# A9Wd The use of F-test to test whether the predictor variable is a significant contributor

The F-test is an alternative to the t-test which can be used to determine whether the predictor variable is a significant contributor to the dependent variable (y). Recall in textbook Section 10.1.3 that the total deviation in y, SST, can be partitioned between the deviation explained by the regression (SSR) and the unexplained deviation (SSE). If the regression model fits the sample data, then we would find that the value of the deviation explained by the regression (SSR) is larger than the value of the unexplained deviation (SSE).

If we take the mean squares and divide by their degrees of freedom, then the ratio MSR/MSE follows an F distribution with k degrees of freedom in the numerator, and n – (k + 1) degrees of freedom in the denominator as defined by equations (W10.5) –(W10.7).

$F\_{cal}=\frac{Mean square for model}{Mean square for error}=\frac{MSR}{MSE}$ (W10.5)

$F\_{cal}=\frac{\left(\frac{SSR}{k}\right)}{\left(\frac{SSE}{n-(k+1)}\right)}$ (W10.6)

$F\_{cal}=\frac{\left(\frac{COD}{k}\right)}{\left(\frac{1-COD}{n-(k+1)}\right)}$ (W10.7)

Where n is the total number of paired values and k is the number of predictor variables. If the regression line fits the sample data (little scatter about line) then the value of F will be quite large. Conversely, if the regression line does not fit the sample data (increased scatter about line) then the value of F will approach zero.

**Example W10.2**

Re-consider textbook Example 10.1 data and conduct an F-test to test whether or not the dependent variable is a significant contributor to the dependent variable. Figures W10.3 and W10.4 illustrate the Excel solution.



Figure W10.3

**Excel solution**

x: Cells C3:C50 Values

y: Cells D3:D50 Values

b0 = Cell C57 Formula:=INTERCEPT (C5:C54,B5:B54)

b1 = Cell C58 Formula:=SLOPE (D5:D54,C5:C54)

ŷ= Cell E3 Formula:=$C$57+$C$58\*B5

 Copy formula down E4:E54



Figure W10.4

**Excel solution**

Level = Cell J12 Value

n = Cell J14 Formula:=COUNT (B5:B54)

k = Cell J15 Value

COD = Cell J16 Formula:=RSQ (D5:D54,C5:C54)

F-cal = Cell J17 Formula:=(J16/J15)/((1-J16)/(J14-(J15+1)))

df num = Cell J19 Formula:=J15

df denom = Cell J20 Formula:=J14-(J15+1)

F-critical = Cell J21 Formula:=F.INV.RT(J11,J19,J20)

p-value = Cell J22 Formula:=F.DIST.RT(J17,J19,J20)

**Step 1 - State hypothesis**

H0: β1 = 0 no linear relationship

H1: β1 0 linear relationship exists and since we believe that the relationship is not zero

**Step 2 - Select test**

Two tail test – which we know is F-test testing whether the predictor variable is a significant contributor

**Step 3 - Set the level of significance** (α = 0.05) (see Cell J11)

**Step 4 - Extract relevant statistic**

Calculate the test statistic. From Excel, Fcal = 110.255

Critical F value, Fcri

Significance = 5% = 0.05

Number of degrees on numerator, dfn = k = 1

Number of degrees on denominator, dfd = n - (k+1) = 48 - (1+1) = 48

From Excel, Fcri = 4.043 and p-value = 5.0007E-14.

**Step 5 - Make decision**

Figure W10.5 illustrates the shape of the F distribution and the relationship between the critical F value and H0 and H1 being true.



**Fcri = 4.043**

**Fcal = 110.255**

**F distribution dfA=1, dfB=48**

Figure W10.5

Since Fcal > Fcri (110.255 > 4.043), we reject H0 and accept H1. Alternatively, use the p-value (5.0007E-14) < 0.05, and conclude that the alternative hypothesis is accepted.

Conclude that the model is useful in predicting the number of UK visits abroad from the percentage of employed in the UK. In fact, we are 95% confident (5% level of significance) that this is the case.

Note that for one predictor models, the t-test and F-test is essentially the same test. In fact, for a one predictor regression model the relationship between F and t is . Check, t = 10.50…., F = 110.255…., $t=\sqrt{F}=\sqrt{110.255}=10.50$ .

The format for the ANOVA table is:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | df | Sum of Squares | Mean Square (Variance) | F |
| Regression | k | SSR |  |  |
| Error | n – k - 1 | SSE |  |  |
| Total | n - 1 | SST |  |  |

Table W10.1 ANOVA table

Where COD = SSR/SST = 1 – SSE/SST.

The completed ANOVA table is part of the Excel Data > Data Analysis > Regression solution.

In accordance with the Example W10.2, k = 1, n = 50 and the ANOVA table would be as presented in Table W10.2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | df | Sum of Squares | Mean Square (Variance) | F |
| Regression | 1 | 5,180,283 | 5,180,283 | 110.255 |
| Error | 48 | 2,255,255 | 46,984.49 |  |
| Total | 49 | 7,435,538 |  |  |

Table W10.2 ANOVA table