The Significance of the Model

Lecture 50
Section 13.10

Robb T. Koether

Hampden-Sydney College

Wed, Apr 23, 2008

Outline

Introduction

- To wrap up Chapter 13, we will 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.

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

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

- 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

- A standard assumption is that for any given value of x, the values of y are normally distributed about true mean value $\alpha + \beta x$ of y, given by the true regression line.
- Our estimate of the true mean 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 $\hat{y}$.
- We assume that σ is independent 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 True Standard Deviation

- The estimated standard deviation s is given by

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

- This is similar to the formula

$$s = \sqrt{\frac{\text{SSX}}{n-1}}$$

  that we had in Chapter 5, except that now we have $n-2$ degrees of freedom instead of $n-1$ degrees of freedom.
- The difference is the 1 degree of freedom that is assigned to the regression line.

The True Correlation Coefficient

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

Linear Regression Analysis

- The components of linear regression analysis are
  - Estimate the unknown parameters α, β, and σ using sample data.
  - Use the estimated regression equation to make predictions.
  - Use the correlation coefficient to measure the strength and direction of the relationship.
  - Determine whether the linear model is statistically significant.
  - Calculate interval estimates for the predicted values of y.

Linear Regression Analysis

- The components of linear regression analysis are
  - Estimate the unknown parameters α, β, and σ using sample data.
  - Use the estimated regression equation to make predictions.
  - Use the correlation coefficient to measure the strength and direction of the relationship.
  - Determine whether the linear model is statistically significant.
  - Calculate interval estimates for the predicted values of y.

Linear Regression Analysis

- The components of linear regression analysis are
  - Estimate the unknown parameters α, β, and σ using sample data.
  - Use the estimated regression equation to make predictions.
  - Use the correlation coefficient to measure the strength and direction of the relationship.
  - Determine whether the linear model is statistically significant.
  - Calculate interval estimates for the predicted values of y.

Linear Regression Analysis

- The components of linear regression analysis are
  - Estimate the unknown parameters α, β, and σ using sample data.
  - Use the estimated regression equation to make predictions.
  - Use the correlation coefficient to measure the strength and direction of the relationship.
  - Determine whether the linear model is statistically significant.
  - Calculate interval estimates for the predicted values of y.

Linear Regression Analysis

- The components of linear regression analysis are
  - Estimate the unknown parameters α, β, and σ using sample data.
  - Use the estimated regression equation to make predictions.
  - Use the correlation coefficient to measure the strength and direction of the relationship.
  - Determine whether the linear model is statistically significant.
  - Calculate interval estimates for the predicted values of y.

Linear Regression Analysis

- The components of linear regression analysis are
  - Estimate the unknown parameters α, β, and σ using sample data.
  - Use the estimated regression equation to make predictions.
  - Use the correlation coefficient to measure the strength and direction of the relationship.
  - Determine whether the linear model is statistically significant.
  - Calculate interval estimates for the predicted values of y.

Linear Regression Analysis

- The components of linear regression analysis are
  - Estimate the unknown parameters α, β, and σ using sample data. Pretty much done.
  - Use the estimated regression equation to make predictions.
  - Use the correlation coefficient to measure the strength and direction of the relationship.
  - Determine whether the linear model is statistically significant.
  - Calculate interval estimates for the predicted values of y.

Linear Regression Analysis

- The components of linear regression analysis are
  - Estimate the unknown parameters α, β, and σ using sample data. Pretty much done.
  - Use the estimated regression equation to make predictions. Done.
  - Use the correlation coefficient to measure the strength and direction of the relationship.
  - Determine whether the linear model is statistically significant.
  - Calculate interval estimates for the predicted values of y.

Linear Regression Analysis

- The components of linear regression analysis are
  - Estimate the unknown parameters α, β, and σ using sample data. Pretty much done.
  - Use the estimated regression equation to make predictions. Done.
  - Use the correlation coefficient to measure the strength and direction of the relationship. Done.
  - Determine whether the linear model is statistically significant.
  - Calculate interval estimates for the predicted values of y.

Linear Regression Analysis

- The components of linear regression analysis are
  - Estimate the unknown parameters α, β, and σ using sample data. Pretty much done.
  - Use the estimated regression equation to make predictions. Done.
  - Use the correlation coefficient to measure the strength and direction of the relationship. Done.
  - Determine whether the linear model is statistically significant. Will do shortly.
  - Calculate interval estimates for the predicted values of y.

Linear Regression Analysis

- The components of linear regression analysis are
  - Estimate the unknown parameters α, β, and σ using sample data. Pretty much done.
  - Use the estimated regression equation to make predictions. Done.
  - Use the correlation coefficient to measure the strength and direction of the relationship. Done.
  - Determine whether the linear model is statistically significant. Will do shortly.
  - Calculate interval estimates for the predicted values of y. Will not do.

Linear Regression Analysis

- Consider again our sample data.

x	y
1	8
3	12
4	9
5	14
8	16
9	20
11	17
15	24

Linear Regression Analysis

- We found
  - The regression line $\hat{y} = 7.3 + 1.1x$,
  - The correlation coefficient $r = 0.9387$,
  - $\text{SSE} = 24.5$.
- Therefore,
  - $\hat{\alpha} = 7.3$.
  - $\hat{\beta} = 1.1$.
  - $\hat{\rho} = 0.9387$.
  - $\hat{\sigma} = \sqrt{\dfrac{\text{SSE}}{n-2}} = \sqrt{\dfrac{24.5}{6}} = 2.021$.

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, however weak or strong it may be.
- If there is no relationship whatsoever between x and y, then $r = 0$.

Testing the Significance of the Model

- Recall that
$$r = \frac{\sum (x - \overline{x})(y - \overline{y})}{\sqrt{\mathrm{SSX}}\sqrt{\mathrm{SSY}}}.$$
- So if $r = 0$, then
$$\sum (x - \overline{x})(y - \overline{y}) = 0.$$
- Also recall that
$$b = \frac{\sum (x - \overline{x})(y - \overline{y})}{\mathrm{SSX}}.$$
- So, if $r = 0$, then $b = 0$, too.

Testing the Significance of the Model

- Also recall that *r* and *b* are estimators of ρ and β.
- So the real question is whether $\rho = 0$ and $\beta = 0$.
- We will describe a test that will determine 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

- The hypotheses:

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

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

- $\alpha = 0.05$

Testing the Significance of the Model

- The test statistic is

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

where

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

and

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

and the number of degrees of freedom is $n - 2$.

Testing the Significance of the Model

- Calculate the value of the test statistic.

$$s = \sqrt{\frac{\text{SSE}}{n-2}} = 2.021.$$

$$\text{SE}(b) = \frac{s}{\sqrt{\text{SSX}}} = \frac{2.021}{\sqrt{150}} = 0.1650.$$

$$t = \frac{b-0}{\text{SE}(b)} = \frac{1.1}{0.1650} = 6.667.$$

Testing the Significance of the Model

- The p-value is

$$\texttt{tcdf(6.667,E99,6)} = 5.510 \times 10^{-4}.$$

- Reject H_0.
- The model is significant.

Testing the Significance on the TI-83

- The TI-83 will perform this test.
  - Enter the x values into list $\mathtt{L}_1$.
  - Enter the y values into list $\mathtt{L}_2$.
  - Compute the regression equation and store it in $\mathtt{Y}_1$.
  - Select `STAT > TESTS > LinRegTTest`.
  - Enter the `XList`.
  - Enter the `YList`.
  - Choose the alternative hypothesis $\beta \neq 0$.
  - Enter the regression function $\mathtt{Y}_1$.
  - Select `Calculate`.

Testing the Significance on the TI-83

- 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.

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.

- We will use the Excel files Correlation1.xls and Correlation2.xls to experiment.

Summary

The End