STA 4033
Mathematical Statistics
with Computer Applications
Lecture 18*

Brett Presnell
Department of Statistics
University of Florida

13 June 2001

*Copyright ©2001 by Brett Presnell
Example. The file flcrime.txt contains data on crime in Florida counties. The variables are

**Crime:** Crime rate, recorded as number of crimes per 1000 residents.

**Income:** Median income in thousands of dollars.

**HS:** Percentage of residents with at least a high school education (of those aged at least 25).

**Urban:** Percentage of residents living in an urban environment.

See R transcript.
7.6. Comparing Models

A *t* test for a single regression parameter $\beta_j$ can be viewed as a test of whether $x_j$ explains a significant proportion of the variation in $Y$ left unexplained by regression on the other predictors in the model.

A more general way to compare “nested” models is through analysis of variance $F$-tests.

Suppose as an example we have three predictors in the “full model”

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

and we wish to test

$H_0 : \beta_2 = \beta_3 = 0 \ vs \ H_a : \text{either } \beta_2 \neq 0 \ or \ \beta_3 \neq 0 \ or \ both$

The null hypothesis corresponds to the “reduced model”

$$Y = \beta_0 + \beta_1 x_1 + \epsilon$$

Note that these models are nested in the sense that the reduced model is a special case ($\beta_2 = \beta_3 = 0$) of the full model.
The general test statistic for comparing nested regression models is

\[ F = \frac{(\text{SSE}_{\text{Reduced}} - \text{SSE}_{\text{Full}})}{(\text{df}_{\text{Reduced}} - \text{df}_{\text{Full}})} \frac{\hat{\sigma}^2_{\text{Full}}}{\text{SSE}_{\text{Full}}/\text{df}_{\text{Full}}} \]

where \( \text{df}_{\text{Full}} \) and \( \text{df}_{\text{Reduced}} \) are the residual degrees of freedom for the full and reduced models, respectively. Note that \( \text{df}_{\text{Reduced}} - \text{df}_{\text{Full}} \) is equal to the difference in the number of parameters in the two models.

Recall that \( \hat{\sigma}^2_{\text{Full}} = \text{SSE}_{\text{Full}}/\text{df}_{\text{Full}} \), so an another way to write the \( F \) statistic is

\[ F = \frac{(\text{SSE}_{\text{Reduced}} - \text{SSE}_{\text{Full}})}{(\text{df}_{\text{Reduced}} - \text{df}_{\text{Full}})} \frac{\text{SSE}_{\text{Full}}/\text{df}_{\text{Full}}}{\text{SSE}_{\text{Full}}/\text{df}_{\text{Full}}} \]

In the specific example above, with three predictors in the full model and one predictor in the reduced model,

\[ \text{df}_{\text{Full}} = n - 4, \quad \text{df}_{\text{Reduced}} = n - 2, \quad \text{and} \quad \text{df}_{\text{Reduced}} - \text{df}_{\text{Full}} = 2. \]

Large values of the \( F \) statistic favor the alternative hypothesis over the null. To carry out the test this statistic is referred to an \( F \) distribution with

\( \text{df}_{\text{Reduced}} - \text{df}_{\text{Full}} \) numerator degrees of freedom, and

\( \text{df}_{\text{Full}} \) denominator degrees of freedom.
7.7. An Omnibus Test

Suppose we have \( p \) predictors, \( x_1, \ldots, x_p \), and consider the model

\[
Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon
\]

To establish that at least one of these explanatory variables (predictors) is useful in predicting \( Y \), we may do an omnibus test of the hypotheses

\[
H_0 : \beta_1 = \cdots = \beta_p = 0 \quad \text{vs} \quad H_a : \text{at least one } \beta_j \neq 0
\]

Thus we are comparing the full model above with the reduced model

\[
Y = \beta_0 + \varepsilon
\]

Note that for this reduced model, \( \hat{\beta}_0 = \bar{y} \).

This is just a special case of our prescription for comparing nested models. If we let a subscript of 0 denote the reduced model for this special case, then for this omnibus test,

\[
df_{\text{Full}} = n - (p + 1), \quad df_{\text{Reduced}} = df_0 = n - 1, \quad \text{and} \quad df_0 - df_{\text{Full}} = p.
\]

Also, in this particular case, the difference in residual sums of squares between the reduced and full models is called the regression sum of squares (or the sum of squares due to or explained by regression), and is given by

\[
SS_{\text{Reg}} = SSE_{\text{Reduced}} - SSE_{\text{Full}} = \sum_{i=1}^{n} (y_i - \bar{y})^2 - SSE_{\text{Full}} = S_{yy} - SSE_{\text{Full}}
\]

See R transcript.
7.8. Coefficient of Determination

The proportion of variation in \( Y \) explained by the regression is called the *coefficient of determination*, denoted \( R^2 \). Thus \( R^2 \) is given by

\[
R^2 = \frac{SS_{\text{Reg}}}{S_{yy}} = \frac{S_{yy} - SSE_{\text{Full}}}{S_{yy}} = 1 - \frac{SSE_{\text{Full}}}{S_{yy}}
\]

- Big \( R^2 \) (one is the maximum possible) is generally considered a good thing.
- Note that \( R^2 \) always increases as we add predictors to the model, so it is questionable as a device for comparing models.

*Adjusted \( R^2 \)*, sometimes denoted \( R^2_{\text{adj}} \), attempts to adjust for this by penalizing \( R^2 \) for the number of predictors in the model. \( R^2_{\text{adj}} \) is often used in model selection. Look it up if you are interested in more details.