4 Multiple Linear Regression 1 139

4.1 The Linear Regression Model 140

$$y = \mu_{y|x_1.x_2,...,x_k} + \varepsilon = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \varepsilon = \beta_0 + \sum_{j=1}^{k} \beta_j x_j$$

4.2 The Least Squares Estimates, and Point Estimation and Prediction 148

$\hat{y} = b_0 + b_1x_1 + b_2x_2 + ... + b_kx_k$

We will rely on Excel (or SAS or SPSS) to calculate the b's

4.3 The Mean Square Error and the Standard Error 154

$$SSE = \sum_{i=1}^{n} (y - \hat{y})^2 = \sum_{i=1}^{n} \left(y - \left[\beta_0 + \sum_{j=1}^{k} \beta_j x_j \right] \right)^2$$

$$MSE = s^2 = \frac{SSE}{n-k-1} \quad s = \sqrt{\frac{SSE}{n-k-1}}$$

4.4 Model Utility: R^2, Adjusted R^2, and the Overall F-Test 155

$SST = SS_{yy}$ just as in simple linear regression

SSR = SST - SSE .

This is equal to $\sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2 = \sum_{i=1}^{n} \left(\left[\beta_0 + \sum_{j=1}^{k} \beta_j x_j \right] - \bar{y} \right)^2$ but SST-SSE is just as conceptually relevant and easier to compute.

$R^2 = \frac{SSR}{SST}$

R^2 Adjusted for degrees of freedom = $\bar{R}^2 = \left(R^2 - \frac{k}{n-1}\right)\left(\frac{n-1}{n-k-1}\right)$

Overall F-Test

We can reject $H_o : \beta_1 = \beta_2 = ... = \beta_k = 0$ in favor of H_1: at keast one $\beta_i \neq 0$ at significance level α if $F = \frac{SSR/k}{SSE/(n-k-1)} > F_\alpha^{k,n-k-1}$

p-Value for Overall F-test $p = \Pr\left(F^{k,n-k-1} \geq \frac{SSR/k}{SSE/(n-k-1)}\right)$

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

4.9 Using Dummy Variables to Model Qualitative Independent Variables 183

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

Exercises 197