Unit6

humberto

10/21/2014

Unit 5

Introduction

Relationship between two numerical variables.

- Correlation.
- Modeling.
- Model Diagnostics.
- Inference.

Part 1: Correlation

Describes the strength of the *linear association* between two variables.

Denoted as R

Properties

The magnitude (absolute value) of the correlation coefficient measures the strength of the linear association between two numerical variables.

The sign of the correlation coefficient indicates the direction of association.

The correlation coefficient is always between -1 (perfect negative linear association) and 1 (perfect positive linear association).

R = 0 indicates no linear relationship.

The correlation coefficient is unitless, and is not affected by changes in the center or scale of either variable. (such as unit conversions)

The correlation of X with Y is the same as of Y with X.

The correlation coefficient is sensitive to outliers/

Part 2: Residuals

Diference between the observed and predicted y.

residuals:$e_i = y_i - \hat{y}$

Part 2: Least Squares Line

Why least squares ?

- Most commonly used.
- Easier to compute.
- In many applicaitons, a residual twice as large as another is more than twice as bad.

Least squares line $\hat{y} = \beta_0 + \beta_1 x$

x: explanatory.

β_0: intercept.

β_1: slope.

$\hat{y}$: predicted response.

Point Estimates

Slope: For each unit increase in x, y is expected to be higher/lower on average by "the slope".

$b_1 = \frac{S_y}{S_x} R$

Intercept: when x = 0, y is expected to equal "the intercept".

$b_0 = \bar{y} - b_1 \bar{x}$

Part 2: Prediction and Extrapolation

Prediction

- Using the linear model to predict the value of the response variable for a given value of the explanatory variable is called **prediction**