# Chapter 14 – Inference on the Least-Squares Regression Model

## OUTLINE

* 1. Testing the Significance of the Least-Squares Regression Model

1. Using Randomization Techniques on Slope of the Least-Squares Regression Line
   1. Confidence and Prediction Intervals

## Putting It Together

In Chapter 4, we learned methods for describing the relation between bivariate quantitative data. We also learned to perform diagnostic tests, such as determining whether a linear model is appropriate, identifying outliers, and identifying influential observations.

In this chapter, we begin by extending hypothesis testing and confidence intervals to least-squares regression models. In Section 14.1, we test whether a linear relation exists between two quantitative variables using methods based on those presented in Chapter 10. In Section 14.2, we construct confidence intervals about the predicted value of the response variable.

## Section 14.1 Testing the Significance of the Least-Squares Regression Model

### Objectives

1. State the Requirements of the Least-Squares Regression Model
2. Compute the Standard Error of the Estimate
3. Verify That the Residuals Are Normally Distributed
4. Conduct Inference on the Slope of the Least-Squares Regression Model
5. Construct a Confidence Interval about the Slope of the Least-Squares Regression Model

Introduction, Page 2

**Example 1 *A Review of Least-Squares Regression***

A family doctor is interested in examining the relationship between a patient's age and total cholesterol (in milligrams per deciliter). He randomly selects 14 of his female patients and obtains the data presented in Table 1. The data are based on results obtained from the National Center for Health Statistics. Draw a scatter diagram, compute the correlation coefficient, find the least-squares regression equation, and determine the coefficient of determination.

**Table 1**

| Age, *x* | Total Cholesterol, *y* |
| --- | --- |
| 25 | 180 |
| 25 | 195 |
| 28 | 186 |
| 32 | 180 |
| 32 | 210 |
| 32 | 197 |
| 38 | 239 |
| 42 | 183 |
| 48 | 204 |
| 51 | 221 |
| 51 | 243 |
| 58 | 208 |
| 62 | 228 |
| 65 | 269 |

#### Objective 1: State the Requirements of the Least-Squares Regression Model

Objective 1, Page 1

Because  and  are statistics, their values vary from sample to sample; so a sampling distribution is associated with each. We use this sampling distribution to perform inference on  and . For example, we might want to test whether  is different from 0. If we have sufficient evidence to this effect, we conclude that there is a linear relation between the explanatory variable, *x*, and response variable, *y*.

1. What does the notation  represent?

Objective 1, Page 2

 *Answer the following after watching the video.*

1. State the two requirements of the least-squares regression model.
2. What does the first requirement mean?

OBJECTIVE 1, PAGE 2 (CONTINUED)

1. How can you check if the first requirement is satisfied?

Objective 1, Page 3

1. State the equation that the least-squares regression model is given by.

#### Objective 2: Compute the Standard Error of the Estimate

Objective 2, Page 1

1. State the formula for the standard error of the estimate, .

OBJECTIVE 2, PAGE 2

**Example 2 *Determining the Standard Error***

Compute the standard error of the estimate for the age and total cholesterol data in Table 1.

**Table 1**

| Age, *x* | Total Cholesterol, *y* |
| --- | --- |
| 25 | 180 |
| 25 | 195 |
| 28 | 186 |
| 32 | 180 |
| 32 | 210 |
| 32 | 197 |
| 38 | 239 |
| 42 | 183 |
| 48 | 204 |
| 51 | 221 |
| 51 | 243 |
| 58 | 208 |
| 62 | 228 |
| 65 | 269 |

#### Objective 3: Verify That the Residuals Are Normally Distributed

Objective 3, Page 1

1. How do we verify that the residuals are normally distributed?

Objective 3, Page 2

**Example 3 *Verify that Residuals Are Normally Distributed***

Verify that the residuals obtained in Example 1 are normally distributed.

| *x* | *y* | Residual |
| --- | --- | --- |
| 25 | 180 |  |
| 25 | 195 | 8.67 |
| 28 | 186 |  |
| 32 | 180 |  |
| 32 | 210 | 13.88 |
| 32 | 197 | 0.88 |
| 38 | 239 | 34.48 |
| 42 | 183 |  |
| 48 | 204 |  |
| 51 | 221 |  |
| 51 | 243 | 20.29 |
| 58 | 208 |  |
| 62 | 228 |  |
| 65 | 269 | 26.70 |

#### Objective 4: Conduct Inference on the Slope of the Least-Squares Regression Model

Objective 4, Page 1

1. If there is no linear relation between the response and explanatory variables, what will the slope of the true regression line be? Why?

OBJECTIVE 4, PAGE 1 (CONTINUED)

1. To conclude that a linear relation exists between two variables, what null and alternative hypotheses can be used?

Objective 4, Page 2

1. State the test statistic for the slope in a least-squares regression model. What distribution does it follow?

Objective 4, Page 3

1. What are the two conditions that must be satisfied before testing a hypothesis regarding the slope coefficient, ?

Objective 4, Page 3 (continued)

1. State the five steps for testing a hypothesis regarding the slope coefficient, .

Step 1

Step 2

Step 3 (By Hand)

Step 3 (Using Technology)

Step 4

Step 5

Objective 4, Page 4

**Note: Handling Departures from Normality**

Because these procedures are robust, minor departures from normality will not adversely affect the results of the test.

In fact, for large samples , inferential procedures regarding can be used even with significant departures from normality.

Objective 4, Page 5

**Example 4 *Testing for a Linear Relation***

Test whether a linear relation exists between age and total cholesterol at the  level of significance using the data in Table 1 from Example 1.

**Table 1**

| **Age, x** | **Total Cholesterol, y** |
| --- | --- |
| 25 | 180 |
| 25 | 195 |
| 28 | 186 |
| 32 | 180 |
| 32 | 210 |
| 32 | 197 |
| 38 | 239 |
| 42 | 183 |
| 48 | 204 |
| 51 | 221 |
| 51 | 243 |
| 58 | 208 |
| 62 | 228 |
| 65 | 269 |

#### Objective 5: Construct a Confidence Interval about the Slope of the Least-Squares Regression Model

Objective 5, Page 1

1. State the formulas for the lower bound and upper bound of a  confidence interval for the slope of the regression line, .

Objective 5, Page 2

**Example 5 *Constructing a Confidence Interval for the Slope of the Least-Squares Regression Line***

Determine a 95% confidence interval for the slope of the least-squares regression line for the data in Table 1 from Example 1.

**Table 1**

| **Age, *x*** | **Total Cholesterol, *y*** |
| --- | --- |
| 25 | 180 |
| 25 | 195 |
| 28 | 186 |
| 32 | 180 |
| 32 | 210 |
| 32 | 197 |
| 38 | 239 |
| 42 | 183 |
| 48 | 204 |
| 51 | 221 |
| 51 | 243 |
| 58 | 208 |
| 62 | 228 |
| 65 | 269 |

Objective 5, Page 4

1. As the value of  gets larger, what is the impact on the value of ?
2. What does this result imply?

Objective 5, Page 5

1. State the two basic reasons we are intentionally avoiding a discussion of inference on the correlation coefficient.

## Section 14.1A Using Randomization Techniques on the Slope of the Least-Squares Regression Line

### Objectives

1. Use Randomization to Test the Significance of the Slope of the Least-Squares Regression Model

#### Objective 1: Use Randomization to Test the Significance of the Slope of the Least-Squares Regression Model

Objective 1, Page 1

1. State the correlation and the least-squares regression equation, where *x* is the Zestimate and *y* is the selling price.

Objective 1, Page 2

We can use the randomization techniques of Chapter 11 to judge whether two quantitative variables are significantly associated.

However, rather than using the correlation coefficient to judge whether a linear relation exists between two quantitative variables, we are going to use the slope of the least-squares regression equation.

Objective 1, Page 3

1. We would like to know if the slope of 1.3059 suggests that higher Zestimates correspond with a higher selling price, or is it possible the two variables are not positively associated and the slope is 0? State the two possibilities in formulating the judgement.
2. State the null and alternative hypotheses associated with these two possibilities.

Objective 1, Page 3 (continued)

1. State the slope and correlation coefficient for the randomly assigned data from StatCrunch.
2. What do the slope and correlation coefficient suggest about the association between the two variables?

Objective 1, Page 4

We will use the Randomization Test for Slope applet in StatCrunch in order to determine how likely it is to observe a sample slope as extreme or more extreme than the one actually observed.

Objective 1, Page 5

1. How many of the 5000 samples resulted in a slope of 1.3059 or higher? What is the *P*-value?
2. What conclusion can be drawn about the association between the two variables?

Objective 1, Page 6

1. What is the shape of the distribution in the null model? Where is it centered?

Objective 1, Page 7

1. State the five steps for testing hypotheses regarding the slope of the Least-Squares Regression Using Randomization.

Step 1

Step 2

Step 3

Step 4

Step 5

Objective 1, Page 8

**Example 1 *Testing Hypotheses Regarding the Slope of the Least-Squares Regression Line Using Randomization***

A family doctor is interested in examining the relationship between a patient's age and total cholesterol (in milligrams per deciliter). He randomly selects 14 of his female patients and obtains the data presented in Table 2. The data are based on results obtained from the National Center for Health Statistics. Does the sample evidence suggest a linear relation exists between age and total cholesterol?

**Table 2**

| **Age, *x*** | **Total Cholesterol, *y*** |
| --- | --- |
| 25 | 180 |
| 25 | 195 |
| 28 | 186 |
| 32 | 180 |
| 32 | 210 |
| 32 | 197 |
| 38 | 239 |
| 42 | 183 |
| 48 | 204 |
| 51 | 221 |
| 51 | 243 |
| 58 | 208 |
| 62 | 228 |
| 65 | 269 |

Step 1

Step 2

Step 3

Step 4

Step 5

## Section 14.2 Confidence and Prediction Intervals

### Objectives

1. Construct Confidence Intervals for a Mean Response
2. Construct Prediction Intervals for an Individual Response

Introduction, Page 1

1. State the two interpretations for the predicted value of total cholesterol  for a given age *x*.

Introduction, Page 2

1. What is a confidence interval for a mean response?
2. What is a prediction interval for an individual response?

**Note:** Confidence intervals are used for a mean response and prediction intervals are used for an individual result.

#### Objective 1: Construct Confidence Intervals for a Mean Response

Objective 1, Page 1

1. State the formulas for the lower bound and upper bound for a  confidence interval for , the mean response of *y* for a specified value of *x*.
2. What are the required conditions for constructing a  confidence interval for , the mean response of *y* for a specified value of *x*?

Objective 1, Page 2

**Example 1 *Constructing a Confidence Interval for a Mean Response***

Construct a 95% confidence interval for the predicted mean total cholesterol of all 42-year-old females, using the data in Table 1.

**Table 1**

| **Age, *x*** | **Total Cholesterol, *y*** |
| --- | --- |
| 25 | 180 |
| 25 | 195 |
| 28 | 186 |
| 32 | 180 |
| 32 | 210 |
| 32 | 197 |
| 38 | 239 |
| 42 | 183 |
| 48 | 204 |
| 51 | 221 |
| 51 | 243 |
| 58 | 208 |
| 62 | 228 |
| 65 | 269 |

#### Objective 2: Construct Prediction Intervals for an Individual Response

Objective 2, Page 1

1. State the formulas for the lower bound and upper bound for a  prediction interval for , the individual response of *y* for a specified value of *x*.
2. What are the required conditions for constructing a  confidence interval for , the individual response of *y* for a specified value of *x*?

Objective 2, Page 2

**Example 2 *Constructing a Prediction Interval for an Individual Response***

Construct a 95% prediction interval for the predicted total cholesterol for a 42-year-old females, using the data in Table 1.

**Table 1**

| **Age, *x*** | **Total Cholesterol, *y*** |
| --- | --- |
| 25 | 180 |
| 25 | 195 |
| 28 | 186 |
| 32 | 180 |
| 32 | 210 |
| 32 | 197 |
| 38 | 239 |
| 42 | 183 |
| 48 | 204 |
| 51 | 221 |
| 51 | 243 |
| 58 | 208 |
| 62 | 228 |
| 65 | 269 |

1. Explain why the interval about the individual (prediction interval for an individual response) is wider than the interval about the mean (confidence interval for a mean response).