1 Multiple Linear Regression Review

The multiple linear regression model is of the form:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + \varepsilon_i,$$

where the errors ε_i are assumed to be independent and identically distributed (i.i.d.) with mean 0 and variance σ^2. This model is often called the **full model** or the **complete model** when compared to other models that set one or more of the β_i terms to 0. A general type of hypothesis that can be tested with this model is that

$H_0 : \beta_{g+1} = \beta_{g+2} = ... = \beta_k = 0$ versus H_a : not all $\beta_{g+1}, \beta_{g+2}, ..., \beta_k$ terms are 0

If H_0 above is true, then the model is

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_g X_{gi} + \varepsilon_i,$$

this is often called the **reduced model**. The null hypothesis H_0 is tested with an F test, sometimes called an MS_{drop} test:

$$F = \frac{[\text{SS(REG, complete)} - \text{SS(REG, reduced)}] / (k - g)}{\text{SS(RESID, complete)}/(n - (k + 1))}.$$

If the errors are normally distributed, then the F test above has an F distribution with $(k - g)$ and $(n - (k + 1))$ degrees of freedom. Special cases of the above test are when the null hypothesis is $H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$, called the **global model test** (denoted $H_{0(M)}$ in the text, with the F statistic denoted F_M) and when the null hypothesis is $H_0 : \beta_j = 0$, called an **individual variable test** (denoted $H_{0(j)}$ in the text, with the F statistic denoted F_j). Also note that the individual variable test F statistic (F_j) is the square of the usual t statistic that is shown on computer printouts for multiple regression analyses.

2 Bootstrap Multiple Regression tests

When the errors are not normally distributed, then the F statistics above do not have an F distribution. An alternative to the parametric F test is to obtain bootstrap P values for these tests by the following procedure:

1. Fit the regression model and obtain the residuals $e_i = Y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 X_{1i} + \widehat{\beta}_2 X_{2i} + ... + \widehat{\beta}_k X_{ki}) = Y_i - \widehat{Y}_{i.}$

2. For a single bootstrap iteration, obtain a bootstrap sample of the residuals e_i, and form a fixed-X bootstrap sample $(X_{1i}, X_{2i}, ..., X_{ki})$ and $e_{i,b}$, $i = 1, 2, ..., n$. Use the F statistic above to obtain a value F_e for any test of interest.

3. Repeat step 2 many times.

4. The bootstrap P value for the test of interest is the fraction of F_e values that are equal to or greater than the observed F value.

3 Bootstrap Confidence Intervals for an individual parameter β_j

To obtain a bootstrap confidence interval for β_j, follow the above procedure and obtain t_e statistics from each bootstrap sample, then use:

$$\widehat{\beta}_j - t_{je,1-\alpha/2}\widehat{SE}(\widehat{\beta}_j) < \beta_j < \widehat{\beta}_j - t_{je,\alpha/2}\widehat{SE}(\widehat{\beta}_j)$$

Also note that later in this section is given a general procedure for conducting a bootstrap test of the general linear hypothesis $H_0 : C\beta = 0$.