Fox Module 15: Advanced interactions

(The attached PDF file has better formatting.)

Homework assignment: F test with interactions

Tables 7.1 and 7.2 on page 139 are tested on the final exam. This homework assignment explains the computations for the F test in these tables.

The variables mean: I = income, E = education, and T = type

The regression sums of squares are

Model	*Terms*	*Sum of Squares*	*df*
1	I, E, T, I × T, E × T	24,794	8
2	I, E, T, I × T	24,556	6
3	I, E, T, E × T	23,842	6
4	I, E, T	23,666	4
5	I, E	23,074	2
6	I, T, I × T	23,488	5
7	E, T, E × T	22,710	5

Table 7.2 shows the degrees of freedom and sum of squares in the numerator of the F test.

Source	*Models*	*Sum of Squares*	*df*	*F*
Income	3 – 7	1,132	1	28.35
Education	2 – 6	1,068	1	26.75
Type	4 – 5	592	2	7.41
Income × Type	1 – 3	952	2	11.92
Education × Type	1 – 2	238	2	2.98
Residuals		3,553	89	
Total		28,347	97	

For each model,

- The residual sum of squares is $\sum\left(Y-\hat{\hat{Y}}\right)^2$.
- The regression sum of squares is $\sum\left(\overline{Y}-\hat{Y}\right)^2$.
- The total sum of squares is $\sum\left(\overline{Y}-Y\right)^2$.

A. Why does the total sum of squares (TSS) not depend on the model? What is the TSS in this illustration?
B. Which model has the smallest residual sum of squares (RSS)? How do we know this even without computing any figures?
C. How do we test the significance of income? What is the null hypothesis? How the F-ratio is computed? (Show the calculations.)
D. How do we test the significance of education × type? What is the null hypothesis? How the F-ratio is computed? (Show the calculations.)

The following comments may help you understand the exhibits:

The degrees of freedom in Table 7.1 on page 139 are the number of explanatory variables in the model (k). The degrees of freedom are actually N-k-1. But this illustration focuses on the degrees of freedom for the numerator of the F test, which is the difference in the number of variables in the full vs reduced models. N-1 is the same for all models, so it drops out of the difference.

For the number of explanatory variables:

- I and E are one explanatory variable each.
- T, I × T, and E × T are two explanatory variables each.

The total sum of squares is 28,347. The sample has 98 data points, so the total sum of squares has 98 – 1 = 97 degrees of freedom. The full model (Model 1) has a regression sum of squares of 24,794, so it has a residual sum of squares of 28,347 – 24,794 = 3,553. This residual sum of squares has 98 – 8 – 1 = 89 degrees of freedom.

Show the calculation of the F-ratio for Parts C and D.