Chapter 5 Model Building and Residual Analysis 221

5.1 Model Building and the Effects of Multicollinearity 222

Multicollinearity 222

Correlation Matrix

Variance Inflation Factor

Let R_j^2 = the coefficient of determination of the regression model with x_j as dependent variable and all the other x's as independent variables.

$VIF_j = \frac{1}{1-R_j^2}$

Multicollinearity "is considered a problem" if the largest VIF is greater than 10 or the average VIF is substantially greater than 1

Picket Fence Diagram.

Difficulty interpreting the t statistic;	
	common variance is allocated in subtle ways
	Value, significance, and even sign of a coefficient is unstable to small changes in the sample
	This is what the book is trying, obscurely, to say in the discussion (page 225) about the effect of VIF_j on "our ability to use the t statistics and related p values to assess the importance of the independent variables."

Comparing Regression Models 226

R^2	Complete model may be better than reduced model if $R_c^2 > R_r^2$
s and Adjusted R^2	Complete model may be better than reduced model if $s_c<s_r$ Note $s_c<s_r$ if and only if $\overline{R}_c^2 > \overline{R}_r^2$
Prediction Interval Length	Complete model may be better than reduced model if $t_{\frac{\alpha}{2}}^{n-(k+1)} s_c < t_{\frac{\alpha}{2}}^{n-k} s_r$ Note that $t_{\frac{\alpha}{2}}^{n-(k+1)} > t_{\frac{\alpha}{2}}^{n-k}$
The C Statistic	Skip (bottom of page 230, top of page 231)

Stepwise Regression: Skip (pages 231-234)

Backward Elimination (a.k.a. stepwise deletion)

Delete the "weakest link" variable as long as this reduces s