1

1.1 Goals

Goals of this section

- Learn in detail how to estimate the relationship between one or more variables.
- Learn how to decompose the variance into variability that is explained and unexplained.

2 Multiple Regression

Multiple Regression

- Multiple regression line (population):
  \[ y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots + \beta_k x_{k,i} + \epsilon_i \]
- Multiple regression line (sample):
  \[ y_i = b_0 + b_1 x_{1,i} + b_2 x_{2,i} + \ldots + b_k x_{k,i} + e_i \]
- \( k \): number of parameters you are estimating.
- \( \epsilon_i \): error term, since linear relationship between the \( x \) variables and \( y \) are not perfect.
- \( e_i \): residual = the difference between the predicted value \( \hat{y} \) and the actual value \( y_i \).

3 Sum of Squares

3.1 SSR and SSE

Sum of Squares Measures of Variation
3.2 SST
Sum of Squares Measures of Variation (continued)

- **Sum of Squares Total (SST)**: measure of the total variability in the dependent variable. Does the formula below look familiar?
  
  \[
  SST = \sum_{i=1}^{n} (y_i - \bar{y})^2
  \]

- SST = SSR + SSE.

3.3 Degrees of Freedom
Degrees of Freedom

- Each measure of variability has its own degrees of freedom.
- Degrees of freedom regression = \(df_R = k - 1\).
- Degrees of freedom error = \(df_E = n - k\).
- Degrees of freedom total = \(df_T = n - 1\) (Look familiar?).

4 Mean Squared Measures
Mean Squared Measures of Variation

- **Mean Squared Regression (MSR)**: Measure of the average amount unexplained variability in the dependent variable. of variability in the dependent (Y) variable that is explained by the independent variables (X’s).
  
  \[
  MSR = \frac{SSR}{df_R}
  \]
• **Mean Squared Error (MSE):** Measure of the *average* amount of unexplained variability in the dependent variable.

\[ MSE = \frac{SSE}{df_E} \]

• No textbook ever talks about a mean squared total, what do you think this would equal?

## 5 Coefficient of determination

### 5.1 $R^2$

**Coefficient of determination**

• The **coefficient of determination** is the percentage of variability in $y$ that is explained by $x$.

\[ R^2 = \frac{SSR}{SST} \]

• This *is not the same* as the correlation coefficient.

• In the case of single variable regression, actually equal to the square of the correlation coefficient.

• $R^2$ will always be between 0 and 1.

• The closer $R^2$ is to 1, the better $x$ is able to explain $y$.

### 5.2 Adjusted $R^2$

**Adjusted $R^2$**

• The more variables you add to the regression, the higher $R^2$ will be.

• Adding new variables is not necessarily good, when the new variables have nothing to do with the dependent variable.

• The Adjusted $R^2$ penalizes additional variables.

\[ R^2_{adj} = 1 - \frac{n - 1}{n - k - 1} (1 - R^2) \]

• When the adjusted $R^2$ increases when adding a variable, then the additional variable really did help explain the dependent variable.

• When the adjusted $R^2$ decreases when adding a variable, then the additional variable does not help explain the dependent variable.
Homework

- Problems computing regression:

- Computing Coefficient of determination.
  - Section 13.3, pages 528-529, problems 13.16 through 13.19.