SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.866441198
R Square	0.75072035
Adjusted R Square	0.719560393
Standard Error	223.0681781
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1198827.203	1198827	24.09247121	0.00118116
Residual	8	398075.2967	49759.41		
Total	9	1596902.5			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	2227.955922	370.1487713	6.019082	0.000316596	1374.391324	3081.520519
Items Produced	53.8835069	10.97779658	4.908408	0.00118116	28.5686626	79.1983512

Figure 13.5.4

The *P*-value measures the probability that the test statistic is as large as it is (in magnitude) under the assumption that the null hypothesis is true. Specifically, a *P*-value is the probability of observing a test statistic as large or larger (in absolute value) than what has been observed, given that the null hypothesis is true. In Example 13.5.4, the value of the test statistic was 4.908. The probability of observing a test statistic this large (in absolute value) or larger, given that the true value of the slope is zero, is very small. Fortunately, virtually all statistical analysis programs that perform regression analysis calculate *P*-values for the two-tailed test of hypothesis

$$H_0\colon \beta_1 = 0$$

$$H_a\colon \beta_1 \neq 0.$$

Figure 13.5.4 shows the *P*-value of b_1 to be approximately 0.0012. A *P*-value of 0.0012 is persuasive evidence that $\beta_1 \neq 0$. The null hypothesis is rejected if the *P*-value is less than or equal to α. In Example 13.5.4, the significance level of the test was set at $\alpha = 0.05$. Since the *P*-value $=$ $0.0012 \leq 0.05$, the null hypothesis is rejected in favor of the alternative hypothesis ($H_a\colon \beta_1 \neq 0$).

If the *P*-value is used, how does the procedure for testing a hypothesis change? In **Step 4** all that is necessary is to specify α. It serves as a critical value. In **Step 6**, the *P*-value is compared to α. Everything else remains the same, provided the *P*-value has been calculated for you.

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 the computer output that will be examined.

13.5 Exercises

Basic Concepts

1. Give an example of a practical application of the confidence interval for β_1.
2. Identify two purposes that confidence intervals for the estimated regression coefficients serve.
3. What is the sampling distribution for b_1? Give the mean and standard deviation.
4. What is the expression for determining the $100(1-\alpha)\%$ confidence interval for β_1?

5. Suppose a 95% confidence interval for β_1 is found to be $(15.11,\ 20.11)$. Give two interpretations of this interval.

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 associated with the simple linear regression model?

9. Can we make inferences about β_0? Explain why we are more interested in inferences about β_1.

10. Describe why the P-value corresponding to b_1, which is displayed by many regression summary outputs, is one of the first values examined by data analysts.

Exercises

11. Consider the summary output for the monthly sales data given in Exercise 7 in Section 13.4.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.949341195
R Square	0.901248704
Adjusted R Square	0.891373575
Standard Error	51.20789475
Observations	12

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	239318.1818	239318.1818	91.26449427
Residual	10	26222.48485	2622.248485	
Total	11	265540.6667		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	403.7575758	31.51628057	12.81107949	1.57569E-07
Age	40.9090901	4.282219283	9.55324522	2.41268E-06

a. Compute a 90% confidence interval for β_1.

b. Interpret this interval.

12. Consider the 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 ($)
1	11,875
1	10,995
2	9995
2	8500
3	8995
4	6995
5	4450
5	5500
6	4400
6	4800

a. Using statistical software, determine the sample estimate of β_1.

b. What is the standard error of b_1?

c. Find a 99% confidence interval for β_1.

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

13. 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)	IRA Contribution (Thousands of Dollars)
28	0.3
25	0
34	1.0
43	1.3
48	3.3
39	2.2
74	8.5

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.

$$\text{IRA Contribution} = \beta_0 + \beta_1(\text{Income}) + \varepsilon_i$$

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

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

e. Use the confidence interval you obtained to test the hypothesis that the IRA contribution increases with the increase in income of the subject.

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

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 significance level?

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

d. What percentage of the variation in annual salary is explained by the model?

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

GPA and Starting Salary

GPA	Starting Salary (Thousands of Dollars)
2.2	35.1
3.5	45.2
2.1	36.3
2.8	39.3
3.2	41.4
2.5	37.6
2.4	34.8
2.9	25.7
3.1	40.1
3.7	39.5

a. Plot the data. Describe the relationship you observe between GPA and starting salary.

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

$$\text{Starting Salary} = \beta_0 + \beta_1(\text{GPA}) + \varepsilon_i.$$

c. Is there evidence of a linear relationship between GPA and starting salary? Test at the 0.05 significance 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 fraction of the variation in starting salaries is explained by GPA?

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

16. An experienced census official feels that she can accurately estimate the number of inhabitants of a city block by simply noting the size of the block and the types of buildings (single family homes, apartments, businesses, etc.) that are found on the block. This procedure, if accurate, would be much quicker and cheaper than visiting each residence and taking a survey of its inhabitants. The data below are estimates of block populations provided by the official for 10 blocks in a large city. Also given are the actual numbers of inhabitants for the same 10 blocks. These were found at a later point in time by conventional methods.

Inhabitants of City Blocks

Estimate	115	234	215	97	78	134	78	129	170	67
Actual	100	225	190	99	92	125	75	130	155	82

a. Draw a scatterplot of points of the actual number against the estimated number of inhabitants. Does the relationship appear to be linear?

b. Estimate the slope and intercept of the following regression equation using statistical software.

$$\text{Actual Inhabitants} = \beta_0 + \beta_1(\text{Estimated Inhabitants}) + \varepsilon_i$$

c. Is there evidence of a linear relationship between the actual number and the estimated number? Test at the $\alpha = 0.01$ significance level.

d. Interpret the regression coefficient for the estimated number of inhabitants.

e. Construct a 95% confidence interval for the slope of the regression equation. Interpret the interval.

f. Compute and interpret the R^2 value.

g. Predict the actual number of inhabitants on a block when the estimated number is 150. Round your answer to the nearest whole number.

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

$$\text{Second Test Grade} = \beta_0 + \beta_1(\text{First Test Grade}) + \varepsilon_i.$$

b. What fraction of the variation in the grades on the second test is explained by the grades on the first test?

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

d. 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.

13.6 Inference Concerning the Model's Prediction

Many regression models are developed for predictive purposes. For example, if you built the model relating the number of items produced to total cost, it was probably because you want to use it to predict total cost. While it is important to evaluate b_1, the estimate of the slope, the real concern of the model builder is the accuracy of the model's predictions. In the case of the production model, how accurate are the costs 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 production data in Table 13.5.1 reveals two weeks in which 30 items were produced. For a given value of items produced (say 30) the costs of producing 30 items were \$3800 and \$3600. For anyone who has observed a production process, price variation is not unexpected. If you use the model