Dougherty

Introduction to Econometrics, 5th edition

Chapter 2: Properties of the Regression Coefficients and Hypothesis Testing

F TEST OF GOODNESS OF FIT

Model $$Y = \beta_1 + \beta_2 X + u$$

$$\sum \left(Y - \overline{Y}\right)^2 = \sum \left(\hat{Y} - \overline{Y}\right)^2 + \sum \hat{u}^2 \qquad TSS = ESS + RSS$$

In an earlier sequence it was demonstrated that the sum of the squares of the actual values of *Y* (*TSS*: total sum of squares) could be decomposed into the sum of the squares of the fitted values (*ESS*: explained sum of squares) and the sum of the squares of the residuals.

F TEST OF GOODNESS OF FIT

Model $$Y = \beta_1 + \beta_2 X + u$$

$$\sum \left(Y - \bar{Y}\right)^2 = \sum \left(\hat{Y} - \bar{Y}\right)^2 + \sum \hat{u}^2 \qquad TSS = ESS + RSS$$

$$R^2 = \frac{ESS}{TSS} = \frac{\sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \left(Y_i - \bar{Y}\right)^2}$$

R^2, the usual measure of goodness of fit, was then defined to be the ratio of the explained sum of squares to the total sum of squares.

F TEST OF GOODNESS OF FIT

Model $$Y = \beta_1 + \beta_2 X + u$$

$$\sum\left(Y - \bar{Y}\right)^2 = \sum\left(\hat{Y} - \bar{Y}\right)^2 + \sum \hat{u}^2 \qquad TSS = ESS + RSS$$

$$R^2 = \frac{ESS}{TSS} = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2}$$

The null hypothesis that we are going to test is that the model has no explanatory power.

F TEST OF GOODNESS OF FIT

Model $$Y = \beta_1 + \beta_2 X + u$$

Null hypothesis: $$H_0 \colon \beta_2 = 0$$

Alternative hypothesis: $$H_1 \colon \beta_2 \neq 0$$

$$\sum \left(Y - \bar{Y} \right)^2 = \sum \left(\hat{Y} - \bar{Y} \right)^2 + \sum \hat{u}^2 \qquad TSS = ESS + RSS$$

$$R^2 = \frac{ESS}{TSS} = \frac{\sum \left(\hat{Y}_i - \bar{Y} \right)^2}{\sum \left(Y_i - \bar{Y} \right)^2}$$

Since X is the only explanatory variable at the moment, the null hypothesis is that Y is not determined by X. Mathematically, we have H_0: $\beta_2 = 0$

F TEST OF GOODNESS OF FIT

Model $$Y = \beta_1 + \beta_2 X + u$$

Null hypothesis: $$H_0: \beta_2 = 0$$

Alternative hypothesis: $$H_1: \beta_2 \neq 0$$

$$\sum \left(Y - \bar{Y}\right)^2 = \sum \left(\hat{Y} - \bar{Y}\right)^2 + \sum \hat{u}^2 \qquad TSS = ESS + RSS$$

$$R^2 = \frac{ESS}{TSS} = \frac{\sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \left(Y_i - \bar{Y}\right)^2}$$

$$F(k-1, n-k) = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{\dfrac{ESS}{TSS}\Big/(k-1)}{\dfrac{RSS}{TSS}\Big/(n-k)} = \frac{R^2/(k-1)}{\left(1-R^2\right)/(n-k)}$$

Hypotheses concerning goodness of fit are tested via the *F* statistic, defined as shown. *k* is the number of parameters in the regression equation, which at present is just 2.

F TEST OF GOODNESS OF FIT

Model $Y = \beta_1 + \beta_2 X + u$

Null hypothesis: $H_0: \beta_2 = 0$

Alternative hypothesis: $H_1: \beta_2 \neq 0$

$$\sum \left(Y - \bar{Y}\right)^2 = \sum \left(\hat{Y} - \bar{Y}\right)^2 + \sum \hat{u}^2 \qquad TSS = ESS + RSS$$

$$R^2 = \frac{ESS}{TSS} = \frac{\sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \left(Y_i - \bar{Y}\right)^2}$$

$$F(k-1, n-k) = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{\frac{ESS}{TSS}\Big/(k-1)}{\frac{RSS}{TSS}\Big/(n-k)} = \frac{R^2/(k-1)}{\left(1-R^2\right)/(n-k)}$$

$n - k$ is, as with the t statistic, the number of degrees of freedom (number of observations less the number of parameters estimated). For simple regression analysis, it is $n - 2$.

F TEST OF GOODNESS OF FIT

Model $Y = \beta_1 + \beta_2 X + u$

Null hypothesis: $H_0 : \beta_2 = 0$

Alternative hypothesis: $H_1 : \beta_2 \neq 0$

$$\sum \left(Y - \bar{Y} \right)^2 = \sum \left(\hat{Y} - \bar{Y} \right)^2 + \sum \hat{u}^2 \qquad TSS = ESS + RSS$$

$$R^2 = \frac{ESS}{TSS} = \frac{\sum \left(\hat{Y}_i - \bar{Y} \right)^2}{\sum \left(Y_i - \bar{Y} \right)^2}$$

$$F(k-1, n-k) = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{\dfrac{ESS}{TSS}\Big/(k-1)}{\dfrac{RSS}{TSS}\Big/(n-k)} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

The *F* statistic may alternatively be written in terms of R^2. First divide the numerator and denominator by *TSS*.

F TEST OF GOODNESS OF FIT

Model $Y = \beta_1 + \beta_2 X + u$

Null hypothesis: $H_0: \beta_2 = 0$

Alternative hypothesis: $H_1: \beta_2 \neq 0$

$$\sum \left(Y - \bar{Y}\right)^2 = \sum \left(\hat{Y} - \bar{Y}\right)^2 + \sum \hat{u}^2 \qquad TSS = ESS + RSS$$

$$R^2 = \frac{ESS}{TSS} = \frac{\sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \left(Y_i - \bar{Y}\right)^2}$$

$$F(k-1, n-k) = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{\dfrac{ESS}{TSS}\Big/(k-1)}{\dfrac{RSS}{TSS}\Big/(n-k)} = \frac{R^2/(k-1)}{\left(1-R^2\right)/(n-k)}$$

We can now rewrite the F statistic as shown. The R^2 in the numerator comes straight from the definition of R^2.

$$\frac{RSS}{TSS} = \frac{TSS - ESS}{TSS} = 1 - \frac{ESS}{TSS} = 1 - R^2$$

$$\sum\left(Y - \bar{Y}\right)^2 = \sum\left(\hat{Y} - \bar{Y}\right)^2 + \sum \hat{u}^2 \qquad TSS = ESS + RSS$$

$$R^2 = \frac{ESS}{TSS} = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2}$$

$$F(k-1, n-k) = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{\frac{ESS}{TSS}\Big/(k-1)}{\frac{RSS}{TSS}\Big/(n-k)} = \frac{R^2/(k-1)}{\left(1 - R^2\right)/(n-k)}$$

It is easily demonstrated that *RSS*/*TSS* is equal to 1 – R^2.

F TEST OF GOODNESS OF FIT

Model $$Y = \beta_1 + \beta_2 X + u$$

Null hypothesis: $$H_0\!: \beta_2 = 0$$

Alternative hypothesis: $$H_1\!: \beta_2 \neq 0$$

$$\sum \left(Y - \bar{Y}\right)^2 = \sum \left(\hat{Y} - \bar{Y}\right)^2 + \sum \hat{u}^2 \qquad TSS = ESS + RSS$$

$$R^2 = \frac{ESS}{TSS} = \frac{\sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \left(Y_i - \bar{Y}\right)^2}$$

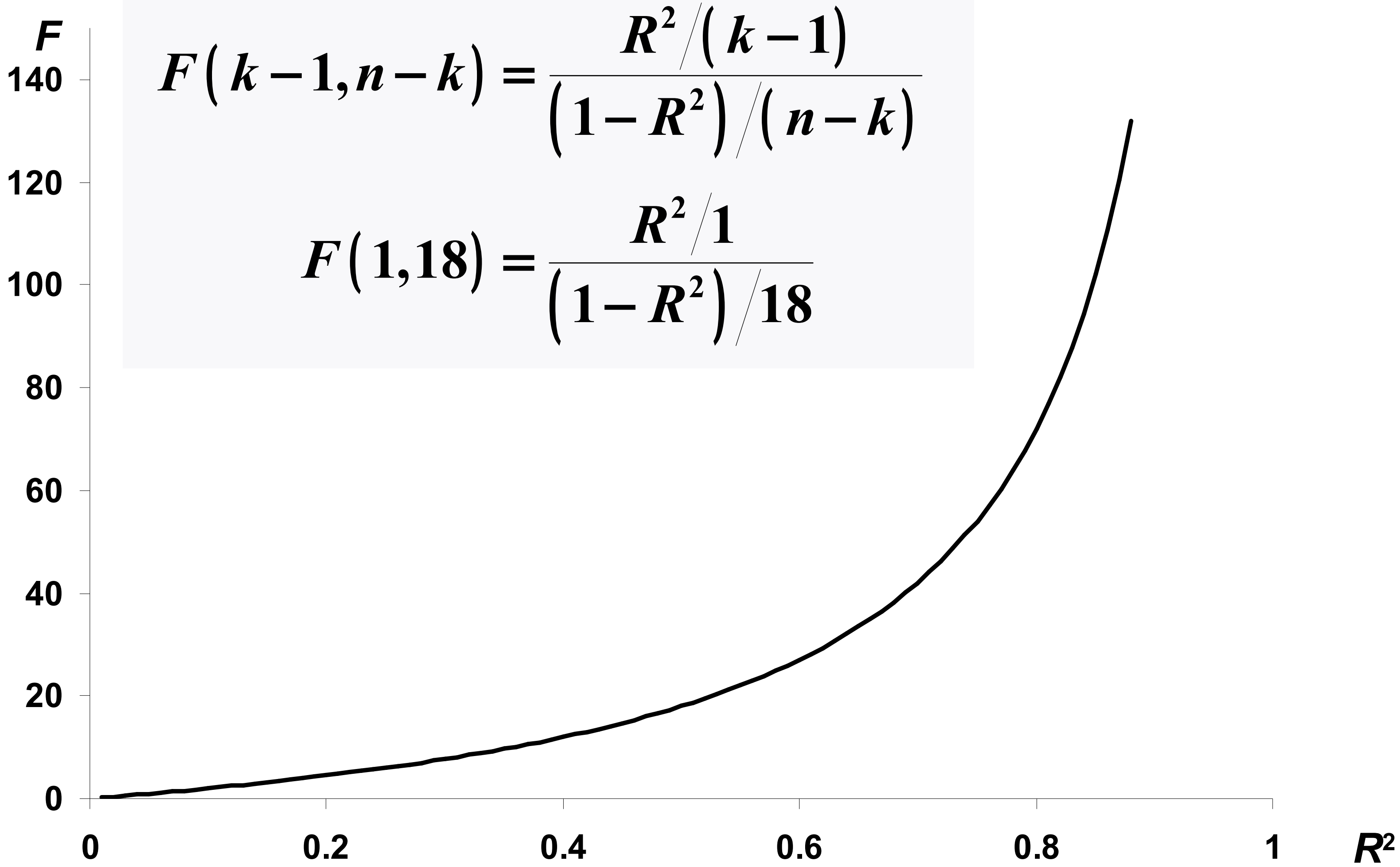
$$F\left(k-1, n-k\right) = \frac{ESS / \left(k-1\right)}{RSS / \left(n-k\right)} = \frac{\dfrac{ESS}{TSS} \Big/ \left(k-1\right)}{\dfrac{RSS}{TSS} \Big/ \left(n-k\right)} = \frac{R^2 / \left(k-1\right)}{\left(1-R^2\right) / \left(n-k\right)}$$

F **is a monotonically increasing function of R^2. As R^2 increases, the numerator increases and the denominator decreases, so for both of these reasons *F* increases.**

F TEST OF GOODNESS OF FIT

$$F(k-1, n-k) = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

$$F(1,18) = \frac{R^2/1}{(1-R^2)/18}$$

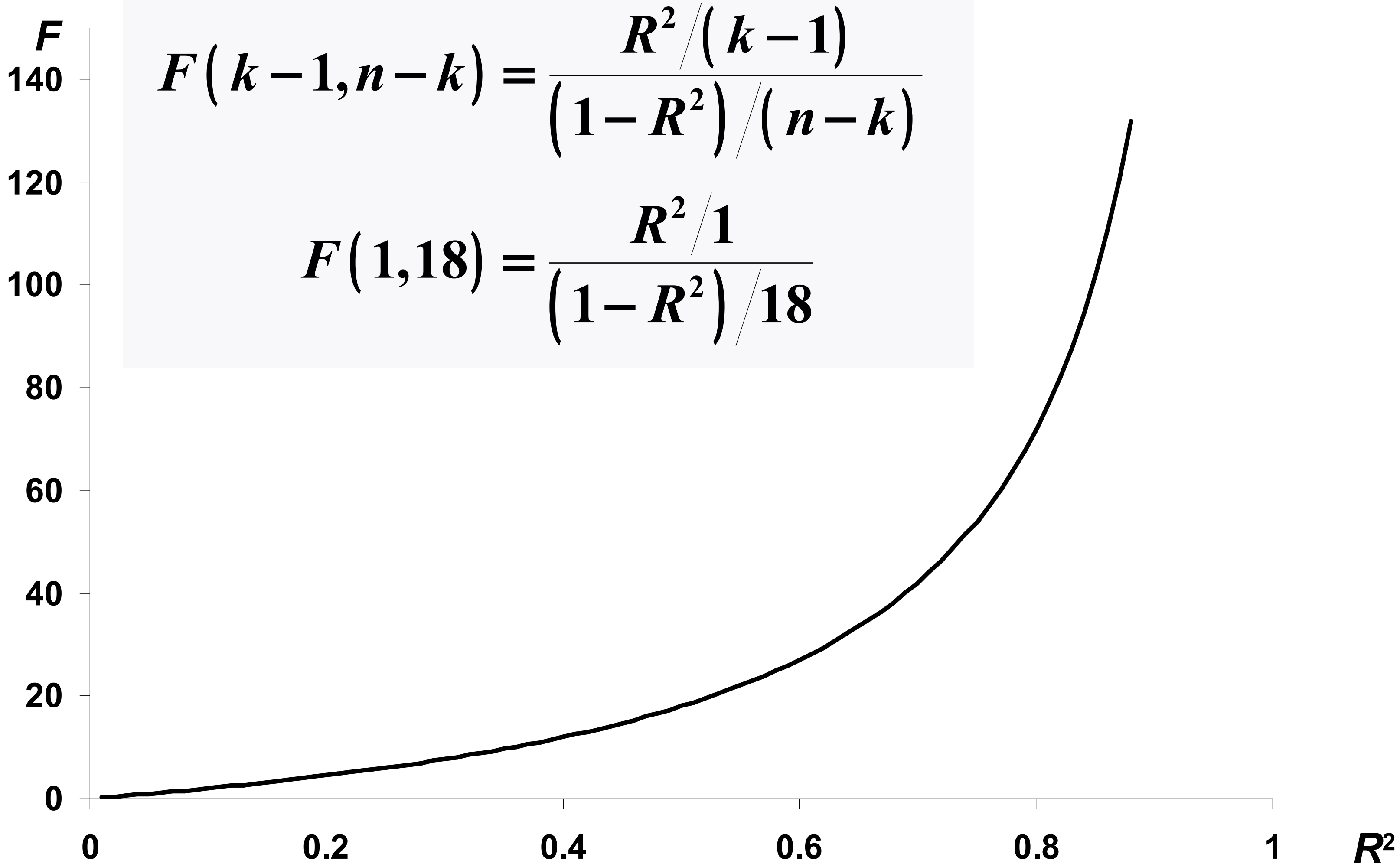


Here is F plotted as a function of R^2 for the case where there is 1 explanatory variable and 20 observations. Since k = 2, $n - k$ = 18.

F TEST OF GOODNESS OF FIT

$$F(k-1, n-k) = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

$$F(1,18) = \frac{R^2/1}{(1-R^2)/18}$$

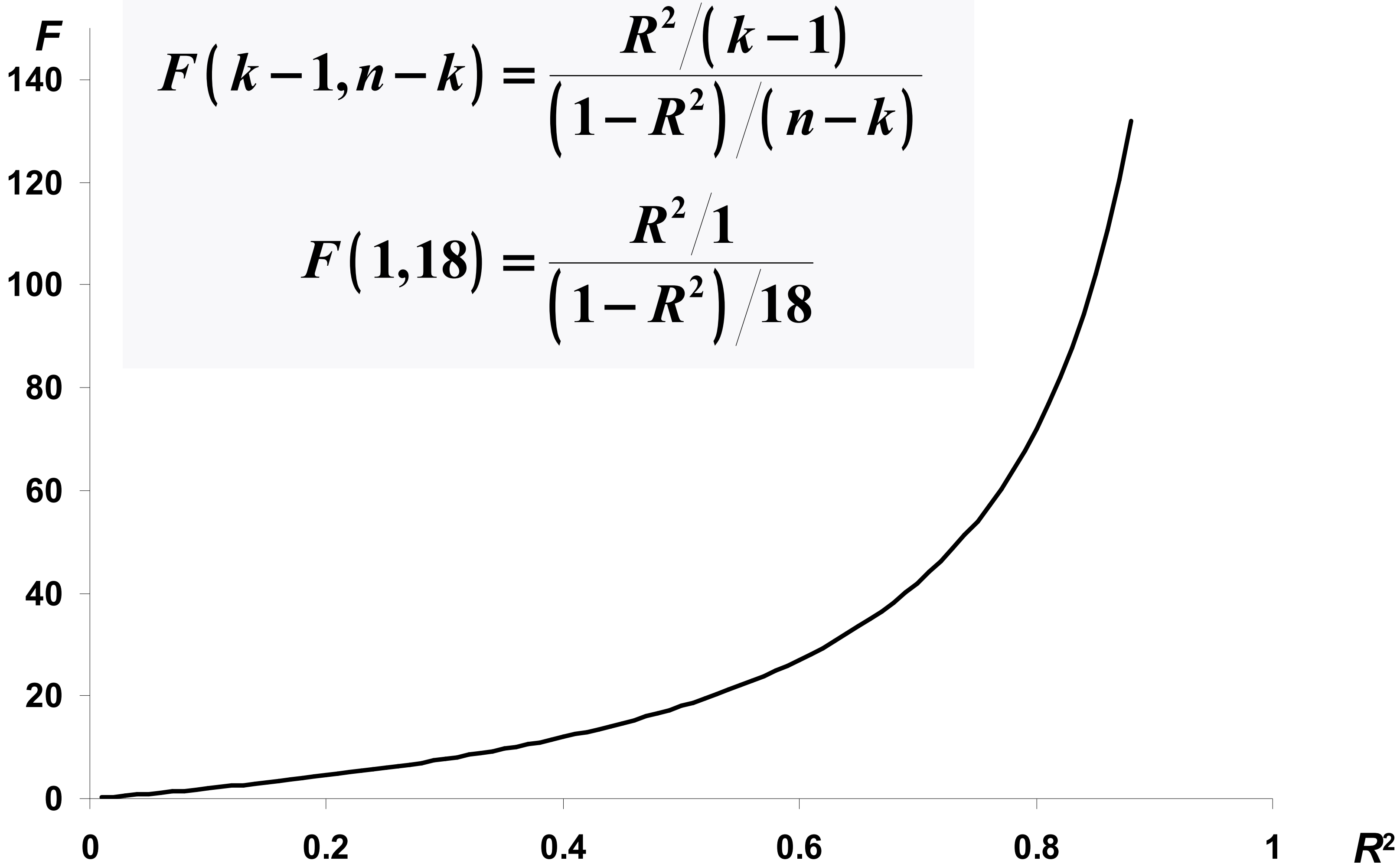


If the null hypothesis is true, F will have a random distribution.

F TEST OF GOODNESS OF FIT

$$F(k-1, n-k) = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

$$F(1,18) = \frac{R^2/1}{(1-R^2)/18}$$

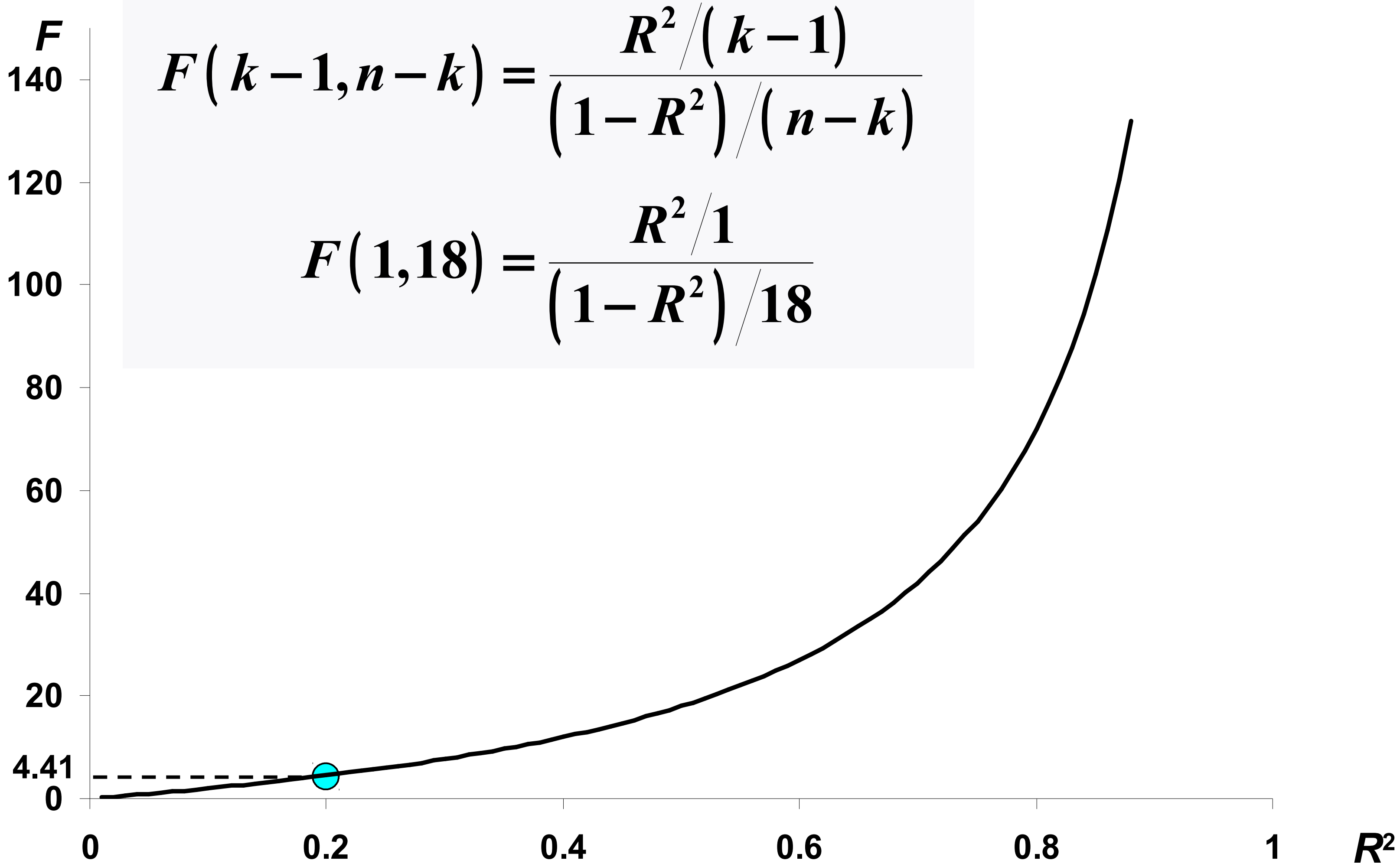


There will be some critical value which it will exceed, as a matter of chance, only 5 percent of the time. If we are performing a 5 percent significance test, we will reject H_0 if the F statistic is greater than this critical value.

F TEST OF GOODNESS OF FIT

$$F(k-1, n-k) = \frac{R^2 / (k-1)}{(1-R^2) / (n-k)}$$

$$F(1,18) = \frac{R^2 / 1}{(1-R^2) / 18}$$



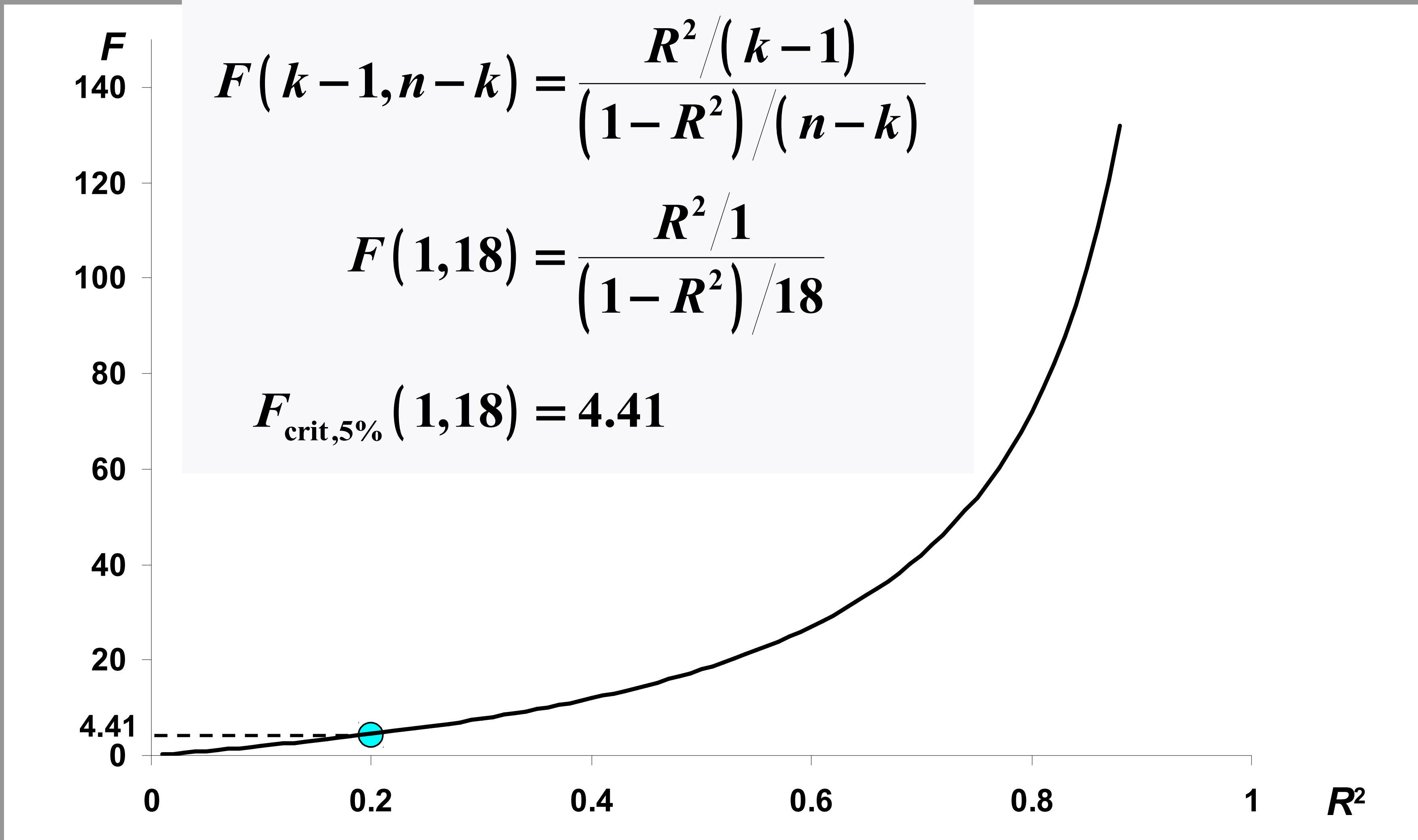
In the case of an F test, the critical value depends on the number of explanatory variables as well as the number of degrees of freedom. When there is one explanatory variable and 18 degrees of freedom, the critical value of F at the 5 percent significance level is 4.41.

F TEST OF GOODNESS OF FIT

$$F(k-1, n-k) = \frac{R^2 / (k-1)}{(1-R^2) / (n-k)}$$

$$F(1,18) = \frac{R^2 / 1}{(1-R^2) / 18}$$

$$F_{\text{crit},5\%}(1,18) = 4.41$$



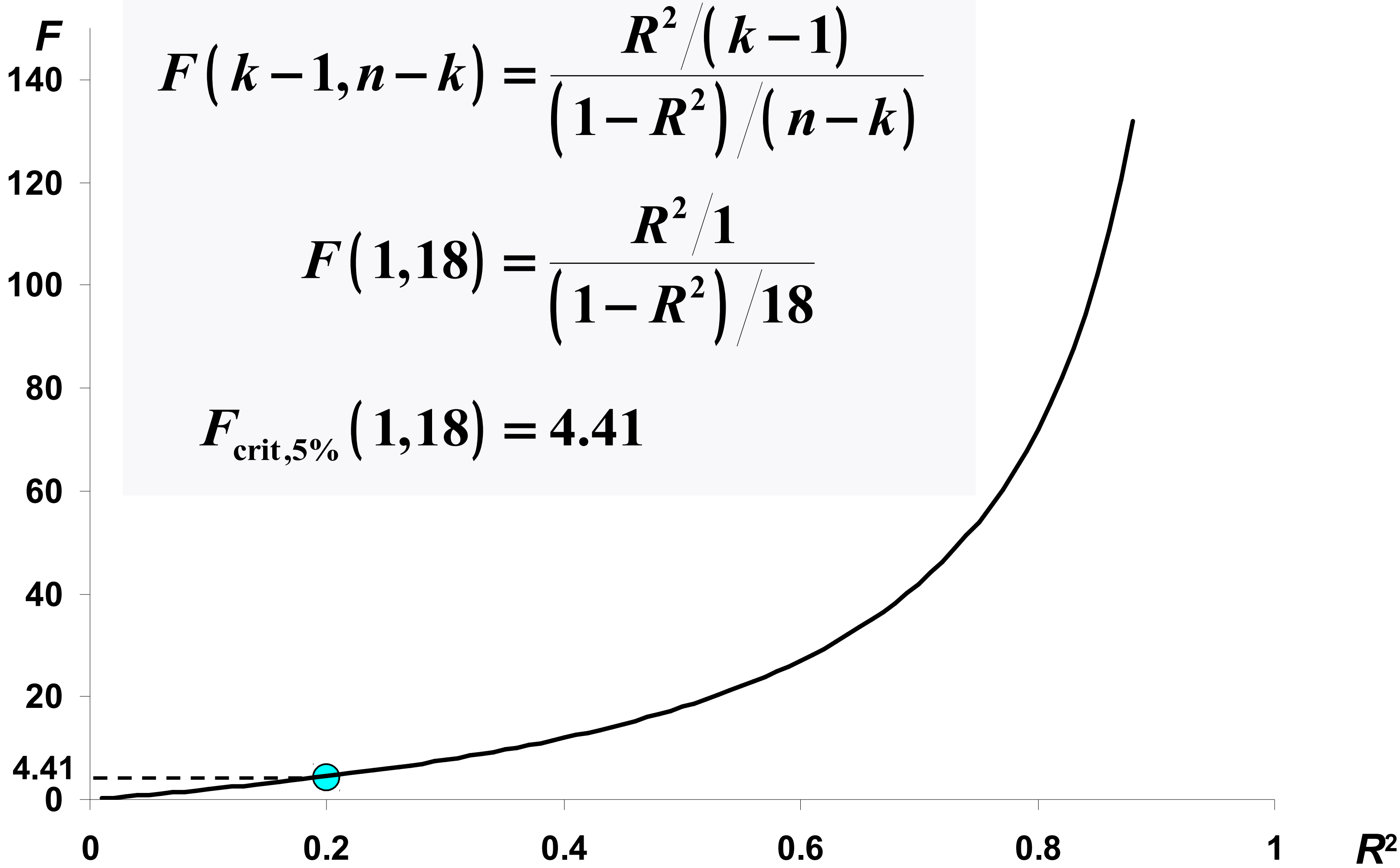
For one explanatory variable and 18 degrees of freedom, F = 4.41 when R^2 = 0.20.

F TEST OF GOODNESS OF FIT

$$F\left(k-1,n-k\right)=\frac{R^2\big/\left(k-1\right)}{\left(1-R^2\right)\big/\left(n-k\right)}$$

$$F\left(1,18\right)=\frac{R^2\big/1}{\left(1-R^2\right)\big/18}$$

$$F_{\text{crit},5\%}\left(1,18\right)=4.41$$



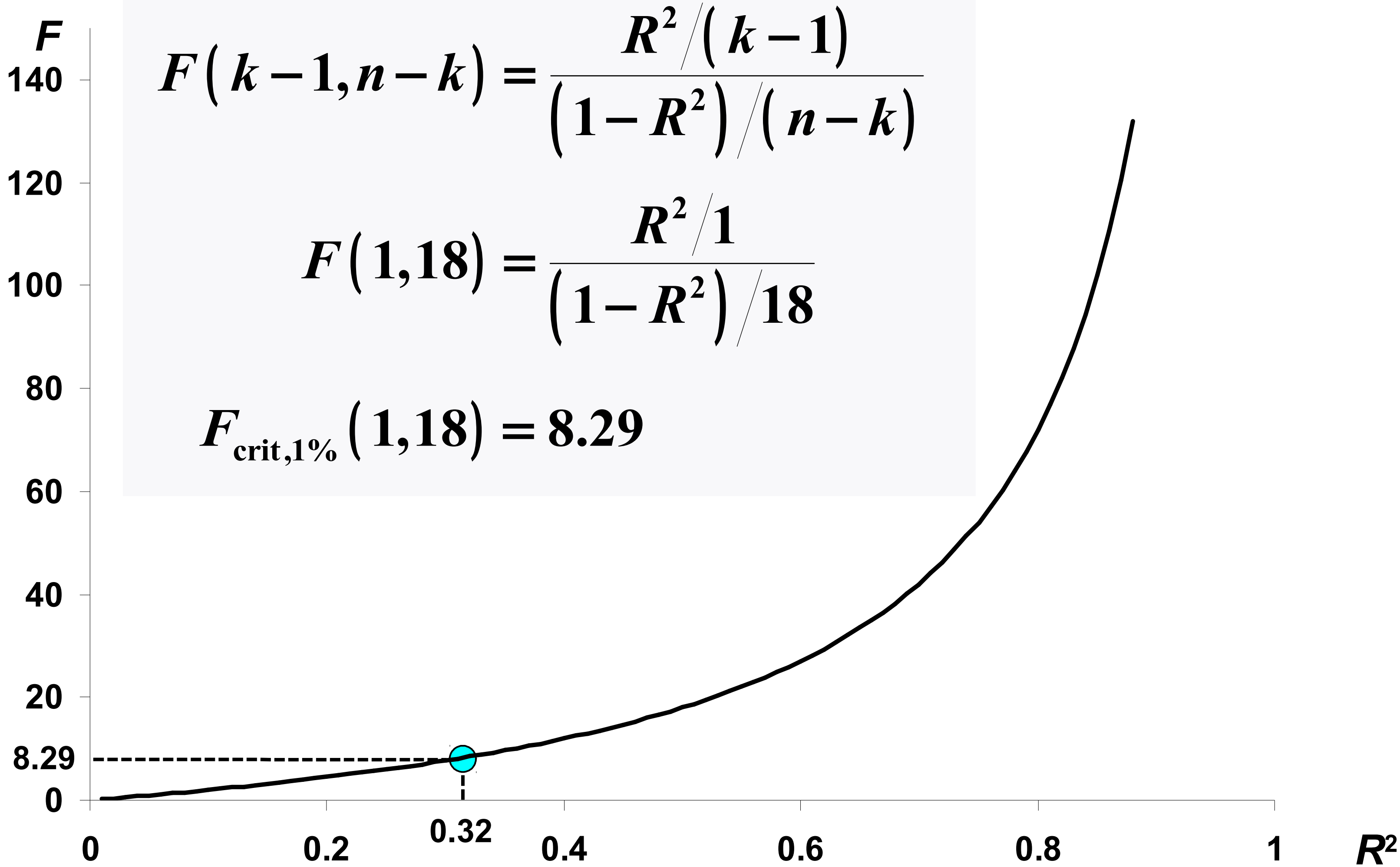
If R^2 is higher than 0.20, F will be higher than 4.41, and we will reject the null hypothesis at the 5 percent level.

F TEST OF GOODNESS OF FIT

$$F(k-1, n-k) = \frac{R^2 / (k-1)}{(1-R^2)/(n-k)}$$

$$F(1,18) = \frac{R^2/1}{(1-R^2)/18}$$

$$F_{\text{crit},1\%}(1,18) = 8.29$$



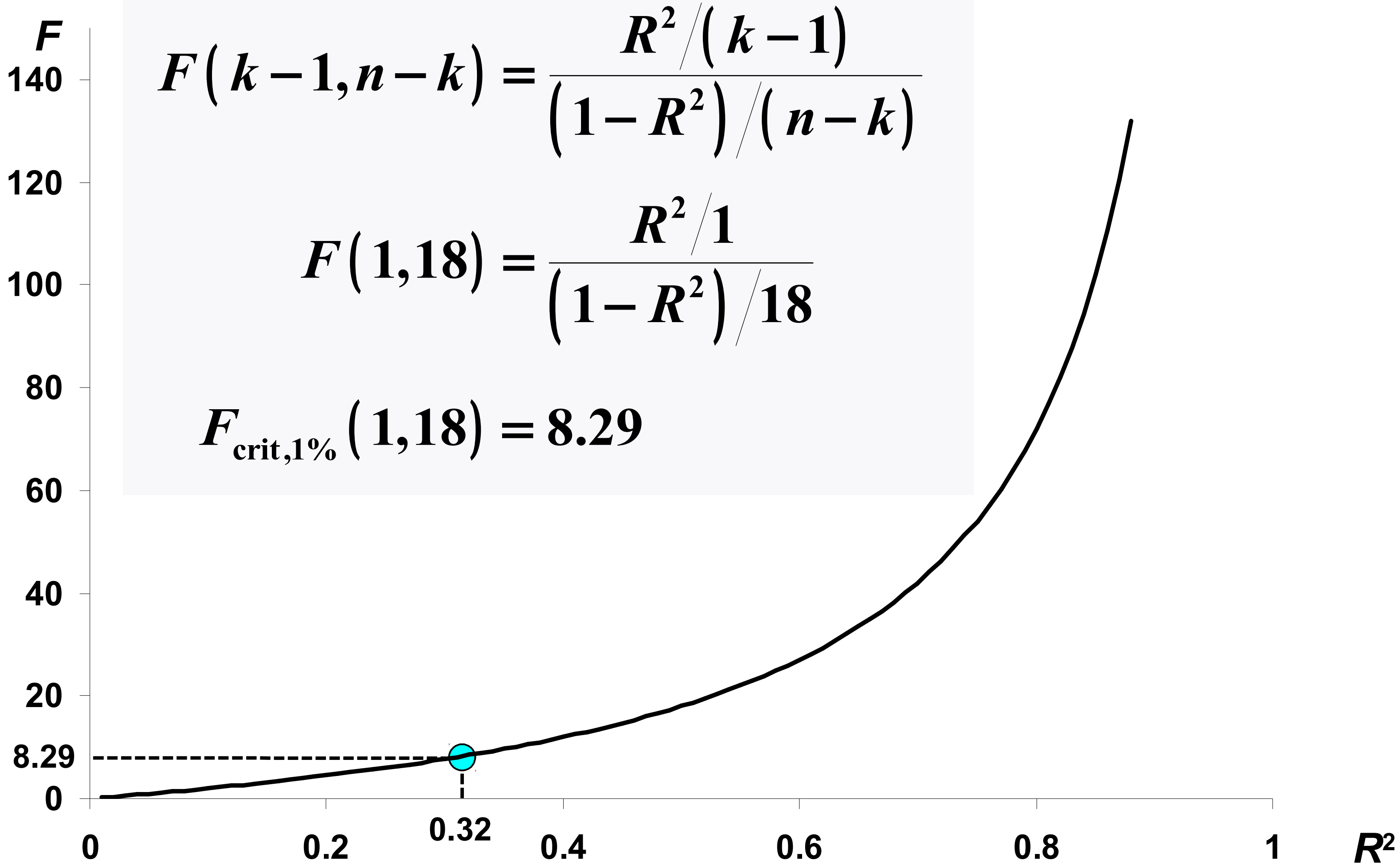
If we were performing a 1 percent test, with one explanatory variable and 18 degrees of freedom, the critical value of F would be 8.29. F = 8.29 when, R^2 = 0.32.

F TEST OF GOODNESS OF FIT

$$F\left(k-1, n-k\right) = \frac{R^2 / (k-1)}{(1-R^2)/(n-k)}$$

$$F\left(1, 18\right) = \frac{R^2 / 1}{(1-R^2)/18}$$

$$F_{\text{crit},1\%}\left(1, 18\right) = 8.29$$



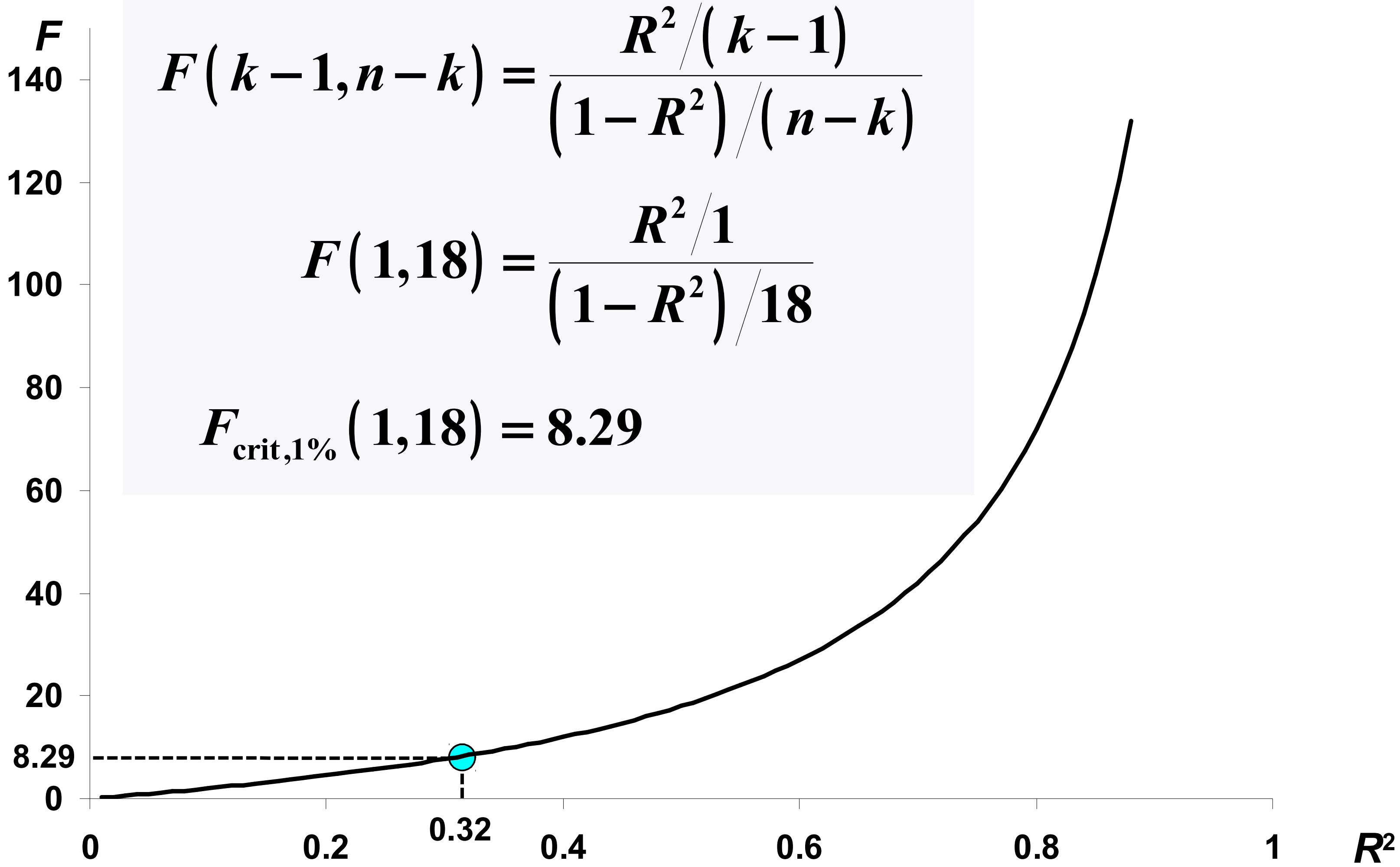
If R^2 is higher than 0.32, F will be higher than 8.29, and we will reject the null hypothesis at the 1 percent level.

F TEST OF GOODNESS OF FIT

$$F(k-1, n-k) = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

$$F(1,18) = \frac{R^2/1}{(1-R^2)/18}$$

$$F_{\text{crit},1\%}(1,18) = 8.29$$



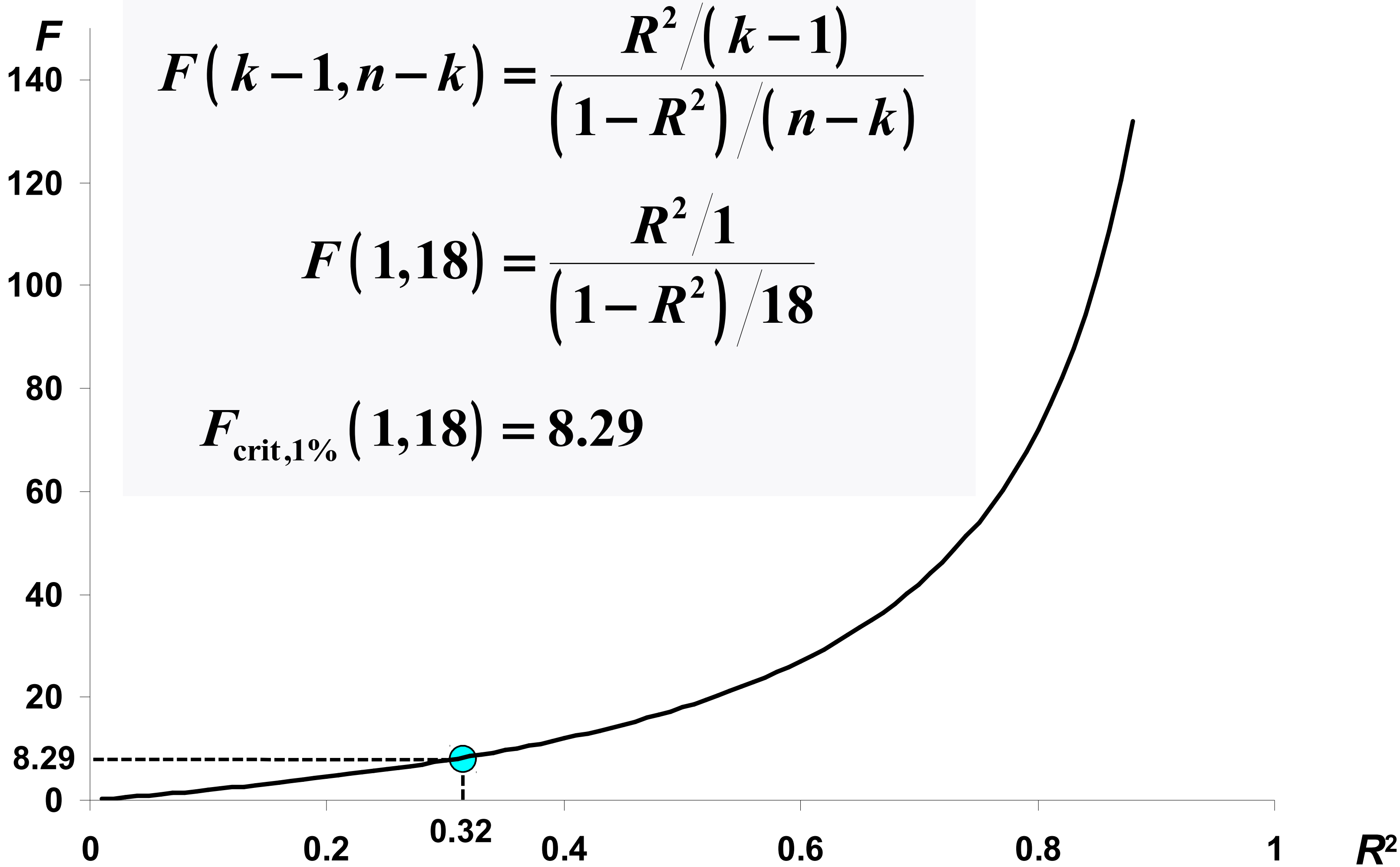
Why do we perform the test indirectly, through F, instead of directly through R^2? After all, it would be easy to compute the critical values of R^2 from those for F.

F TEST OF GOODNESS OF FIT

$$F(k-1, n-k) = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

$$F(1,18) = \frac{R^2/1}{(1-R^2)/18}$$

$$F_{\text{crit},1\%}(1,18) = 8.29$$



The reason is that an *F* test can be used for several tests of analysis of variance. Rather than have a specialized table for each test, it is more convenient to have just one.

F TEST OF GOODNESS OF FIT

Model $Y = \beta_1 + \beta_2 X + u$

Null hypothesis: $H_0: \beta_2 = 0$

Alternative hypothesis: $H_1: \beta_2 \neq 0$

$$\sum \left(Y - \bar{Y}\right)^2 = \sum \left(\hat{Y} - \bar{Y}\right)^2 + \sum \hat{u}^2 \qquad TSS = ESS + RSS$$

$$R^2 = \frac{ESS}{TSS} = \frac{\sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \left(Y_i - \bar{Y}\right)^2}$$

$$F\left(k-1, n-k\right) = \frac{ESS/\left(k-1\right)}{RSS/\left(n-k\right)} = \frac{\frac{ESS}{TSS}\Big/\left(k-1\right)}{\frac{RSS}{TSS}\Big/\left(n-k\right)} = \frac{R^2/\left(k-1\right)}{\left(1-R^2\right)/\left(n-k\right)}$$

Note that, for simple regression analysis, the null and alternative hypotheses are mathematically exactly the same as for a two-tailed *t* test. Could the *F* test come to a different conclusion from the *t* test?

F TEST OF GOODNESS OF FIT

Model $Y = \beta_1 + \beta_2 X + u$

Null hypothesis: $H_0: \beta_2 = 0$

Alternative hypothesis: $H_1: \beta_2 \neq 0$

$$\sum \left(Y - \bar{Y}\right)^2 = \sum \left(\hat{Y} - \bar{Y}\right)^2 + \sum \hat{u}^2 \qquad TSS = ESS + RSS$$

$$R^2 = \frac{ESS}{TSS} = \frac{\sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \left(Y_i - \bar{Y}\right)^2}$$

$$F(k-1, n-k) = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{\frac{ESS}{TSS}\Big/(k-1)}{\frac{RSS}{TSS}\Big/(n-k)} = \frac{R^2/(k-1)}{\left(1-R^2\right)/(n-k)}$$

The answer, of course, is no. We will demonstrate that, for simple regression analysis, the *F* statistic is the square of the *t* statistic.

Demonstration that $F = t^2$

$$F = \frac{ESS}{RSS/(n-2)} = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \hat{u}_i^2/(n-2)}$$

$$= \frac{\sum\left(\left(\hat{\beta}_1 + \hat{\beta}_2 X_i\right) - \left(\hat{\beta}_1 + \hat{\beta}_2 \bar{X}\right)\right)^2}{\hat{\sigma}_u^2} = \frac{1}{\hat{\sigma}_u^2}\sum \hat{\beta}_2^2\left(X_i - \bar{X}\right)^2$$

$$= \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2}\sum\left(X_i - \bar{X}\right)^2 = \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2 / \sum\left(X_i - \bar{X}\right)^2} = \frac{\hat{\beta}_2^2}{\left(\text{s.e.}\left(\hat{\beta}_2\right)\right)^2} = t^2$$

We start by replacing *ESS* and *RSS* by their mathematical expressions.

F TEST OF GOODNESS OF FIT

Demonstration that $F = t^2$

$$F = \frac{ESS}{RSS/(n-2)} = \frac{\sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \hat{u}_i^2 / (n-2)}$$

$$= \frac{\sum \left(\left(\hat{\beta}_1 + \hat{\beta}_2 X_i\right) - \left(\hat{\beta}_1 + \hat{\beta}_2 \bar{X}\right) \right)^2}{\hat{\sigma}_u^2} = \frac{1}{\hat{\sigma}_u^2} \sum \hat{\beta}_2^2 \left(X_i - \bar{X} \right)^2$$

$$= \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2} \sum \left(X_i - \bar{X} \right)^2 = \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2 / \sum \left(X_i - \bar{X} \right)^2} = \frac{\hat{\beta}_2^2}{\left(\text{s.e.}\left(\hat{\beta}_2\right) \right)^2} = t^2$$

The denominator is the expression for $\hat{\sigma}_u^2$, the estimator of σ_u^2, for the simple regression model. We expand the numerator using the expression for the fitted relationship.

Demonstration that $F = t^2$

$$F = \frac{ESS}{RSS/(n-2)} = \frac{\sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \hat{u}_i^2 / (n-2)}$$

$$= \frac{\sum \left(\left(\hat{\beta}_1 + \hat{\beta}_2 X_i \right) - \left(\hat{\beta}_1 + \hat{\beta}_2 \bar{X} \right) \right)^2}{\hat{\sigma}_u^2} = \frac{1}{\hat{\sigma}_u^2} \sum \hat{\beta}_2^2 \left(X_i - \bar{X} \right)^2$$

$$= \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2} \sum \left(X_i - \bar{X} \right)^2 = \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2 / \sum \left(X_i - \bar{X} \right)^2} = \frac{\hat{\beta}_2^2}{\left(\text{s.e.}\left(\hat{\beta}_2 \right) \right)^2} = t^2$$

The $\hat{\beta}_1$ terms in the numerator cancel. The rest of the numerator can be grouped as shown.

F TEST OF GOODNESS OF FIT

Demonstration that $F = t^2$

$$F = \frac{ESS}{RSS/(n-2)} = \frac{\sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \hat{u}_i^2 / (n-2)}$$

$$= \frac{\sum \left(\left(\hat{\beta}_1 + \hat{\beta}_2 X_i \right) - \left(\hat{\beta}_1 + \hat{\beta}_2 \bar{X} \right) \right)^2}{\hat{\sigma}_u^2} = \frac{1}{\hat{\sigma}_u^2} \sum \hat{\beta}_2^2 \left(X_i - \bar{X} \right)^2$$

$$= \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2} \sum \left(X_i - \bar{X} \right)^2 = \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2 / \sum \left(X_i - \bar{X} \right)^2} = \frac{\hat{\beta}_2^2}{\left(\text{s.e.}\left(\hat{\beta}_2 \right) \right)^2} = t^2$$

We take the $\hat{\beta}_2^2$ term out of the summation as a factor.

Demonstration that $F = t^2$

$$F = \frac{ESS}{RSS/(n-2)} = \frac{\sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \hat{u}_i^2 / (n-2)}$$

$$= \frac{\sum \left(\left(\hat{\beta}_1 + \hat{\beta}_2 X_i\right) - \left(\hat{\beta}_1 + \hat{\beta}_2 \bar{X}\right) \right)^2}{\hat{\sigma}_u^2} = \frac{1}{\hat{\sigma}_u^2} \sum \hat{\beta}_2^2 \left(X_i - \bar{X} \right)^2$$

$$= \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2} \sum \left(X_i - \bar{X} \right)^2 = \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2 / \sum \left(X_i - \bar{X} \right)^2} = \frac{\hat{\beta}_2^2}{\left(\text{s.e.}\left(\hat{\beta}_2\right) \right)^2} = t^2$$

We move the term involving *X* to the denominator.

F TEST OF GOODNESS OF FIT

Demonstration that *F* = *t*²

$$F = \frac{ESS}{RSS/(n-2)} = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \hat{u}_i^2/(n-2)}$$

$$= \frac{\sum\left(\left(\hat{\beta}_1 + \hat{\beta}_2 X_i\right) - \left(\hat{\beta}_1 + \hat{\beta}_2 \bar{X}\right)\right)^2}{\hat{\sigma}_u^2} = \frac{1}{\hat{\sigma}_u^2}\sum \hat{\beta}_2^2\left(X_i - \bar{X}\right)^2$$

$$= \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2}\sum\left(X_i - \bar{X}\right)^2 = \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2 / \sum\left(X_i - \bar{X}\right)^2} = \frac{\hat{\beta}_2^2}{\left(\text{s.e.}\left(\hat{\beta}_2\right)\right)^2} = t^2$$

The denominator is the square of the standard error of $\hat{\beta}_2$.

F TEST OF GOODNESS OF FIT

Demonstration that $F = t^2$

$$F = \frac{ESS}{RSS/(n-2)} = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \hat{u}_i^2/(n-2)}$$

$$= \frac{\sum\left(\left(\hat{\beta}_1 + \hat{\beta}_2 X_i\right) - \left(\hat{\beta}_1 + \hat{\beta}_2 \bar{X}\right)\right)^2}{\hat{\sigma}_u^2} = \frac{1}{\hat{\sigma}_u^2}\sum \hat{\beta}_2^2\left(X_i - \bar{X}\right)^2$$

$$= \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2}\sum\left(X_i - \bar{X}\right)^2 = \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2 / \sum\left(X_i - \bar{X}\right)^2} = \frac{\hat{\beta}_2^2}{\left(\text{s.e.}\left(\hat{\beta}_2\right)\right)^2} = t^2$$

Hence we obtain $\hat{\beta}_2^2$ divided by the square of the standard error of $\hat{\beta}_2$. This is the *t* statistic, squared.

F TEST OF GOODNESS OF FIT

Demonstration that $F = t^2$

$$F = \frac{ESS}{RSS/(n-2)} = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \hat{u}_i^2/(n-2)}$$

$$= \frac{\sum\left(\left(\hat{\beta}_1 + \hat{\beta}_2 X_i\right) - \left(\hat{\beta}_1 + \hat{\beta}_2 \bar{X}\right)\right)^2}{\hat{\sigma}_u^2} = \frac{1}{\hat{\sigma}_u^2}\sum \hat{\beta}_2^2\left(X_i - \bar{X}\right)^2$$

$$= \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2}\sum\left(X_i - \bar{X}\right)^2 = \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2/\sum\left(X_i - \bar{X}\right)^2} = \frac{\hat{\beta}_2^2}{\left(\text{s.e.}\left(\hat{\beta}_2\right)\right)^2} = t^2$$

It can also be shown that the critical value of *F*, at any significance level, is equal to the square of the critical value of *t*. We will not attempt to prove this.

Demonstration that $F = t^2$

$$F = \frac{ESS}{RSS/(n-2)} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum \hat{u}_i^2/(n-2)}$$

$$= \frac{\sum \left((\hat{\beta}_1 + \hat{\beta}_2 X_i) - (\hat{\beta}_1 + \hat{\beta}_2 \bar{X}) \right)^2}{\hat{\sigma}_u^2} = \frac{1}{\hat{\sigma}_u^2} \sum \hat{\beta}_2^2 (X_i - \bar{X})^2$$

$$= \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2} \sum (X_i - \bar{X})^2 = \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2 / \sum (X_i - \bar{X})^2} = \frac{\hat{\beta}_2^2}{\left(\text{s.e.}(\hat{\beta}_2) \right)^2} = t^2$$

Since the *F* test is equivalent to a two-sided *t* test in the simple regression model, there is no point in performing both tests. In fact, if justified, a one-sided *t* test would be better than either because it is more powerful (lower risk of Type II error if H_0 is false).

Demonstration that $F = t^2$

$$F = \frac{ESS}{RSS/(n-2)} = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \hat{u}_i^2/(n-2)}$$

$$= \frac{\sum\left(\left(\hat{\beta}_1 + \hat{\beta}_2 X_i\right) - \left(\hat{\beta}_1 + \hat{\beta}_2 \bar{X}\right)\right)^2}{\hat{\sigma}_u^2} = \frac{1}{\hat{\sigma}_u^2} \sum \hat{\beta}_2^2 \left(X_i - \bar{X}\right)^2$$

$$= \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2} \sum\left(X_i - \bar{X}\right)^2 = \frac{\hat{\beta}_2^2}{\hat{\sigma}_u^2 / \sum\left(X_i - \bar{X}\right)^2} = \frac{\hat{\beta}_2^2}{\left(\text{s.e.}\left(\hat{\beta}_2\right)\right)^2} = t^2$$

The *F* test will have its own role to play when we come to multiple regression analysis.

F TEST OF GOODNESS OF FIT

```
. reg EARNINGS S

----------------------------------------------------------------------------
    Source |       SS       df       MS              Number of obs =     500
-----------+------------------------------           F(  1,   498) =   46.57
     Model |  6014.04474     1  6014.04474           Prob > F      =  0.0000
  Residual |  64314.9215   498  129.146429           R-squared     =  0.0855
-----------+------------------------------           Adj R-squared =  0.0837
     Total |  70328.9662   499  140.939812           Root MSE      =  11.364
----------------------------------------------------------------------------
  EARNINGS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------+----------------------------------------------------------------
         S |   1.265712   .1854782     6.82   0.000     .9012959    1.630128
     _cons |   .7646844   2.803765     0.27   0.785    -4.743982    6.273351
----------------------------------------------------------------------------
```

Here is the output for the regression of hourly earnings on years of schooling for the sample of 500 respondents from the National Longitudinal Survey of Youth 1997–.

F TEST OF GOODNESS OF FIT

```
. reg EARNINGS S

------------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =   46.57
       Model |  6014.04474     1  6014.04474           Prob > F      =  0.0000
    Residual |  64314.9215   498  129.146429           R-squared     =  0.0855
-------------+------------------------------           Adj R-squared =  0.0837
       Total |  70328.9662   499  140.939812           Root MSE      =  11.364
------------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           S |   1.265712   .1854782     6.82   0.000     .9012959    1.630128
       _cons |   .7646844   2.803765     0.27   0.785    -4.743982    6.273351
------------------------------------------------------------------------------
```

$$F(1, n-2) = \frac{ESS}{RSS/(n-2)} = \frac{6014}{64315/(500-2)} = \frac{6014}{129.15} = 46.57$$

We shall check that the *F* statistic has been calculated correctly. The explained sum of squares (described in Stata as the model sum of squares) is 6014.

F TEST OF GOODNESS OF FIT

```
. reg EARNINGS S

------------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =   46.57
       Model |  6014.04474     1  6014.04474           Prob > F      =  0.0000
    Residual |  64314.9215   498  129.146429           R-squared     =  0.0855
-------------+------------------------------           Adj R-squared =  0.0837
       Total |  70328.9662   499  140.939812           Root MSE      =  11.364
------------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           S |   1.265712   .1854782     6.82   0.000     .9012959    1.630128
       _cons |   .7646844   2.803765     0.27   0.785    -4.743982    6.273351
------------------------------------------------------------------------------
```

$$F(1, n-2) = \frac{ESS}{RSS/(n-2)} = \frac{6014}{64315/(500-2)} = \frac{6014}{129.15} = 46.57$$

The residual sum of squares is 64315.

F TEST OF GOODNESS OF FIT

```
. reg EARNINGS S

----------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =   46.57
       Model |  6014.04474     1  6014.04474           Prob > F      =  0.0000
    Residual |  64314.9215   498  129.146429           R-squared     =  0.0855
-------------+------------------------------           Adj R-squared =  0.0837
       Total |  70328.9662   499  140.939812           Root MSE      =  11.364
----------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+--------------------------------------------------------------
           S |   1.265712   .1854782     6.82   0.000     .9012959    1.630128
       _cons |   .7646844   2.803765     0.27   0.785    -4.743982    6.273351
----------------------------------------------------------------------------
```

$$F(1, n-2) = \frac{ESS}{RSS/(n-2)} = \frac{6014}{64315/(500-2)} = \frac{6014}{129.15} = 46.57$$

The number of degrees of freedom is 500 – 2 = 498.

F TEST OF GOODNESS OF FIT

```
. reg EARNINGS S

------------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =   46.57
       Model |  6014.04474     1  6014.04474           Prob > F      =  0.0000
    Residual |  64314.9215   498  129.146429           R-squared     =  0.0855
-------------+------------------------------           Adj R-squared =  0.0837
       Total |  70328.9662   499  140.939812           Root MSE      =  11.364
------------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           S |   1.265712   .1854782     6.82   0.000     .9012959    1.630128
       _cons |   .7646844   2.803765     0.27   0.785    -4.743982    6.273351
------------------------------------------------------------------------------
```

$$F(1, n-2) = \frac{ESS}{RSS/(n-2)} = \frac{6014}{64315/(500-2)} = \frac{6014}{129.15} = 46.57$$

The denominator of the expression for *F* is therefore 129.15. Note that this is an estimate of σ_u^2. Its square root, denoted in Stata by Root MSE, is an estimate of the standard deviation of *u*.

F TEST OF GOODNESS OF FIT

```
. reg EARNINGS S

----------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =   46.57
       Model |  6014.04474     1  6014.04474           Prob > F      =  0.0000
    Residual |  64314.9215   498  129.146429           R-squared     =  0.0855
-------------+------------------------------           Adj R-squared =  0.0837
       Total |  70328.9662   499  140.939812           Root MSE      =  11.364

----------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+--------------------------------------------------------------
           S |   1.265712   .1854782     6.82   0.000     .9012959    1.630128
       _cons |   .7646844   2.803765     0.27   0.785    -4.743982    6.273351
----------------------------------------------------------------------------
```

$$F(1, n-2) = \frac{ESS}{RSS/(n-2)} = \frac{6014}{64315/(500-2)} = \frac{6014}{129.15} = 46.57$$

Our calculation of *F* agrees with that in the Stata output.

F TEST OF GOODNESS OF FIT

```
. reg EARNINGS S

------------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =   46.57
       Model |  6014.04474     1  6014.04474           Prob > F      =  0.0000
    Residual |  64314.9215   498  129.146429           R-squared     =  0.0855
-------------+------------------------------           Adj R-squared =  0.0837
       Total |  70328.9662   499  140.939812           Root MSE      =  11.364
------------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           S |   1.265712   .1854782     6.82   0.000     .9012959    1.630128
       _cons |   .7646844   2.803765     0.27   0.785    -4.743982    6.273351
------------------------------------------------------------------------------
```

$$F(1, n-2) = \frac{R^2}{(1-R^2)/(n-2)} = \frac{0.0855}{(1-0.0855)/(500-2)} = 46.56$$

We will also check the *F* statistic using the expression for it in terms of R^2. We see again that it agrees, apart from rounding error.

```
. reg EARNINGS S

----------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =   46.57
       Model |  6014.04474     1  6014.04474           Prob > F      =  0.0000
    Residual |  64314.9215   498  129.146429           R-squared     =  0.0855
-------------+------------------------------           Adj R-squared =  0.0837
       Total |  70328.9662   499  140.939812           Root MSE      =  11.364
----------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval]
-------------+--------------------------------------------------------------
           S |   1.265712   .1854782     6.82   0.000     .9012959    1.630128
       _cons |   .7646844   2.803765     0.27   0.785    -4.743982    6.273351
----------------------------------------------------------------------------
```

We will also check the relationship between the *F* statistic and the *t* statistic for the slope coefficient.

F TEST OF GOODNESS OF FIT

```
. reg EARNINGS S

----------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =   46.57
       Model |  6014.04474     1  6014.04474           Prob > F      =  0.0000
    Residual |  64314.9215   498  129.146429           R-squared     =  0.0855
-------------+------------------------------           Adj R-squared =  0.0837
       Total |  70328.9662   499  140.939812           Root MSE      =  11.364
----------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+--------------------------------------------------------------
           S |   1.265712   .1854782     6.82   0.000     .9012959    1.630128
       _cons |   .7646844   2.803765     0.27   0.785    -4.743982    6.273351
----------------------------------------------------------------------------
```

$$6.82^2 = 46.51$$

Obviously, this is correct as well, apart from rounding error.

F TEST OF GOODNESS OF FIT

```
. reg EARNINGS S

----------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =   46.57
       Model |  6014.04474     1  6014.04474           Prob > F      =  0.0000
    Residual |  64314.9215   498  129.146429           R-squared     =  0.0855
-------------+------------------------------           Adj R-squared =  0.0837
       Total |  70328.9662   499  140.939812           Root MSE      =  11.364
----------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t    P>|t|      [95% Conf. Interval]
-------------+--------------------------------------------------------------
           S |   1.265712   .1854782     6.82   0.000      .9012959    1.630128
       _cons |   .7646844   2.803765     0.27   0.785     -4.743982    6.273351
----------------------------------------------------------------------------
```

$$F_{\text{crit, 0.1\%}}(1{,}500) = 10.96 \qquad t_{\text{crit, 0.1\%}}(500) = 3.31 \qquad 10.96 = 3.31^2$$

And the critical value of *F* is the square of the critical value of *t*. (We are using the values for 500 degrees of freedom because those for 498 do not appear in the table.)

F TEST OF GOODNESS OF FIT

```
. reg EARNINGS S

----------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =   46.57
       Model |  6014.04474     1  6014.04474           Prob > F      =  0.0000
    Residual |  64314.9215   498  129.146429           R-squared     =  0.0855
-------------+------------------------------           Adj R-squared =  0.0837
       Total |  70328.9662   499  140.939812           Root MSE      =  11.364
----------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+--------------------------------------------------------------
           S |   1.265712   .1854782     6.82   0.000     .9012959    1.630128
       _cons |   .7646844   2.803765     0.27   0.785    -4.743982    6.273351
----------------------------------------------------------------------------
```

$$F_{\text{crit, 0.1\%}}(1,500) = 10.96 \qquad t_{\text{crit, 0.1\%}}(500) = 3.31 \qquad 10.96 = 3.31^2$$

The relationship is shown for the 0.1% significance level, but obviously it is also true for any other significance level.

The content of this slideshow comes from Section 2.7 of C. Dougherty, *Introduction to Econometrics*, fifth edition 2016, Oxford University Press. Additional (free) resources for both students and instructors may be downloaded from the OUP Online Resource Centre
http://www.oxfordtextbooks.co.uk/orc/dougherty5e/

Individuals studying econometrics on their own who feel that they might benefit from participation in a formal course should consider the London School of Economics summer school course
EC212 Introduction to Econometrics
http://www2.lse.ac.uk/study/summerSchools/summerSchool/Home.aspx
or the University of London International Programmes distance learning course
EC2020 Elements of Econometrics
www.londoninternational.ac.uk/lse.