

Answers to Exercises:

Chapter 4: F-Tests

4.1.1 Exercise:

- Suppose the results from the unrestricted regression described above (of inflation on money supply and ten country dummies) are as follows: $RSS_U = 8000$, $n = 212$, k (the number of coefficients including the intercept) = 12. And suppose the results from the restricted regression (inflation regressed on money supply only) are: $RSS_R = 10,000$, $n = 212$, $k = 2$. State the null and alternative hypotheses and calculate the value of F. Use the F-table provided in the **Ftests.xls** template (open the spreadsheet and click on the "F-table extract" worksheet) to complete the hypothesis test at the 5% significance level.

Unrestricted Regression:

$$INFL = a + b MS + g_1 \text{ COUNTRY1} + \dots + g_{10} \text{ COUNTRY10} \quad [1]$$

Restricted Regression:

$$INFL = a + b MS \quad [2]$$

$$H_0: g_1 = g_2 = g_3 = \dots = g_{10} = 0$$

(i.e. ten restrictions imposed in the restricted regression)

$$H_1: g_1, g_2, g_3, \dots, g_{10} \neq 0$$

$$F_{\frac{df_{\text{numerator}}}{df_{\text{denominator}}}} = F_{df_U}^r = \frac{(RSS_R - RSS_U) / r}{RSS_U / df_U}$$

Where:

$$RSS_U = \text{unrestricted residual sum of squares} = \text{RSS under } H_1 = 8000$$

$$RSS_R = \text{restricted residual sum of squares} = \text{RSS under } H_0 = 10,000$$

$$r = \text{number of restrictions} = 10$$

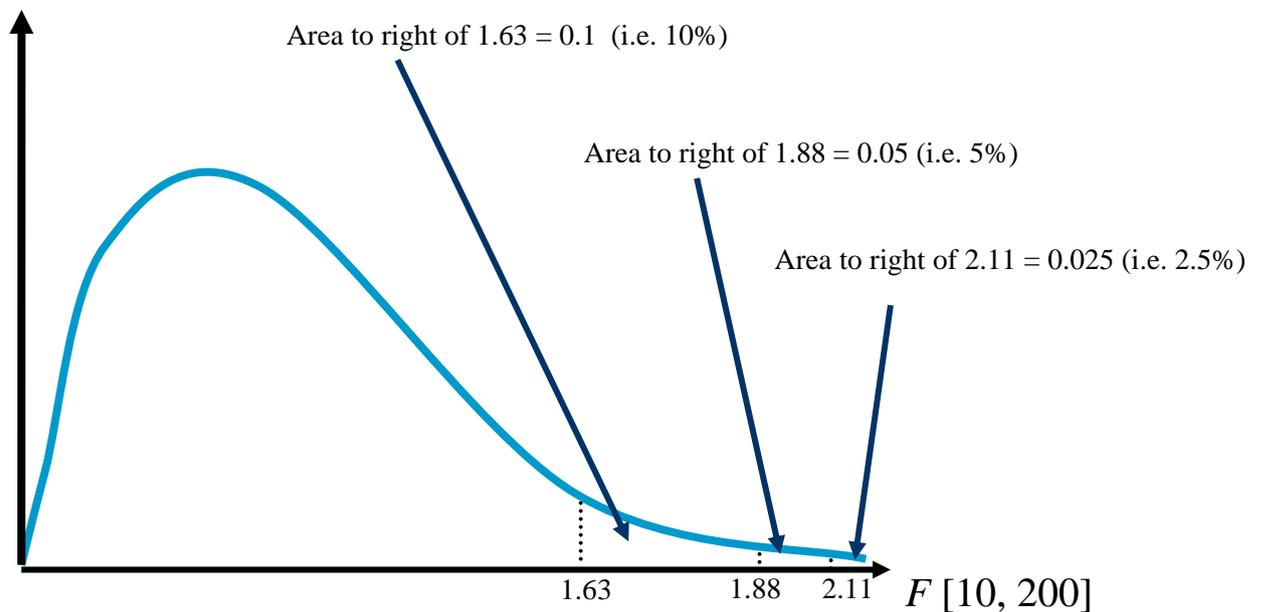
$$df_u = n - k \text{ from unrestricted regression,} = 212 - 12 = 200.$$

$$F [10, 200] = ((10000 - 8000) / 10) / (8000 / 200) \\ = 200 / 40 = \underline{\underline{5}}$$

From the F-table we can see that for 10 degrees of freedom in the numerator and 200 degrees of freedom in the denominator, the critical value of F associated with $p = \text{sig.} = 5\% = 0.050$ is $F=1.88$. In other words, if F is greater than 1.88, the area in the right-hand tail will be less than 0.05 and so the probability of false rejection of the null will be less than 5%.

	Numerator df
p	10
0.100	1.63
0.050	1.88
0.025	2.11
0.010	2.41
0.001	3.12

These values are depicted in the stylised F-distribution below:



So we can reject the null hypothesis if F is greater than 1.88. We have computed F to be 5 and so we can confidently reject the null hypothesis that the coefficients on the ten country dummies all equal zero.

2. Suppose you repeat the above exercise using a different dataset and estimate $RSS_U = 2,000$ and $RSS_R = 2,100$ (sample sizes and explanatory variables are the same as before). Conduct a hypothesis test of no country effects at the 1% significance level.

Unrestricted Regression:

$$INFL = a + b MS + g_1 \text{ COUNTRY1} + \dots + g_{10} \text{ COUNTRY10} \quad [1]$$

Restricted Regression:

$$INFL = a + b MS \quad [2]$$

$$H_0: g_1 = g_2 = g_3 = \dots = g_{10} = 0$$

(i.e. ten restrictions imposed in the restricted regression)

$$H_1: g_1, g_2, g_3, \dots, g_{10} \neq 0$$

$$F_{\frac{df_{\text{numerator}}}{df_{\text{denominator}}}}^r = F_{df_U}^r = \frac{(RSS_R - RSS_U) / r}{RSS_U / df_U}$$

Where:

$$RSS_U = \text{unrestricted residual sum of squares} = \text{RSS under } H_1 = 2,000$$

$$RSS_R = \text{restricted residual sum of squares} = \text{RSS under } H_0 = 2,100$$

$$r = \text{number of restrictions} = 10$$

$$df_u = n - k \text{ from unrestricted regression, } = 212 - 12 = 200.$$

$$F [10, 200] = ((2100 - 2000) / 10) / (2000 / 200) \\ = 10 / 10 = \underline{1}$$

At the 1% level of significance, F has to be above 2.41 to reject the null hypothesis. Since we have computed F to be 1, which is less than 2.41, we cannot reject the null hypothesis that the country effect coefficients all equal zero. I.e. the restriction appears to be valid (or at least we cannot say that it is invalid).

4.1.2 Exercise:

1. Open the `sovdebt.sav` data. Create dummies for the first **ten** countries.

We created dummies before by simply creating a new variable with all values equal to 0, then using the If statement to set the values equal to one if the property is satisfied. Note that if you run all the commands as a block, you only need to include “EXECUTE” after the last command:

```
COMPUTE COUNTRY1 = 0.  
IF(COUNTRY = 1) COUNTRY1 = 1.  
COMPUTE COUNTRY2 = 0.  
IF(COUNTRY = 2) COUNTRY2 = 1.  
COMPUTE COUNTRY3 = 0.  
IF(COUNTRY = 3) COUNTRY3 = 1.  
COMPUTE COUNTRY4 = 0.  
IF(COUNTRY = 4) COUNTRY4 = 1.  
COMPUTE COUNTRY5 = 0.  
IF(COUNTRY = 5) COUNTRY5 = 1.  
COMPUTE COUNTRY6 = 0.  
IF(COUNTRY = 6) COUNTRY6 = 1.  
COMPUTE COUNTRY7 = 0.  
IF(COUNTRY = 7) COUNTRY7 = 1.  
COMPUTE COUNTRY8 = 0.  
IF(COUNTRY = 8) COUNTRY8 = 1.  
COMPUTE COUNTRY9 = 0.  
IF(COUNTRY = 9) COUNTRY9 = 1.  
COMPUTE COUNTRY10 = 0.  
IF(COUNTRY = 10) COUNTRY10 = 1.  
EXECUTE.
```

The disadvantage with this approach is that it does not take account of missing values. If a missing value exists in the variable we are trying to recode, then the dummy variable will have the value zero. However, strictly speaking, this is incorrect, since it suggests that we know that the observation does not have the attribute in question (in this case, the country is not Argentina), when in reality we do not actually know what the country is. An additional problem is that the above procedure did not encourage us to include a variable label. This means that if we do not go back and label the newly created dummy variable we can end up not knowing what it means or how it was calculated (remember that a variable *name* can only have eight characters and it can sometimes be difficult to describe a variable with so few letters).

A better approach would be to use the RECODE syntax or to use the Transform/Recode/Into Different Variables... window. The advantage of using the syntax rather than the windows procedure is that if you are creating lots of dummies you can copy and paste the syntax and only alter the parts of the syntax that need to change. It also allows you to keep a record of your work.

The recode syntax for creating the first three country dummies is as follows (the shaded syntax indicates the what needs to be amended when you create a new country dummy variable) :

```
RECODE
  country
  (1=1) (MISSING=SYSMIS) (ELSE=0) INTO Country1 .
VARIABLE LABELS Country1 'Dummy for Argentina'.
EXECUTE .
```

```
RECODE
  country
  (2=1) (MISSING=SYSMIS) (ELSE=0) INTO Country2 .
VARIABLE LABELS Country2 'Dummy for Bolivia'.
EXECUTE .
```

```
RECODE
  country
  (3=1) (MISSING=SYSMIS) (ELSE=0) INTO Country3 .
VARIABLE LABELS Country3 'Dummy for Brazil'.
EXECUTE .
```

You can copy and paste these into the SPSS syntax window and run as normal.

2. Test whether there are country effects for Argentina, Bolivia and Brazil in a regression of money supply on inflation:

$$INFL = a + b MS + g_1 COUNTRY1 + g_2 COUNTRY2 + g_3 COUNTRY3$$

- (i) Run the *Unrestricted* regression and copy the ANOVA table from this regression. Place the cursor in cell E17 of the spreadsheet (the blue shaded cell) and paste as text (in Excel, select Edit, Paste Special, Text). This will displace any existing entry with your new ANOVA *Unrestricted* regression results.

```
DATASET ACTIVATE DataSet1.
REGRESSION
  /DEPENDENT infl
  /METHOD=ENTER ms COUNTRY1 COUNTRY2 COUNTRY3 COUNTRY4 COUNTRY5 COUNTRY6
  COUNTRY7 COUNTRY8 COUNTRY9 COUNTRY10.
```

Unrestricted Model

Model	Sum of Squares	ANOVA(b) df
1 Regression	302.068	11
Residual	1830.514	504
Total	2132.583	515

- (ii) Now run the *Restricted* regression and copy the ANOVA table from this regression. Place the cursor in cell E6 of the spreadsheet (the green shaded cell) and paste as text (in Excel, select Edit, Paste Special, Text). This will again displace any existing entry with your new ANOVA *Restricted* regression results.

```
REGRESSION
  /DEPENDENT infl
  /METHOD=ENTER ms.
```

Restricted Model

Model	Sum of Squares	ANOVA(b) df	Mean Square	F	Sig.
1 Regression	34.86	1	34.9	8.542	.004a
Residual	2097.722	514	4.08		
Total	2132.583	515			

- (iii) Excel should automatically calculate the r , df_U , F and significance level (see cells N25 to N35) but you should check the answer manually the first few times to ensure that you are using the spreadsheet correctly and that you understand the procedure. Remember that the null hypothesis is that there is no difference between the restricted and unrestricted regressions. Therefore, if F is large and the significance level small, you can reject the null with only a small probability of a Type I Error (i.e. that you have rejected H_0 incorrectly).

F Test Statistic: $F_{df_U}^r = \frac{(RSS_R - RSS_U) / r}{RSS_U / df_U}$	RSS_R	=	2097.722	
	RSS_U	=	1830.514	
	r	= $k_U - k_R$	=	10
	k_U	=	12	
	n	=	516	
	df_u	= $n - k_U$	=	504
			=	504
	F	=	7.357105	
	Sig F	=	6.3E-11	

F = 7.357, with sig. = 0.000, so we can reject the null hypothesis of no country effects (i.e. we should keep the country dummies in the regression).

3. (Optional) In a regression explaining imports per capita (mp_pc), test the hypothesis that the coefficients on GDP per capita (gdp_pc) and exports per capita (xp_pc) sum to one. That is, for:

$$mp_pc = b_1 + b_2 \text{ gdp_pc} + b_3 \text{ xp_pc}$$

test the null hypothesis that $b_2 + b_3 = 1$

Hint: first you need to create the new variables:

```
compute xp_pc = xp/pop.
VARIABLE LABELS xp_pc 'exports per capita'.
execute.

compute gdp_pc = gdp/pop.
VARIABLE LABELS gdp_pc 'GDP per capita'.
execute.

compute mp_pc = mp/pop.
VARIABLE LABELS mp_pc 'Imports per capita'.
execute.
```

Then run the unrestricted regression ($mp_pc = b_1 + b_2 \text{ gdp_pc} + b_3 \text{ xp_pc}$).

Then you need to derive the restricted regression. Before you know what regression to run, you need to work through the algebraic implications of the null hypothesis:

- substitute the restrictions into the equation (i.e. substitute $b_2 + b_3 = 1$ into the original unrestricted equation: $mp_pc = b_1 + b_2 \text{ gdp_pc} + b_3 \text{ xp_pc}$).
- rearrange the equation so that each parameter appears only once
- create new variables where necessary and estimate by OLS

Once you've done this, and you know what the restricted regression looks like, run the restricted regression.

Then paste the ANOVA tables from the restricted and unrestricted regressions into the Excel template. If the significance level is small, you can reject the

null that $b_2 + b_3 = 1$. You should check your results by calculating the F -ratio manually.

4. (Optional) Test whether, in the previous question, the two slope coefficients are equal.