F tests

Math 463, Spring 2009, University of Oregon

David A. Levin

University of Oregon

April 29, 2009

We have that

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

$$\mathrm{SS}_{\mathrm{Tot}} = \mathrm{SS}_{\mathrm{Res}} + \mathrm{SS}_{\mathrm{Reg}} \,.$$

We have that

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

$$\mathrm{SS}_{\mathrm{Tot}} = \mathrm{SS}_{\mathrm{Res}} + \mathrm{SS}_{\mathrm{Reg}} .$$

$$F = \frac{\mathrm{SS}_{\mathrm{Reg}}/p}{\mathrm{SS}_{\mathrm{Res}}/(n-(p+1))} = \frac{(\mathrm{SS}_{\mathrm{Tot}} - \mathrm{SS}_{\mathrm{Res}})/p}{\mathrm{SS}_{\mathrm{Res}}/(n-(p+1))} .$$

We have that

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

$$\mathrm{SS}_{\mathrm{Tot}} = \mathrm{SS}_{\mathrm{Res}} + \mathrm{SS}_{\mathrm{Reg}} \,.$$

$$F = \frac{\mathrm{SS}_{\mathrm{Reg}}/p}{\mathrm{SS}_{\mathrm{Res}}/(n-(p+1))} = \frac{(\mathrm{SS}_{\mathrm{Tot}} - \mathrm{SS}_{\mathrm{Res}})/p}{\mathrm{SS}_{\mathrm{Res}}/(n-(p+1))} \,.$$

Write

$$\mathrm{SS}_{\mathrm{Res}}(\boldsymbol{\beta}) = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

where $\hat{Y}$ is the vector of fitted values when parameters $\boldsymbol{\beta}$ are in model. The above statistic can be written as

$$F = \frac{(\mathrm{SS}_{\mathrm{Res}}(\beta_0) - \mathrm{SS}_{\mathrm{Res}}(\beta_0, \beta_1, \ldots, \beta_p))/p}{\mathrm{SS}_{\mathrm{Res}}/(n-(p+1))} \,.$$

Let $\boldsymbol{\beta}_0 = (\beta_0, \ldots, \beta_{p_1})$ and $\boldsymbol{\beta}_1 = (\beta_{p_1+1}, \ldots, \beta_p)$, etc. That is,

$$Y = X_0\boldsymbol{\beta}_0 + X_1\boldsymbol{\beta}_1 + \varepsilon\,.$$

Let

$$\mathrm{SS}_{\mathrm{Reg}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2,$$

where $\hat{Y}$ are the fitted values when the parameters in $\boldsymbol{\beta}$ are included in the model. We note that

$$\begin{aligned}\mathrm{SS}_{\mathrm{Reg}}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1) - \mathrm{SS}_{\mathrm{Reg}}(\boldsymbol{\beta}_0) &= \mathrm{SS}_{\mathrm{Tot}} - \mathrm{SS}_{\mathrm{Res}}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1) - [\mathrm{SS}_{\mathrm{Tot}} - \mathrm{SS}_{\mathrm{Res}}(\boldsymbol{\beta}_0)] \\ &= \mathrm{SS}_{\mathrm{Res}}(\boldsymbol{\beta}_0) - \mathrm{SS}_{\mathrm{Res}}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)\end{aligned}$$

is the gain in the regression sum-of-squares, when the parameters $\boldsymbol{\beta}_1$ is added to $\boldsymbol{\beta}_2$. If $\boldsymbol{\beta}_1$ has dimension p_1, then

$$F = \frac{(\mathrm{SS}_{\mathrm{Reg}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) - \mathrm{SS}_{\mathrm{Reg}}(\boldsymbol{\beta}_2))/p_1}{S^2}$$

has a F distribution with p_1 and $n - p$ degrees of freedom.