4 Multiple Linear Regression Part 3

4.1 The Linear Regression Model 140
4.2 The Least Squares Estimates, and Point Estimation and Prediction 148
4.3 The Mean Square Error and the Standard Error 154
4.4 Model Utility: R^2, Adjusted R^2, and the Overall F-Test 155

4.5 Testing the Significance of an Independent Variable 160
4.6 Confidence and Prediction Intervals 163
4.7 The Quadratic Regression Model 167

4.8 Interaction 175

If the effect of one predictor variable depends on the value of another,

One way to approximate this is by including a third variable

which is the product of the other two multiplied by each other.

Examples: Alcohol & barbiturates, Poison & antidote, Soil & Fertilizer acidity

4.9 Using Dummy Variables to Model Qualitative Independent Variables 183

Turn one qualitative predictor with c categories into c-1 binary variables.

Choose one category is "all zeros"

the choice does not affect predictive power, but can affect clarity

4.10 The Partial F-Test: Testing the Significance of a Portion of a Regression Model 193

Suppose we have two competing models of the same dependent variable y:

The "complete model"

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_g x_g + \beta_{g+1} x_{g+1} + \ldots + \beta_k x_k + \varepsilon = \beta_0 + \sum_{j=1}^{k} \beta_j x_j$$

and the "reduced model"

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_g x_g + \varepsilon = \beta_0 + \sum_{j=1}^{g} \beta_j x_j$$

The complete model has all the predictors in the reduced model plus more besides

$H_o : \beta_{g+1} = \beta_{g+2} = \ldots = \beta_k = 0$ is the hypothesis that the complete model

does not add any real explanatory model over and above the reduced model.

We can reduce H_0 in favor of H_1: at least one of $\beta_{g+1}, \beta_{g+2}, \ldots, \beta_k \neq 0$

at significance level α if F = $\frac{(SSE_r - SSE_c)/(k-g)}{SSE_c/(n-k-1)} > F_\alpha^{k-g,n-k-1}$

p-Value for Partial F-test p = $\Pr\left(F^{g,n-k-1} \geq \frac{(SSE_r - SSE_c/(k-g)}{SSE_c/(n-k-1)}\right)$

Exercises 197