1342.03 Statistics
Singleton
Notes by Caxton Petri
page 1

2/28/2020

Main Ideas

- Define the coefficient of determination, R^2
- Explain the meaning of R^2.

2/28/2020

Section 5.3

Ok. So, we now know how to look at bivariate data with a scatterplot. Using Excel, we can then determine a "best fit" line to model the data. As discussed, the model will not be exact. That is, there will be some error. In fact, just because we can find a model does not necessarily mean we want to use that model.

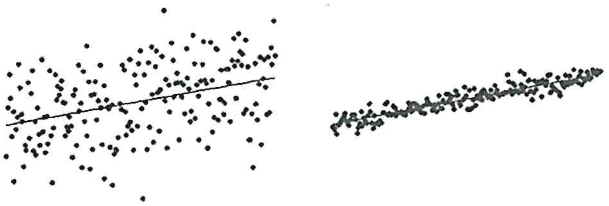
Thus, we come to the question: Is this model a good fit? How well does this model accurately predict the dependent variable?

The **coefficient of determination**, denoted R^2, summarizes the degree of fit of the linear model. That is, how well can the x-variables be used to predict the y-variables.

In other words, R^2 represents the percent of data that is closest to the line of best fit. For example, if $R^2 = 0.850$, this means that 85% of the total variation in y can be explained by the linear relationship between x and y. The other 15% of the total variation remains unexplained.

Meaning of R^2:

- $0 \leq R^2 \leq 1$ aka $0\% \leq R^2 \leq 100\%$
- The higher the R^2, the better the model fits the data
- The lower the R^2, the more variation
- $R^2 = 1$ The linear model represents all the data
- $R^2 = 0$ The linear model represents none of the data



Difference between r and R^2:

- r measures the strength of the relationship between two variables. Is there a correlation between the variables? Is it positive? Negative?

- R^2 measures the strength of the linear equation used to predict the dependent variable. Once we generate the model, does the model do a good job of predicting future values?

In Excel: You can find the R^2 value by creating a scatterplot, adding a trendline, and formatting the line to display the R^2 value on the chart.

Correlation coefficient, r

- measure of how strong the linear relationship is

$$-1 \underset{0}{\longleftrightarrow} 1$$

Coefficient of determination, R^2

- tells us how well our linear equation fits the data.
- goodness of fit
- $0 \longleftrightarrow 1$
- ex. $R^2 = 0.80$

  this means that 80% of the variation in the y-variable can be attributed to variation in the x-variable.
- if $R^2 = 1$ then it is called a perfect fit
- if $R^2 = 0$ then it has no fit.