Step 6: State the conclusion in terms of the original problem.

There is overwhelming evidence at the 0.05 level that $H_a : \beta_1 \neq 0$. This implies that it is reasonable to believe (at the 0.05 level) that there is a linear relationship between the age and the price of a Jeep Cherokee. In fact, there appears to be a negative linear relationship between the age and the price of a Jeep Cherokee. However, our hypothesis test did not address the issue of a *negative* relationship, so we cannot make this conclusion.

If a data analyst feels that the assumptions of the simple linear model have been met and decides to make an inference about the model, the *P*-value of b_1 will be one of the first pieces of computer output that will be examined. The analyst will also look at the value of R^2 (R Square in the Excel output) to see what proportion of the variation in the data is explained by the regression model (see Section 5.3).

Thus far, the focus in this chapter has been inference on β_1. What about β_0? Since β_0 is merely a constant term, in most problems its value is not of great concern. However, if a confidence interval or test of hypothesis is needed, the methods used would be virtually identical to those presented for analyzing β_1.

13.2 Exercises

Basic Concepts

1. Identify two purposes that a confidence interval for β_1 serves.
2. What is the formula for the $100(1-\alpha)\%$ confidence interval for β_1?
3. What are the three pieces of information needed to calculate a confidence interval for β_1?
4. A 99% confidence interval for β_1 is found to be $(5.6,\ 10.2)$. Give two interpretations of this interval.
5. For the confidence interval given in the previous question, what is b_1, the sample estimate for β_1?
6. If there is no linear relationship between two variables, what is the value of β_1? Explain.
7. What is the test statistic for testing the hypothesis that $\beta_1 \neq 0$? Describe how this test statistic is similar to other test statistics used in hypothesis testing.
8. What are the degrees of freedom for the test statistic in the previous question?
9. Can we make inferences about β_0? Why are we more interested in inferences about β_1?
10. Explain why the *P*-value corresponding to b_1 is one of the first values examined by data analysts.

©HAWKES LEARNING

Exercises

11. Consider the random sample of data in the following table regarding the age of a particular model of car and the asking price for that car.

Car Data			
Age (Years)	**Asking Price ($)**	**Age (Years)**	**Asking Price ($)**
1	27,288	4	18,998
1	25,984	5	18,800
2	24,858	5	18,500
2	25,551	6	16,897
3	20,199	6	17,997

a. Draw a scatterplot of the data. Describe the relationship you observe in the scatterplot.

b. Using statistical software, estimate the simple linear model relating age to asking price.

c. What is the standard error of b_1?

d. Find a 99% confidence interval for β_1.

e. Interpret the confidence interval found in part **d.**

12. An economist is studying the relationship between income and IRA contributions. He has randomly selected eight subjects and obtained annual income and IRA contribution data from them. He wishes to predict the amount of money contributed to an IRA based on annual income.

Income and IRA Contributions							
Annual Income (Thousands of Dollars)	56	50	66	75	84	70	92
IRA Contribution (Thousands of Dollars)	0.3	0	1.2	1.8	3.3	2.2	5.2

a. Draw a scatterplot of the data. Describe the relationship that you observe between income and IRA contribution.

b. Estimate the parameters of the following model using statistical software.

$$IRA\ Contribution = \beta_0 + \beta_1 Income + \varepsilon_i$$

c. Calculate and interpret a 95% confidence interval for β_1.

d. What assumptions are being made about the error term in the construction of the confidence interval for β_1?

13. Consider the following summary output, which was generated from a random sample of 8 employees relating age to annual salary.

©HAWKES LEARNING

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.732431223
R Square	0.536455496
Adjusted R Square	0.459198079
Standard Error	15.60374155
Observations	8

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	1690.639497	1690.639	6.943741
Residual	6	1460.860503	243.4768	
Total	7	3151.5		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-2.132440745	20.99597109	-0.10156	0.922412
Age	1.564320608	0.593648001	2.635098	0.038794

a. What is the estimated regression equation?

b. Is there evidence of a linear relationship between age and salary at the 0.05 level?

c. Does the decision in part **b.** change at the 0.01 level? Explain.

d. What proportion of the variation in annual salary is explained by the model? (See Section 5.3.)

14. The college placement office is developing a model to relate grade point average (GPA) to starting salary for liberal arts majors. Twenty recent graduates have been randomly selected, and their graduating GPAs and starting salaries were recorded.

GPA and Starting Salary										
GPA	2.2	3.5	2.1	2.8	1.9	3.2	2.5	2.4	2.9	3.1
Starting Salary (Thousands of Dollars)	55.1	65.2	56.3	59.3	54.3	61.4	57.6	54.8	45.7	63.2
GPA	3.7	2.0	3.3	2.7	3.5	2.6	3.4	3.9	3.0	2.8
Starting Salary (Thousands of Dollars)	59.5	47.8	62.5	55.9	72.4	57.6	62.3	72.5	60.3	58.0

a. Draw a scatterplot of the data. Describe the relationship you observe between GPA and starting salary.

b. Using statistical software, estimate the parameters of the model

$$\textit{Starting Salary} = \beta_0 + \beta_1 \textit{GPA} + \varepsilon_i.$$

c. Is there evidence of a linear relationship between GPA and starting salary? Test at the 0.05 level.

d. Predict the starting salary for a student with a GPA of 2.5.

e. Interpret the coefficient of GPA in the model.

f. What proportion of the variation in starting salaries is explained by GPA? (See Section 5.3.)

g. To perform statistical inference on the model, what assumptions are being made?

©HAWKES LEARNING

15. A statistics professor would like to build a model relating student scores on the first test to the scores on the second test. The test scores from a random sample of 21 students who have previously taken the course are given in the table.

Test Scores					
Student	First Test Grade	Second Test Grade	Student	First Test Grade	Second Test Grade
1	69	73	12	54	67
2	66	56	13	57	65
3	69	65	14	85	67
4	75	51	15	75	67
5	57	59	16	79	77
6	75	76	17	44	51
7	75	76	18	82	84
8	82	76	19	57	81
9	91	82	20	75	90
10	66	73	21	69	73
11	88	67			

a. Draw a scatterplot of the two test grades and describe the relationship you observe.

b. Using statistical software, estimate the parameters of the model

$$\textit{Second Test Grade} = \beta_0 + \beta_1 \textit{First Test Grade} + \varepsilon_i.$$

c. What proportion of the variation in the grades on the second test is explained by the grades on the first test?

d. Is there a linear relationship between the first test grades and the second test grades? Test at the 0.05 level.

e. Suppose you're enrolled in the professor's course this semester. If you scored a 75 on the first test, use the model to predict your second test score. Round your answer to the nearest whole number.

Data

stat.hawkeslearning.com
Discovering Statistics and Data, Fourth Edition > Data Sets > Mount Pleasant Real Estate Data

16. Using the Mount Pleasant Real Estate data set, answer the following questions.

a. Using statistical software, estimate the simple linear regression model relating *List Price* (dependent variable) to *Square Footage* (independent variable).

b. Interpret the slope coefficient of the model.

c. Calculate and interpret a 95% confidence interval for β_1.

d. Is there evidence of a linear relationship between *List Price* and *Square Footage* at the 0.05 level?

e. What proportion of the variation in *List Price* is explained by *Square Footage*? (See Section 5.3.)

f. Using the estimated linear regression model in part **a.**, predict the price of a home in Mount Pleasant that has 3000 square feet.

Data

stat.hawkeslearning.com
Discovering Statistics and Data, Fourth Edition > Data Sets > US County Data

17. Using the US County Data data set, answer the following questions.

a. Using statistical software, estimate the simple linear regression model relating *Diabetes.percent* (dependent variable) to *Adult.obesity.percent* (independent variable).

©HAWKES LEARNING

b. Interpret the slope coefficient of the model.

c. Calculate and interpret a 95% confidence interval for β_1.

d. Is there evidence of a linear relationship between *Diabetes.percent* and *Adult.obesity.percent* at the 0.05 level?

e. What proportion of the variation in *Diabetes.percent* is explained by *Adult.obesity.percent*? (See Section 5.3.)

18. Using the Global Statistics by Country data set, answer the following questions regarding birth rate and female literacy rate.

a. Draw a scatterplot of the data. Describe the relationship you observe in the scatterplot.

b. Using statistical software, estimate the simple linear model relating female literacy rate to birth rate.

c. Find a 99% confidence interval for the slope.

d. Interpret the confidence interval found in part **c.** in terms of the problem.

Data

The full data set is available on stat.hawkeslearning.com under **Data Sets > Global Statistics by Country**.

13.3 Inference Concerning the Model's Prediction

The vast majority of regression models are developed for predictive purposes. For example, if you built the model relating the price of a Jeep Cherokee Limited to its age, it was probably because you want to use it to predict prices. While it is important to evaluate b_1, the estimate of the slope, the real concern of the model builder is the accuracy of a model's predictions. In the case of the Jeep Cherokee Limited model, how accurate are the prices that the model predicts? If the assumptions of the linear model (detailed in Section 13.1) have been met, then it is possible to make inferences as to the quality of a model's predictions.

The Regression Line as the Mean Value of *y* Given *x*

Examining the Jeep Cherokee data in Example 13.2.1 reveals two cars that are one-year old. For a given value of age (say one year) the prices of the one-year-old cars were \$44,998 and \$42,768. For anyone who has ever observed the car market, price variation is not unexpected. If you use the model,

$$\textit{Estimated Asking Price of Jeep Cherokee} = \$47{,}030.83 - \$3846.09\,\textit{Age}$$

for predictive purposes, then the predicted value of a one-year-old Jeep Cherokee will be

$$\textit{Asking Price} = \$47{,}030.83 - \$3846.09(1) = \$43{,}184.74.$$

Using this model, all one-year-old Jeep Cherokees will have a predicted value of \$43,184.74. Since the prices of one-year-old Jeep Cherokees vary, how do you interpret the predicted price of \$43,184.74? The model's predicted value when *Age* is set to one is considered to be the average price of a one-year-old Jeep Cherokee. In other words, it is the mean value of *y* (price) when *x* (age) equals one. But wait a minute! The prices