2592	39.1776
15	Susan Welch	44	31	31.4384	–0.4384	0.1922
16	Tom Keller	84	30	41.8704	–11.8704	140.9064
17	**Total**				**0.0000**	**587.1108**

The standard error of estimate is 6.720

$$s_{y \cdot x} = \sqrt{\frac{\Sigma(y-\hat{y})^2}{n-2}} = \sqrt{\frac{587.1108}{15-2}} = 6.720$$

If the standard error of estimate is small, this indicates that the data are relatively close to the regression line and the regression equation can be used. If it is large, the data are widely scattered around the regression line and the regression equation will not provide a precise estimate of y.

Coefficient of Determination

COEFFICIENT OF DETERMINATION The proportion of the total variation in the dependent variable Y that is explained, or accounted for, by the variation in the independent variable X.

- It ranges from 0 to 1.0
- It is the square of the correlation coefficient
- It is found from the following formula

COEFFICIENT OF DETERMINATION $$r^2 = \frac{\text{SSR}}{\text{SS Total}} = 1 - \frac{\text{SSE}}{\text{SS Total}} \quad \textbf{(13–8)}$$

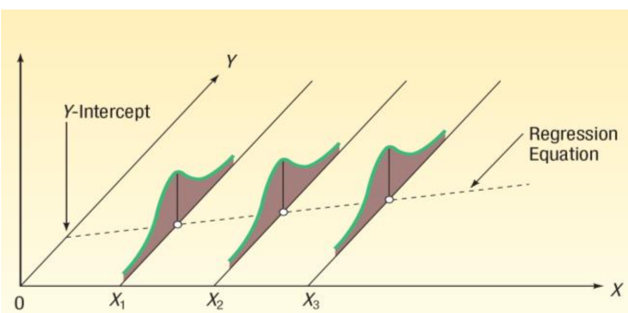
- In the North American Copier Sales example, the correlation coefficient was .865; just square that $(.865)^2 = .748$; this is the coefficient of determination
- This means 74.8% of the variation in the number of copiers sold is explained by the variation in the number of sales calls

Relationships among r, r_2, and $s_{y,x}$

- Recall the standard error of estimate measures how close the actual values are to the regression line
  - When it is small, the two variables are closely related
- The correlation coefficient measures the strength of the linear association between two variables
  - When points on the scatter diagram are close to the line, the correlation coefficient tends to be large
- Therefore, the correlation coefficient and the standard error of estimate are inversely related
- As noted earlier, the coefficient of determination is the correlation coefficient squared

Inference about Linear Regression

- We can predict the number of copiers sold (y) for a selected value of number of sales calls made (x)
- But first, let's review the regression assumptions of each of the distributions in the graph below
  - Follow the normal distribution
  - Has a mean on the regression line
  - Has the same standard error of estimate, $s_{y \cdot x}$
  - Is independent of the others



$\hat{y} \pm s_{y \cdot x}$ will include 68% of the observations
$\hat{y} \pm 2s_{y \cdot x}$ will include 95% of the observations
$\hat{y} \pm 3s_{y \cdot x}$ will include virtually all the observations

Constructing Confidence and Prediction Intervals

- Use a confidence interval when the regression equation is used to predict the mean value of y for a given value of x
- For instance, we would use a confidence interval to estimate the mean salary of all executives in the retail industry based on their years of experience

CONFIDENCE INTERVAL FOR THE MEAN OF *Y*, GIVEN *X*

$$\hat{y} \pm t s_{y \cdot x} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\Sigma(x - \bar{x})^2}} \qquad (13\text{–}11)$$

- Use a prediction interval when the regression equation is used to predict an individual y for a given value of x
- For instance, we would estimate the salary of a particular retail executive who has 20 years of experience

PREDICTION INTERVAL FOR *Y*, GIVEN *X*

$$\hat{y} \pm t s_{y \cdot x} \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\Sigma(x - \bar{x})^2}} \qquad (13\text{–}12)$$

Confidence Interval and Prediction Interval Example

We return to the North American Copier Sales example. Determine a 95% confidence interval for all sales representatives who make 50 calls, and determine a prediction interval for Sheila Baker, a west coast sales representative who made 50 sales calls.

Sales Representative	Sales Calls (x)	Copiers Sold (y)	$(x - \bar{x})$	$(x - \bar{x})^2$
Brian Virost	96	41	0	0
Carlos Ramirez	40	41	−56	3,136
Carol Saia	104	51	8	64
Greg Fish	128	60	32	1,024
Jeff Hall	164	61	68	4,624
Mark Reynolds	76	29	−20	400
Meryl Rumsey	72	39	−24	576
Mike Kiel	80	50	−16	256
Ray Snarsky	36	28	−60	3,600
Rich Niles	84	43	−12	144
Ron Broderick	180	70	84	7,056
Sal Spina	132	56	36	1,296
Soni Jones	120	45	24	576
Susan Welch	44	31	−52	2,704
Tom Keller	84	30	−12	144
Total	1440	675	0	25,600

$$\hat{y} = 19.9632 + 0.2608x = 19.9632 + 0.2608(50) = 33.0032$$

$$\text{Confidence Interval} = \hat{y} \pm t s_{y \cdot x}\sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\Sigma(x - \bar{x})^2}}$$
$$= 33.0032 \pm 2.160(6.720)\sqrt{\frac{1}{15} + \frac{(50 - 96)^2}{25{,}600}}$$
$$= 33.0032 \pm 5.6090$$

$$\text{Prediction Interval} = \hat{y} \pm t s_{y \cdot x}\sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\Sigma(x - \bar{x})^2}}$$
$$= 33.0032 \pm 2.160(6.720)\sqrt{1 + \frac{1}{15} + \frac{(50 - 96)^2}{25{,}600}}$$
$$= 33.0032 \pm 15.5612$$

The 95% confidence interval for all sales representatives is 27.3942 up to 38.6122
The 95% prediction interval for Sheila Baker is 17.442 up to 48.5644 copiers