The Significance of the Model

Lecture 47
Section 13.10

Robb T. Koether

Hampden-Sydney College

Wed, Apr 18, 2012

Outline

Outline

Introduction

- We finally address the question of whether the relationship between x and y is significant.

Introduction

- We finally address the question of whether the relationship between x and y is significant.
- If it is not significant, then any correlation we observe is purely accidental.

Introduction

- We finally address the question of whether the relationship between x and y is significant.
- If it is not significant, then any correlation we observe is purely accidental.
- This is relatively likely in small samples, where randomness dominates.

Introduction

- We finally address the question of whether the relationship between x and y is significant.
- If it is not significant, then any correlation we observe is purely accidental.
- This is relatively likely in small samples, where randomness dominates.
- In larger samples, it is not likely, but weak accidental correlations are still possible.

Introduction

- We finally address the question of whether the relationship between x and y is significant.
- If it is not significant, then any correlation we observe is purely accidental.
- This is relatively likely in small samples, where randomness dominates.
- In larger samples, it is not likely, but weak accidental correlations are still possible.
- We will perform a test that will settle the question, up to a level of certainty (α).

Outline

1. Introduction
2. The True Model
3. Testing the Significance of the Relationship
4. Testing the Significance on the TI-83
5. Significance vs. Strength of the Relationship
6. Assignment

The True Regression Line

- The estimated regression line

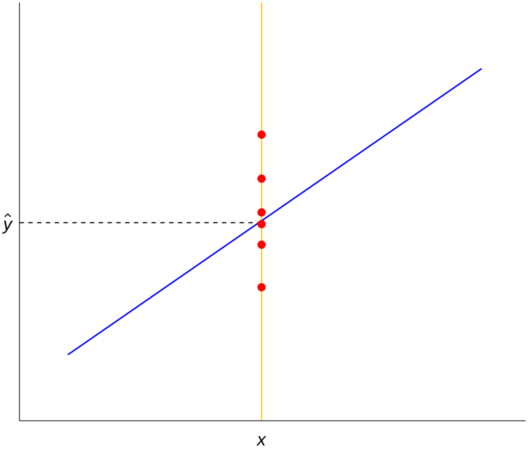
$$\hat{y} = a + bx$$

  is based on the data.
- Therefore, it is an approximation to the true regression line

$$y = \alpha + \beta x,$$

  had we surveyed the entire population.
- The coefficients a and b are estimators of the true coefficients α and β.
- In symbols, $\hat{\alpha} = a$ and $\hat{\beta} = b$.

The True Mean Value of y



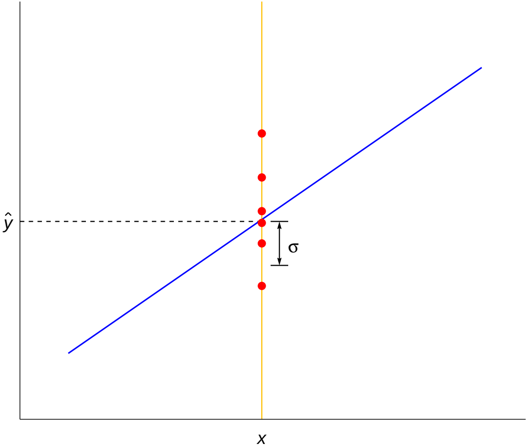
For each value of x, there is an entire distributions of values of y, aligned vertically in the scatterplot.

The True Mean Value of y

- Our estimate of the true mean of this distribution is $\hat{y}$.
- That is,

$$\hat{y} = \hat{\alpha} + \hat{\beta}x.$$

The True Standard Deviation



The true standard deviation σ measures how much the data points deviate from the mean $\hat{y}$.

The True Standard Deviation

- We assume that σ is the same for all values of x.
- The estimated standard deviation s is an estimate of the true standard deviation σ of the model.
- That is, $\hat{\sigma} = s$.
- The estimated standard deviation s is given by

$$s = \sqrt{\frac{\text{SSE}}{n-2}}.$$

The True Correlation Coefficient

- The true correlation coefficient is ρ.
- It is estimated by r.
- That is, $\hat{\rho} = r$.

- Is there really a relationship between height and weight?

Height (x)	Weight (y)
70	185
65	140
71	180
76	220
68	150
67	170
68	185
72	200
74	210
69	160

Linear Regression Analysis

- We find that
  - The regression line $\hat{y} = -310 + 7x$,
  - The correlation coefficient $r = 0.9075$,
  - $\text{SSE} = 1050$.

Linear Regression Analysis

- We find that
  - The regression line $\hat{y} = -310 + 7x$,
  - The correlation coefficient $r = 0.9075$,
  - $\text{SSE} = 1050$.
- Therefore,
  - $a = \hat{\alpha} = -310$.
  - $b = \hat{\beta} = 7$.
  - $r = \hat{\rho} = 0.9075$.
  - $s = \hat{\sigma} = \sqrt{\frac{\text{SSE}}{n-2}} = \sqrt{\frac{1050}{8}} = 11.456$.

Outline

Testing the Significance of the Model

- Now we will test the significance of the model.
- This is a measure of how *sure* we are that there really is a relationship between x and y.
- It is not a measure of how weak or strong the relationship is.
- If there is no relationship whatsoever between x and y, then $r = 0$ and $b = 0$.

Testing the Significance of the Model

- Now we will test the significance of the model.
- This is a measure of how *sure* we are that there really is a relationship between x and y.
- It is not a measure of how weak or strong the relationship is.
- If there is no relationship whatsoever between x and y, then $r = 0$ and $b = 0$.
- Actually, we should say $\rho = 0$ and $\beta = 0$.

Testing the Significance of the Model

- We will describe a test that will decide whether $\rho = 0$ and $\beta = 0$.
- We will use the data of the previous example to demonstrate the procedure.

Testing the Significance of the Model

Example (Test of Significance of the Model)

(1) The hypotheses:

$$H_0 : \beta = 0 \text{ and } \rho = 0$$

$$H_1 : \beta \neq 0 \text{ and } \rho \neq 0$$

(2) $\alpha = 0.05$

Example (Test of Significance of the Model)

(3) The test statistic uses the quantity $\text{SE}(b)$ which turns out to be

$$\text{SE}(b) = \frac{s}{\sqrt{\text{SSX}}}.$$

Testing the Significance of the Model

Example (Test of Significance of the Model)

(3) The test statistic uses the quantity SE(b) which turns out to be

$$\mathrm{SE}(b) = \frac{s}{\sqrt{\mathrm{SSX}}}.$$

Recall that

$$s = \sqrt{\frac{\mathrm{SSE}}{n-2}}.$$

Testing the Significance of the Model

Example (Test of Significance of the Model)

(3) The test statistic uses the quantity SE(b) which turns out to be

$$\mathrm{SE}(b) = \frac{s}{\sqrt{\mathrm{SSX}}}.$$

Recall that

$$s = \sqrt{\frac{\mathrm{SSE}}{n-2}}.$$

The test statistic is

$$t = \frac{b-0}{\mathrm{SE}(b)}.$$

Testing the Significance of the Model

Example (Test of Significance of the Model)

(4) Calculate the value of the test statistic.

$$s = \sqrt{\frac{\text{SSE}}{n-2}} = \sqrt{\frac{1050}{8}} = 11.456.$$

Testing the Significance of the Model

Example (Test of Significance of the Model)

(4) Calculate the value of the test statistic.

$$s = \sqrt{\frac{\text{SSE}}{n-2}} = \sqrt{\frac{1050}{8}} = 11.456.$$

$$\text{SSX} = \sum (x - \overline{x})^2 = 100.$$

Example (Test of Significance of the Model)

(4) Calculate the value of the test statistic.

$$s = \sqrt{\frac{\text{SSE}}{n-2}} = \sqrt{\frac{1050}{8}} = 11.456.$$

$$\text{SSX} = \sum (x - \overline{x})^2 = 100.$$

$$\text{SE}(b) = \frac{s}{\sqrt{\text{SSX}}} = \frac{11.456}{\sqrt{100}} = 1.1456.$$

Testing the Significance of the Model

Example (Test of Significance of the Model)

(4) Calculate the value of the test statistic.

$$s = \sqrt{\frac{\text{SSE}}{n-2}} = \sqrt{\frac{1050}{8}} = 11.456.$$

$$\text{SSX} = \sum (x - \overline{x})^2 = 100.$$

$$\text{SE}(b) = \frac{s}{\sqrt{\text{SSX}}} = \frac{11.456}{\sqrt{100}} = 1.1456.$$

Testing the Significance of the Model

Example (Test of Significance of the Model)

(5) The p-value is

$$2 \times \texttt{tcdf(6.110,E99,8)} = 2.863 \times 10^{-4}.$$

(6) Reject H_0.

(7) The model is significant.

Outline

Testing the Significance on the TI-83

TI-83 Testing the Significance of the Model

- Enter the x values into list L_1.
- Enter the y values into list L_2.
- Compute the regression equation and store it in Y_1.
- Select `STAT > TESTS > LinRegTTest`.
- Enter the `XList`.
- Enter the `YList`.
- Choose the appropriate alternative hypothesis.
- Enter the regression function Y_1.
- Select `Calculate`.

Testing the Significance on the TI-83

TI-83 Testing the Significance of the Model

- The results appear in the window.
  - The 1st title `y=a+bx`.
  - The 2nd title $\beta \neq 0$ and $\rho \neq 0$.
  - The value of t.
  - The p-value.
  - The degrees of freedom.
  - The value of a.
  - The value of b.
  - The value of s.
  - The value of r^2.
  - The value of r.

Testing the Significance on the TI-83

TI-83 Testing the Significance of the Model

- Perform the test of the previous example on the TI-83.

Outline

Significance vs. Strength

- The strength of the correlation and the significance are different concepts.
- We could have a very strong, but insignificant relation.
- We could have a very weak, but highly significant relation.

Significance vs. Strength

- r^2 measures the *strength* of the relationship.

  If r^2 is close to 1, the relationship is strong.

- The p-value of the test measures the *significance* of the relationship.

  If the p-value is close to 0, the relationship is significant.

Outline

Homework

- Read Sections 13.10.1 and 13.10.2, pages 870 - 880, skipping any discussion of confidence intervals. In particular, skip pages 876 and 877.
- Let's Do It! 13.17.
- Exercises 35(abde), 36(abcdf), page 888.
- Review exercises 39, 40, 41, 42, 43, 46, 47(omit f), 52, 54, 57(omit df).