Chapter 12
Simple Regression Analysis and Correlation

LEARNING OBJECTIVES

The overall objective of this chapter is to give you an understanding of bivariate linear regression analysis, thereby enabling you to:

1. Calculate the Pearson product-moment correlation coefficient to determine if there is a correlation between two variables.
2. Explain what regression analysis is and the concepts of independent and dependent variable.
3. Calculate the slope and *y*-intercept of the least squares equation of a regression line and from those, determine the equation of the regression line.
4. Calculate the residuals of a regression line and from those determine the fit of the model, locate outliers, and test the assumptions of the regression model.
5. Calculate the standard error of the estimate using the sum of squares of error, and use the standard error of the estimate to determine the fit of the model.
6. Calculate the coefficient of determination to measure the fit for regression models, and relate it to the coefficient of correlation.
7. Use the *t* and *F* tests to test hypotheses for both the slope of the regression model and the overall regression model.
8. Calculate confidence intervals to estimate the conditional mean of the dependent variable and prediction intervals to estimate a single value of the dependent variable.
9. Determine the equation of the trend line to forecast outcomes for time periods in the future, using alternate coding for time periods if necessary.
10. Use a computer to develop a regression analysis, and interpret the output that is associated with it.

CHAPTER OUTLINE

12.1 Correlation

12.2 Introduction to Simple Regression Analysis

12.3 Determining the Equation of the Regression Line

12.4 Residual Analysis
- Using Residuals to Test the Assumptions of the Regression Model
- Using the Computer for Residual Analysis

12.5 Standard Error of the Estimate

12.6 Coefficient of Determination
- Relationship Between r and r^2

12.7 Hypothesis Tests for the Slope of the Regression Model and Testing the Overall Model
- Testing the Slope
- Testing the Overall Model

12.8 Estimation
- Confidence Intervals to Estimate the Conditional Mean of y: $\mu_{y|x}$
- Prediction Intervals to Estimate a Single Value of y

12.9 Using Regression to Develop a Forecasting Trend Line
- Determining the Equation of the Trend Line
- Forecasting Using the Equation of the Trend Line
- Alternate Coding for Time Periods

12.10 Interpreting Computer Output

KEY TERMS

Coefficient of Determination (r^2)
Confidence Interval
Dependent Variable
Deterministic Model
Heteroscedasticity
Homoscedasticity
Independent Variable
Least Squares Analysis
Outliers
Prediction Interval
Probabilistic Model
Regression Analysis
Residual
Residual Plot
Scatter Plot
Simple Regression
Standard Error of the Estimate (s_e)
Sum of Squares of Error (SSE)

STUDY QUESTIONS

1. _______________ is a measure of the degree of relatedness of two variables.

2. The Pearson product-moment correlation coefficient is denoted by _______________.

3. The value of r varies from _________________________.

4. Perfect positive correlation results in an r value of _______________.

5. The value of the coefficient of correlation from the following data is _______________.

   x: 19, 20, 26, 31, 34, 45, 45, 51
   y: 78, 100, 125, 120, 119, 130, 145, 143

6. The value of r from the following data is __________.

   x: -10, -6, 1, 4, 15
   y: -26, -44, -36, -39, -43

7. The process of constructing a mathematical model or function that can be used to predict or determine one variable by another variable is _______________.

8. Bivariate linear regression is often termed _______________ regression.

9. In regression, the variable being predicted is usually referred to as the _______________ variable.

10. In regression, the predictor is called the _______________ variable.

11. The first step in simple regression analysis often is to graph or construct a _______________.

12. In regression analysis, β_1 represents the population _______________.

13. In regression analysis, b_o represents the sample _______________.

14. A researcher wants to develop a regression model to predict the price of gold by the prime interest rate. The dependent variable is _______________.

15. In an effort to develop a regression model, the following data were gathered:

    x: 2, 9, 11, 19, 21, 25
    y: 26, 17, 18, 15, 15, 8

    The slope of the regression line determined from these data is ____________.
    The y intercept is ____________.

16. A researcher wants to develop a regression line from the data given below:

x: 12, 11, 5, 6, 9
y: 31, 25, 14, 12, 16

The equation of the regression line is ____________________.

17. In regression, the value of $y - \hat{y}$ is called the _______________.

18. Data points that lie apart from the rest of the points are called _______________.

19. The regression assumption of constant error variance is called _______________.

If the error variances are not constant, it is called _______________.

20. Suppose the following data are used to determine the equation of the regression line given below:

x: 2, 5, 11, 24, 31
y: 12, 13, 16, 14, 19

$\hat{y}$ = 12.224 + 0.1764 x

The residual for x = 11 is ____________________.

21. The total of the residuals squared is called the ______________________________.

22. A standard deviation of the error of the regression model is called the __________________________________ and is denoted by _______________.

23. Suppose a regression model is developed for ten pairs of data resulting in S.S.E. = 1,203. The standard error of the estimate is _______________.

24. A regression analysis results in the following data:

$\Sigma x = 276$ $\Sigma x^2 = 12{,}014$ $\Sigma xy = 2{,}438$
$\Sigma y = 77$ $\Sigma y^2 = 1{,}183$ $n = 7$

The value of SSE is _______________.

25. The value of S_e is computed from the data of question 24 is _______________.

26. Suppose a regression model results in a value of $s_e = 27.9$. 95% of the residuals should fall within _______________.

27. Coefficient of determination is denoted by __________.

28. _________________________ is the proportion of variability of the dependent variable accounted for or explained by the independent variable.

29. The value of r^2 always falls between __________ and __________ inclusive.

30. Suppose a regression analysis results in the following:

$b_1 =$ .19364 $\Sigma y =$ 1,019
$b_0 =$ 59.4798 $\Sigma y^2 =$ 134,451
$n =$ 8 $\Sigma xy =$ 378,932

The value of r^2 for this regression model is ________.

31. Suppose the data below are used to determine the equation of a regression line:

x: 18, 14, 9, 6, 2
y: 14, 25, 22, 23, 27

The value of r^2 associated with this model is __________.

32. A researcher has developed a regression model from sixteen pairs of data points. He wants to test to determine if the slope is significantly different from zero. He uses a two-tailed test and $\alpha = .01$. The critical table t value is _______________.

33. The following data are used to develop a simple regression model:

x: 22, 20, 15, 15, 14, 9
y: 31, 20, 12, 9, 10, 6

The observed t value used to test the slope of this regression model is _______________.

34. If $\alpha = .05$ and a two-tailed test is being conducted, the critical table t value to test the slope of the model developed in question 33 is _______________.

35. The decision reached about the slope of the model computed in question 33 is to _______________ the null hypothesis.

36. The equation of the trend line through the following sales data is ____________________; and using this trend line, the predicted sales for year 10 is _________ .

Year	Sales
1	230
2	246
3	251
4	254
5	272
6	283
7	299

ANSWERS TO STUDY QUESTIONS

1. Correlation
2. *r*
3. –1 to 0 to +1
4. +1
5. .876
6. -.581
7. Regression
8. Simple
9. Dependent
10. Independent
11. Scatter Plot
12. Slope
13. *y* Intercept
14. Price of Gold
15. -0.626, 25.575
16. -1.253 + 2.425 x
17. Residual
18. Outliers
19. Homoscedasticity, Heteroscadasticity
20. 1.8356
21. Sum of Squares of Error
22. Standard Error of the Estimate, s_e
23. 12.263
24. 20.015
25. 2.00
26. 0 $\pm$ 55.8
27. r^2
28. Coefficient of Determination
29. 0, 1
30. .900
31. .578
32. 2.977
33. 4.72
34. $\pm$ 2.776
35. Reject
36. 219 + 10.7857x
    326.857

SOLUTIONS TO THE ODD-NUMBERED PROBLEMS IN CHAPTER 12

12.1 $\Sigma x = 80$ $\Sigma x^2 = 1{,}148$ $\Sigma y = 69$

$\Sigma y^2 = 815$ $\Sigma xy = 624$ $n = 7$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} =$$

$$r = \frac{624 - \frac{(80)(69)}{7}}{\sqrt{\left[1{,}148 - \frac{(80)^2}{7}\right]\left[815 - \frac{(69)^2}{7}\right]}} = \frac{-164.571}{\sqrt{(233.714)(134.857)}} =$$

$$r = \frac{-164.571}{177.533} = \mathbf{-0.927}$$

12.3

Delta (x)	SW (y)
47.6	15.1
46.3	15.4
50.6	15.9
52.6	15.6
52.4	16.4
52.7	18.1

$\Sigma x = 302.2$ $\Sigma y = 96.5$ $\Sigma xy = 4{,}870.11$

$\Sigma x^2 = 15{,}259.62$ $\Sigma y^2 = 1{,}557.91$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} =$$

$$r = \frac{4{,}870.11 - \frac{(302.2)(96.5)}{6}}{\sqrt{\left[15{,}259.62 - \frac{(302.2)^2}{6}\right]\left[1{,}557.91 - \frac{(96.5)^2}{6}\right]}} = \mathbf{.6445}$$

12.5 Correlation between Year 1 and Year 2:

$\Sigma x = 17.09$ $\Sigma x^2 = 58.7911$
$\Sigma y = 15.12$ $\Sigma y^2 = 41.7054$
$\Sigma xy = 48.97$ $n = 8$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} =$$

$$r = \frac{48.97 - \frac{(17.09)(15.12)}{8}}{\sqrt{\left[58.7911 - \frac{(17.09)^2}{8}\right]\left[41.7054 - \frac{(15.12)^2}{8}\right]}} =$$

$$r = \frac{16.6699}{\sqrt{(22.28259)(13.1286)}} = \frac{16.6699}{17.1038} = \mathbf{.975}$$

Correlation between Year 2 and Year 3:

$\Sigma x = 15.12$ $\Sigma x^2 = 41.7054$
$\Sigma y = 15.86$ $\Sigma y^2 = 42.0396$
$\Sigma xy = 41.5934$ $n = 8$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} =$$

$$r = \frac{41.5934 - \frac{(15.12)(15.86)}{8}}{\sqrt{\left[41.7054 - \frac{(15.12)^2}{8}\right]\left[42.0396 - \frac{(15.86)^2}{8}\right]}} =$$

$$r = \frac{11.618}{\sqrt{(13.1286)(10.59715)}} = \frac{11.618}{11.795} = \mathbf{.985}$$

Correlation between Year 1 and Year 3:

$\Sigma x = 17.09$ $\Sigma x^2 = 58.7911$
$\Sigma y = 15.86$ $\Sigma y^2 = 42.0396$
$\Sigma xy = 48.5827$ $n = 8$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} =$$

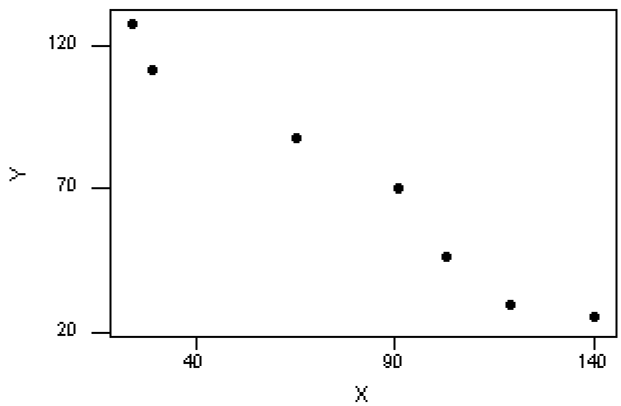
$$r = \frac{48.5827 - \frac{(17.09)(15.86)}{8}}{\sqrt{\left[58.7911 - \frac{(17.09)^2}{8}\right]\left[42.0396 - \frac{(15.86)^2}{8}\right]}}$$

$$r = \frac{14.702}{\sqrt{(22.2826)(10.5972)}} = \frac{14.702}{15.367} = \mathbf{.957}$$

The years 2 and 3 are the most correlated with $r = .985$.

12.7

x	y
140	25
119	29
103	46
91	70
65	88
29	112
24	128



$\Sigma x = 571$ $\Sigma y = 498$ $\Sigma xy = 30{,}099$

$\Sigma x^2 = 58{,}293$ $\Sigma y^2 = 45{,}154$ $n = 7$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{30{,}099 - \frac{(571)(498)}{7}}{58{,}293 - \frac{(571)^2}{7}} = \mathbf{-0.898}$$

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{498}{7} - (-0.898)\frac{571}{7} = \mathbf{144.414}$$

$$\hat{y} = \mathbf{144.414 - 0.898}\ \boldsymbol{x}$$

12.9

(Prime) x	(Bond) y
16	5
6	12
8	9
4	15
7	7

$\Sigma x = 41$ $\Sigma y = 48$ $\Sigma xy = 333$

$\Sigma x^2 = 421$ $\Sigma y^2 = 524$ $n = 5$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{333 - \frac{(41)(48)}{5}}{421 - \frac{(41)^2}{5}} = \mathbf{-0.715}$$

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{48}{5} - (-0.715)\frac{41}{5} = \mathbf{15.460}$$

$$\hat{y} = \mathbf{15.460 - 0.715\, x}$$

12.11

No. of Farms (x)	Avg. Size (y)
5.65	213
4.65	258
3.96	297
3.36	340
2.95	374
2.52	420
2.44	426
2.29	441
2.15	460
2.07	469
2.17	434
2.10	444
2.19	419

$\Sigma x = 38.5$ $\Sigma y = 4{,}995$ $\Sigma x^2 = 129.5892$

$\Sigma y^2 = 2{,}000{,}589$ $\Sigma xy = 13{,}684.32$ $n = 13$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{13{,}684.32 - \frac{(38.5)(4{,}995)}{13}}{129.5892 - \frac{(38.5)^2}{13}} = \mathbf{-71.199}$$

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{4{,}995}{13} - (-71.199)\frac{38.5}{13} = \mathbf{595.089}$$

$\hat{y}$ = **595.089 – 71.199 *x***

12.13

x	y
15	47
8	36
19	56
12	44
5	21

$\hat{y} = 13.625 + 2.303\, x$

Residuals:

x	y	$\hat{y}$	Residuals (y- $\hat{y}$)
15	47	48.1694	-1.1694
8	36	32.0489	3.9511
19	56	57.3811	-1.3811
12	44	41.2606	2.7394
5	21	25.1401	-4.1401

12.15

x	y	Predicted ($\hat{y}$)	Residuals (y- $\hat{y}$)
140	25	18.6597	6.3403
119	29	37.5229	-8.5229
103	46	51.8948	-5.8948
91	70	62.6737	7.3263
65	88	86.0281	1.9720
29	112	118.3648	-6.3648
24	128	122.8561	5.1439

$\hat{y} = 144.414 - 0.898\, x$

12.17

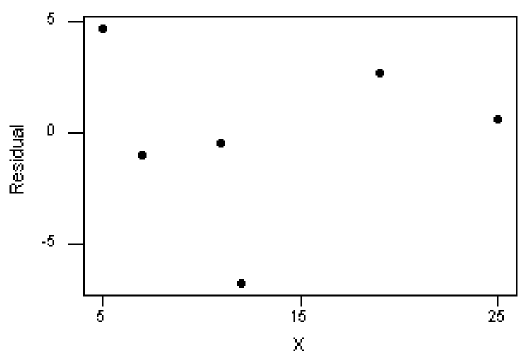
x	y	Predicted ($\hat{y}$)	Residuals (y- $\hat{y}$)
16	5	4.0259	0.9741
6	12	11.1722	0.8278
8	9	9.7429	-0.7429
4	15	12.6014	2.3986
7	7	10.4576	-3.4575

$\hat{y}$ = 15.460 - 0.715 x

12.19

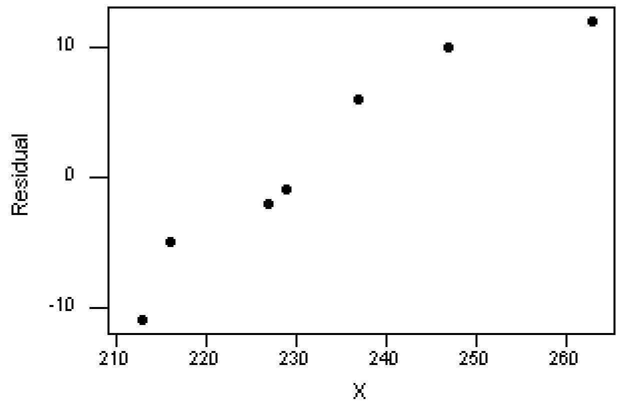
x	y	Predicted ($\hat{y}$)	Residuals (y- $\hat{y}$)
5	47	42.2756	4.7244
7	38	38.9836	-0.9836
11	32	32.3997	-0.3996
12	24	30.7537	-6.7537
19	22	19.2317	2.7683
25	10	9.3558	0.6442

$\hat{y}$ = 50.506 - 1.646 x



No apparent violation of assumptions

12.21



Error terms appear to be non independent

12.23 The Minitab Residuals vs. Fits graphic is strongly indicative of a violation of the homoscedasticity assumption of regression. Because the residuals are very close together for small values of *x*, there is little variability in the residuals at the left end of the graph. On the other hand, for larger values of *x*, the graph flares out indicating a much greater variability at the upper end. Thus, there is a lack of homogeneity of error across the values of the independent variable.

12.25 SSE = $\Sigma y^2 - b_0\Sigma y - b_1\Sigma xy$ = 45,154 - 144.414(498) - (-.89824)(30,099) =

SSE = 272.0

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{272.0}{5}} = \mathbf{7.376}$$

6 out of 7 = 85.7% fall within ± 1s_e
7 out of 7 = 100% fall within ± 2s_e

12.27 SSE = $\Sigma y^2 - b_0\Sigma y - b_1\Sigma xy$ = 524 - 15.46(48) - (-0.71462)(333) **= 19.8885**

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{19.8885}{3}} = \mathbf{2.575}$$

Four out of five (80%) of the estimates are within 2.575 of the actual rate for bonds. This amount of error is probably not acceptable to financial analysts.

12.29

$(y-\hat{y})$	$(y-\hat{y})^2$
4.7244	22.3200
-0.9836	.9675
-0.3996	.1597
-6.7537	45.6125
2.7683	7.6635
0.6442	.4150
	$\Sigma(y-\hat{y})^2$ = 77.1382

$$\text{SSE} = \sum (y-\hat{y})^2 = 77.1382$$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{77.1382}{4}} = \mathbf{4.391}$$

12.31

Volume (x)	Sales (y)
728.6	10.5
497.9	48.1
439.1	64.8
377.9	20.1
375.5	11.4
363.8	123.8
276.3	89.0

n = 7 Σx = 3059.1 Σy = 367.7

Σx^2 = 1,464,071.97 Σy^2 = 30,404.31 Σxy = 141,558.6

b_1 = -.1504 b_0 = 118.257

$\hat{y}$ = 118.257 - .1504x

SSE = $\Sigma y^2 - b_0\Sigma y - b_1\Sigma xy$

= 30,404.31 - (118.257)(367.7) - (-0.1504)(141,558.6) = 8211.6245

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{8211.6245}{5}} = \mathbf{40.526}$$

This is a relatively large standard error of the estimate given the sales values (ranging from 10.5 to 123.8).

12.33 $r^2 = 1 - \dfrac{SSE}{\sum y^2 - \dfrac{(\sum y)^2}{n}} = 1 - \dfrac{272.121}{45{,}154 - \dfrac{(498)^2}{7}} = \mathbf{.972}$

This is a high value of r^2

12.35 $r^2 = 1 - \dfrac{SSE}{\sum y^2 - \dfrac{(\sum y)^2}{n}} = 1 - \dfrac{19.8885}{524 - \dfrac{(48)^2}{5}} = \mathbf{.685}$

This value of r^2 is a modest value.
68.5% of the variation of *y* is accounted for by *x* but 31.5% is unaccounted for.

12.37

CCI	Median Income
116.8	37.415
91.5	36.770
68.5	35.501
61.6	35.047
65.9	34.700
90.6	34.942
100.0	35.887
104.6	36.306
125.4	37.005

$\Sigma x = 323.573$ $\Sigma y = 824.9$ $\Sigma x^2 = 11,640.93413$

$\Sigma y^2 = 79,718.79$ $\Sigma xy = 29,804.4505$ $n = 9$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{29,804.4505 - \frac{(323.573)(824.9)}{9}}{11,640.93413 - \frac{(323.573)^2}{9}} =$$

b_1 = **19.2204**

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{824.9}{9} - (19.2204)\frac{323.573}{9} = \mathbf{-599.3674}$$

$\hat{y}$ = **-599.3674 + 19.2204 *x***

SSE = $\Sigma y^2 - b_0 \Sigma y - b_1 \Sigma xy$ =

79,718.79 – (-599.3674)(824.9) – 19.2204(29,804.4505) = 1283.13435

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{1283.13435}{7}} = \mathbf{13.539}$$

$$r^2 = 1 - \frac{SSE}{\sum y^2 - \frac{(\sum y)^2}{n}} = 1 - \frac{1283.13435}{79,718.79 - \frac{(824.9)^2}{9}} = \mathbf{.688}$$

12.39 $s_b = \dfrac{s_e}{\sqrt{\sum x^2 - \dfrac{(\sum x)^2}{n}}} = \dfrac{7.376}{\sqrt{58{,}293 - \dfrac{(571)^2}{7}}} = .068145$

$b_1 = -0.898$

H$_o$: $\beta = 0$ $\alpha = .01$
H$_a$: $\beta \neq 0$

Two-tail test, $\alpha/2 = .005$ df = n - 2 = 7 - 2 = 5

$t_{.005,5} = \pm 4.032$

$t = \dfrac{b_1 - \beta_1}{s_b} = \dfrac{-0.898 - 0}{.068145} = \mathbf{-13.18}$

Since the observed t = -13.18 < $t_{.005,5}$ = -4.032, the decision is to **reject the null hypothesis**.

12.41 $s_b = \dfrac{s_e}{\sqrt{\sum x^2 - \dfrac{(\sum x)^2}{n}}} = \dfrac{2.575}{\sqrt{421 - \dfrac{(41)^2}{5}}} = .27963$

$b_1 = -0.715$

H$_o$: $\beta = 0$ $\alpha = .05$
H$_a$: $\beta \neq 0$

For a two-tail test, $\alpha/2 = .025$ df = n - 2 = 5 - 2 = 3

$t_{.025,3} = \pm 3.182$

$t = \dfrac{b_1 - \beta_1}{s_b} = \dfrac{-0.715 - 0}{.27963} = \mathbf{-2.56}$

Since the observed t = -2.56 > $t_{.025,3}$ = -3.182, the decision is to **fail to reject the null hypothesis**.

12.43 F = 8.26 with a p-value of .021. The overall model is significant at α = .05 but not at α = .01. For simple regression,

$t = \sqrt{F} = \mathbf{2.874}$

$t_{.05,8}$ = 1.86 but $t_{.01,8}$ = 2.896. The slope is significant at α = .05 but not at α = .01.

12.45 x_0 = 100 For 90% confidence, $\alpha/2$ = .05
df = n - 2 = 7 - 2 = 5 $t_{.05,5} = \pm 2.015$

$$\bar{x} = \frac{\sum x}{n} = \frac{571}{7} = 81.57143$$

Σx= 571 Σx^2 = 58,293 s_e = 7.377

$\hat{y}$ = 144.414 - .0898(100) = 54.614

$$\hat{y} \pm t_{\alpha/2,n\text{-}2}\, s_e \sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}} =$$

$$54.614 \pm 2.015(7.377)\sqrt{1+\frac{1}{7}+\frac{(100-81.57143)^2}{58{,}293-\frac{(571)^2}{7}}} =$$

54.614 ± 2.015(7.377)(1.08252) = 54.614 ± 16.091

38.523 ≤ y ≤ 70.705

For x_0 = 130, $\hat{y}$ = 144.414 - 0.898(130) = 27.674

$$y \pm t_{/2,\text{n-2}}\, s_e \sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}} =$$

$$27.674 \pm 2.015(7.377)\sqrt{1+\frac{1}{7}+\frac{(130-81.57143)^2}{58{,}293-\frac{(571)^2}{7}}} =$$

$27.674 \pm 2.015(7.377)(1.1589) = 27.674 \pm 17.227$

$10.447 \leq y \leq 44.901$

The width of this confidence interval of y for $x_0 = 130$ is <u>wider</u> that the confidence interval of y for $x_0 = 100$ because $x_0 = 100$ is nearer to the value of $x = 81.57$ than is $x_0 = 130$.

12.47 $x_0 = 10$ For 99% confidence $\alpha/2 = .005$
df = $n - 2 = 5 - 2 = 3$ $t_{.005,3} = 5.841$

$$\bar{x} = \frac{\sum x}{n} = \frac{41}{5} = 8.20$$

$\Sigma x = 41$ $\Sigma x^2 = 421$ $s_e = 2.575$

$\hat{y} = 15.46 - 0.715(10) = 8.31$

$$\hat{y} \pm t_{\alpha/2,n\text{-}2}\, s_e \sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum x^2-\frac{(\sum x)^2}{n}}}$$

$$8.31 \pm 5.841(2.575)\sqrt{\frac{1}{5}+\frac{(10-8.2)^2}{421-\frac{(41)^2}{5}}} =$$

$8.31 \pm 5.841(2.575)(.488065) = 8.31 \pm 7.34$

$0.97 \leq E(y_{10}) \leq 15.65$

If the prime interest rate is 10%, we are 99% confident that the average bond rate is between 0.97% and 15.65%.

12.49

Year	Fertilizer
2004	5860
2005	6632
2006	7125
2007	6000
2008	4380
2009	3326
2010	2642

$\Sigma x = 14{,}049$ $\qquad \Sigma y = 35{,}965$ $\qquad \Sigma xy = 72{,}162{,}744$

$\Sigma x^2 = 28{,}196{,}371$ $\qquad \Sigma y^2 = 202{,}315{,}489$ $\qquad n = 7$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} =$$

$$\frac{72{,}162{,}744 - \frac{(14{,}049)(35{,}965)}{7}}{28{,}196{,}371 - \frac{(14{,}049)^2}{7}} = -678.9643$$

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{35{,}965}{7} - (-678.9642857)\frac{14{,}049}{7} = 1{,}367{,}819.18$$

$\hat{y}$ = **1,367,819.18 - 678.9643 x**

$\hat{y}$ (2014) = 1,367,819.18 - 678.9643(2014) = **385.08**

12.51 $\Sigma x = 36$ $\Sigma x^2 = 256$

$\Sigma y = 44$ $\Sigma y^2 = 300$

$\Sigma xy = 188$ $n = 7$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} = \frac{188 - \frac{(36)(44)}{7}}{\sqrt{\left[256 - \frac{(36)^2}{7}\right]\left[300 - \frac{(44)}{7}\right]}}$$

$$r = \frac{-38.2857}{\sqrt{(70.85714)(23.42857)}} = \frac{-38.2857}{40.7441} = \mathbf{-.940}$$

12.53

x	y
53	5
47	5
41	7
50	4
58	10
62	12
45	3
60	11

$\Sigma x = 416$ $\Sigma x^2 = 22,032$

$\Sigma y = 57$ $\Sigma y^2 = 489$ $b_1 = 0.355$

$\Sigma xy = 3,106$ $n = 8$ $b_0 = -11.335$

a) $\hat{y}$ **= -11.335 + 0.355** $\boldsymbol{x}$

b)

$\hat{y}$ (Predicted Values)	(y- $\hat{y}$) residuals
7.48	**-2.48**
5.35	**-0.35**
3.22	**3.78**
6.415	**-2.415**
9.255	**0.745**
10.675	**1.325**
4.64	**-1.64**
9.965	**1.035**

c) $(y-\hat{y})^2$

6.1504
0.1225
14.2884
5.8322
0.5550
1.7556
2.6896
1.0712

SSE = **32.4649**

d) $s_e = \sqrt{\dfrac{SSE}{n-2}} = \sqrt{\dfrac{32.4649}{6}}$ = **2.3261**

e) $r^2 = 1 - \dfrac{SSE}{\sum y^2 - \dfrac{(\sum y)^2}{n}} = 1 - \dfrac{32.4649}{489 - \dfrac{(57)^2}{8}}$ = **.608**

f) H_o: $\beta = 0$ $\alpha = .05$
H_a: $\beta \neq 0$

Two-tailed test, $\alpha/2 = .025$ df = n - 2 = 8 - 2 = 6

$t_{.025,6} = \pm 2.447$

$$s_b = \frac{s_e}{\sqrt{\sum x^2 - \dfrac{(\sum x)^2}{n}}} = \frac{2.3261}{\sqrt{22{,}032 - \dfrac{(416)^2}{8}}} = 0.116305$$

$$t = \frac{b_1 - \beta_1}{s_b} = \frac{0.3555 - 0}{.116305} = \mathbf{3.05}$$

Since the observed $t = 3.05 > t_{.025,6} = 2.447$, the decision is to **reject the null hypothesis**.

The population slope is different from zero.

g) This model produces only a modest $r^2 = .608$. Almost 40% of the variance of y is unaccounted for by x. The range of y values is 12 - 3 = 9 and the standard error of the estimate is 2.33. Given this small range, the s_e is not small.

12.55 a) $x_0 = 60$

$\Sigma x = 524$ $\Sigma x^2 = 36,224$

$\Sigma y = 215$ $\Sigma y^2 = 6,411$ $b_1 = .5481$

$\Sigma xy = 15,125$ $n = 8$ $b_0 = -9.026$

$s_e = 3.201$ 95% Confidence Interval $\alpha/2 = .025$

df = n - 2 = 8 - 2 = 6 $t_{.025,6} = \pm 2.447$

$\hat{y} = -9.026 + 0.5481(60) = 23.86$

$$\bar{x} = \frac{\sum x}{n} = \frac{524}{8} = 65.5$$

$$\hat{y} \pm t_{\alpha/2,n-2}\, s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

$$23.86 \pm 2.447(3.201) \sqrt{\frac{1}{8} + \frac{(60 - 65.5)^2}{36,224 - \frac{(524)^2}{8}}}$$

$23.86 \pm 2.447(3.201)(.375372) = 23.86 \pm 2.94$

$\mathbf{20.92 \le E(y_{60}) \le 26.8}$

b) $x_0 = 70$

$\hat{y}_{70} = -9.026 + 0.5481(70) = 29.341$

$$\hat{y} \pm t_{\alpha/2,n-2}\, s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

$$29.341 \pm 2.447(3.201) \sqrt{1 + \frac{1}{8} + \frac{(70 - 65.5)^2}{36,224 - \frac{(524)^2}{8}}}$$

$29.341 \pm 2.447(3.201)(1.06567) = 29.341 \pm 8.347$

$\mathbf{20.994 \le y \le 37.688}$

c) The confidence interval for (b) is much wider because part (b) is for a single value of y which produces a much greater possible variation. In actuality, $x_0 = 70$ in part (b) is slightly closer to the mean (x) than $x_0 = 60$. However, the width of the single interval is much greater than that of the average or expected y value in part (a).

12.57 $\Sigma y = 267$ $\Sigma y^2 = 15{,}971$

$\Sigma x = 21$ $\Sigma x^2 = 101$

$\Sigma xy = 1{,}256$ $n = 5$

$b_0 = 9.234375$ $b_1 = 10.515625$

SSE = $\Sigma y^2 - b_0\Sigma y - b_1\Sigma xy =$

15,971 - (9.234375)(267) - (10.515625)(1,256) = 297.7969

$$r^2 = 1 - \frac{SSE}{\sum y^2 - \frac{(\sum y)^2}{n}} = 1 - \frac{297.7969}{1{,}713.2} = \mathbf{.826}$$

If a regression model would have been developed to predict number of cars sold by the number of sales people, the model would have had an r^2 of 82.6%. The same would hold true for a model to predict number of sales people by the number of cars sold.

12.59

Sales(y)	Number of Units(x)
34.2	14.1
11.4	24.7
8.5	5.9
8.4	7.2
6.8	5.7
5.4	7.6
4.5	4.8
4.1	1.6
3.7	3.5
3.4	4.9
3.3	1.5
3.0	3.5
3.0	2.2

$\Sigma y = 99.7$ $\Sigma y^2 = 1609.01$ $\Sigma x = 87.2$

$\Sigma x^2 = 1067$ $\Sigma xy = 1034.05$ $n = 13$

$b_1 = 0.75773$ $b_0 = 2.58662$

$\hat{y}$ **= 2.58662 + 0.75773 *x***

SSE = $\Sigma y^2 - b_0\Sigma y - b_1\Sigma xy =$

$1609.01 - (2.58662)(99.7) - (0.75773)(1034.05) = 567.59328$

$$r^2 = 1 - \frac{SSE}{\sum y^2 - \frac{(\sum y)^2}{n}} = 1 - \frac{567.59328}{1609.01 - \frac{(99.7)^2}{13}} = \mathbf{.328}$$

12.61 $\Sigma x = 36.62$ $\Sigma x^2 = 217.137$
$\Sigma y = 57.23$ $\Sigma y^2 = 479.3231$
$\Sigma xy = 314.9091$ $n = 8$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} =$$

$$r = \frac{314.9091 - \frac{(36.62)(57.23)}{8}}{\sqrt{\left[217.137 - \frac{(36.62)^2}{8}\right]\left[479.3231 - \frac{(57.23)^2}{8}\right]}} =$$

$$r = \frac{52.938775}{\sqrt{(49.50895)(69.91399)}} = \mathbf{.8998}$$

There is a strong positive relationship between the inflation rate and the thirty-year treasury yield.

12.63 $\Sigma x = 11.902$ $\Sigma x^2 = 25.1215$
$\Sigma y = 516.8$ $\Sigma y^2 = 61{,}899.06$ $b_1 = 66.36277$
$\Sigma xy = 1{,}202.867$ $n = 7$ $b_0 = -39.0071$

$$\hat{y} = \mathbf{-39.0071 + 66.36277}\ \boldsymbol{x}$$

SSE = $\Sigma y^2 - b_0 \Sigma y - b_1 \Sigma xy$

SSE = 61,899.06 - (-39.0071)(516.8) - (66.36277)(1,202.867) = 2,232.343

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{2{,}232.343}{5}} = \mathbf{21.13}$$

$$r^2 = 1 - \frac{SSE}{\sum y^2 - \frac{(\sum y)^2}{n}} = 1 - \frac{2{,}232.343}{61{,}899.06 - \frac{(516.8)^2}{7}} = 1 - .094 = \mathbf{.906}$$

12.65

Advertising	Revenues
5.0	82.6
3.1	148.9
2.5	110.9
2.5	55.8
2.4	126.7
2.4	97.2
2.1	136.3
2.1	30.0
2.1	28.3

$\Sigma x = 816.7$ $\Sigma y = 24.2$
$\Sigma x^2 = 90{,}185.73$ $\Sigma y^2 = 71.86$
$\Sigma xy = 2237.36$ $n = 9$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} =$$

$$\frac{2237.36 - \frac{(816.7)(24.2)}{9}}{90{,}185.73 - \frac{(816.7)^2}{9}} = 0.00257$$

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{24.2}{9} - (0.00257)\frac{816.7}{9} = 2.4557$$

$\hat{y}$ = **2.4557 + 0.00257 *x***

SSE = $\Sigma y^2 - b_0 \Sigma y - b_1 \Sigma xy$

= 71.86 – (2.4557)(24.2) – (0.00257)(2237.36) = 6.682

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{6.682}{7}} = \mathbf{0.977}$$

$$r^2 = 1 - \frac{SSE}{\sum y^2 - \frac{\left(\sum y\right)^2}{n}} = 1 - \frac{6.682}{71.86 - \frac{(24.2)^2}{9}} = \mathbf{.0157}$$

H_0: $\beta = 0$
H_a: $\beta \neq 0$ $\alpha = .05$ $t_{.025,7} = \pm 2.365$

$$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 90{,}185.73 - \frac{(816.7)^2}{9} = 16{,}074.74$$

$$t = \frac{b_1 - 0}{\frac{s_e}{\sqrt{SS_{xx}}}} = \frac{0.00257}{\frac{0.977}{\sqrt{16{,}074.74}}} = 0.33$$

Since the observed $t = 0.33 < t_{.025,7} = 2.365$, the decision is to **fail to reject the null hypothesis**.

12.67 a) The regression equation is: $\hat{y}$ = **67.2 – 0.0565 *x***

b) For every unit of increase in the value of *x*, the predicted value of *y* will decrease by .0565.

c) The *t* ratio for the slope is –5.50 with an associated *p*-value of .000. This is significant at $\alpha = .10$. The *t* ratio is negative because the slope is negative and the numerator of the *t* ratio formula equals the slope minus zero.

d) r^2 is .627 or 62.7% of the variability of *y* is accounted for by *x*. This is only a modest proportion of predictability. The standard error of the estimate is 10.32. This is best interpreted in light of the data and the magnitude of the data.

e) The *F* value which tests the overall predictability of the model is 30.25. For simple regression analysis, this equals the value of t^2 which is $(-5.50)^2$.

f) The negative is not a surprise because the slope of the regression line is also negative indicating an inverse relationship between *x* and *y*. In addition, taking the square root of r^2 which is .627 yields .7906 which is the magnitude of the value of *r* considering rounding error.

12.69 The Residual Model Diagnostics from Minitab indicate a relatively healthy set of residuals. The Histogram indicates that the error terms are generally normally distributed. This is somewhat confirmed by the semi straight line Normal Plot of Residuals. However, the Residuals vs. Fits graph indicates that there may be some heteroscedasticity with greater error variance for small *x* values.