Joint hypotheses

The null and alternative hypotheses can usually be interpreted as a restricted model (H_0) and an unrestricted model (H_A).

In our example:

$$H_0: score_i = \beta_0 + \beta_3 eng_i + u_i$$

$$H_A: score_i = \beta_0 + \beta_1 strat_i + \beta_2 exppup_i + \beta_3 eng_i + u_i$$

Note that if the unrestricted model "fits" significantly better than the restricted model, we should reject the null.

The difference in "fit" between the model under the null and the model under the alternative leads us to an intuitive formulation of the F-test statistic, for testing joint hypotheses.

Recall that a measure of "fit" is the sum of squared residuals:

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2$$

where

$$\hat{u}_i = Y_i - \hat{Y}_i$$

The F-test statistic may be written as:

$$F = \frac{(SSR_{restricted} - SSR_{unrestricted})/q}{SSR_{unrestricted}/(n - k_{unrestricted} - 1)}$$

where q = number of restrictions.

Notice that if the restrictions are true (if the null is true), $SSR_{restricted} - SSR_{unrestricted}$ will be small, and we'll fail to reject.

Another statistic which uses SSR is the R^2:

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

This gives us yet another formula for the F-test statistic.

$$F = \frac{(R^2_{unrestricted} - R^2_{restricted})/q}{(1 - R^2_{unrestricted})/(n - k_{unrestricted} - 1)}$$

where:

$R^2_{restricted}$ = the R^2 for the restricted regression
$R^2_{unrestricted}$ = the R^2 for the unrestricted regression
q = the number of restrictions under the null
$k_{unrestricted}$ = the number of regressors in the unrestricted regression.

The bigger the difference between the restricted and unrestricted R^2's – the greater the improvement in fit by adding the variables in question – the larger is the F statistic.

Note: the textbook differentiates between homoskedasticity only and heteroskedasticity robust F-tests. We will ignore heteroskedasticity for simplicity.

Example: are the coefficients on *strat* and *exppup* zero?

<u>Unrestricted population regression (under H_A):</u>

$$score_i = \beta_0 + \beta_1 strat_i + \beta_2 exppup_i + \beta_3 eng_i + u_i$$

<u>Restricted population regression (that is, under H_0):</u>

$$TestScore_i = \beta_0 + \beta_3 eng_i + u_i \qquad (why?)$$

- The number of restrictions under H_0 is $q = 2$ (*why*?).
- The fit will be better (R^2 will be higher) in the unrestricted regression (*why*?)

By how much must the R^2 increase for the coefficients on *strat* and *exppup* to be judged statistically significant?

<u>Restricted regression</u>:

$$score = 644.7 - 0.671eng, \quad R^2_{restricted} = 0.4149$$

$$(1.0) \quad (0.032)$$

<u>Unrestricted regression</u>:

$$score = 649.6 - 0.29strat + 3.87exppup - 0.656eng$$

$$(15.5) \quad (0.48) \quad (1.59) \quad (0.032)$$

$$R^2_{unrestricted} = 0.4366, \; k_{unrestricted} = 3, \; q = 2$$

so

$$F = \frac{(R^2_{unrestricted} - R^2_{restricted})/q}{(1 - R^2_{unrestricted})/(n - k_{unrestricted} - 1)}$$

$$= \frac{(.4366 - .4149)/2}{(1 - .4366)/(420 - 3 - 1)} = \mathbf{8.01}$$

A commonly performed F-test is one which assesses whether the chosen model fits *at all*.

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

H_A: any one of the βs not equal to zero

Why shouldn't the intercept be restricted?

This test will be performed by most regression software, and reported as "F-test" in the regression output – usually along with a p-value.

R Code

```
teachdata =
read.csv("http://home.cc.umanitoba.ca/~godwinrt/3180/data/str.csv")

attach(teachdata)
```

```
restricted = lm(score ~ eng)

unrestricted = lm(score ~ strat + exppup + eng)
```

Let’s look at the full (unrestricted) model:

```
summary(unrestricted)
```

Output:

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 649.581422  15.206583  42.717  < 2e-16 ***
strat        -0.286117   0.480548  -0.595  0.55190
exppup        0.003867   0.001412   2.738  0.00644 **
eng          -0.655976   0.039113 -16.771  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.35 on 416 degrees of freedom
Multiple R-squared: 0.4365,     Adjusted R-squared: 0.4324
F-statistic: 107.4 on 3 and 416 DF,  p-value: < 2.2e-16
```

Now, to perform the *F*-test of whether school spending matters or not:

```
anova(unrestricted, restricted)
```

Output:

```
Analysis of Variance Table


Model 1: score ~ strat + exppup + eng
Model 2: score ~ eng
  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1    416 85716
2    418 89014 -2   -3298.2 8.0034 0.0003885 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The $F_{q,n-k-1}$ distribution:

- The F distribution is tabulated many places
- As $n \to \infty$, the $F_{q,n-k-1}$ distribution asymptotes to the χ^2_q/q distribution:

The $F_{q,\infty}$ and χ^2_q/q distributions are the same.

- For q not too big and $n \geq 100$, the $F_{q,n-k-1}$ distribution and the χ^2_q/q distribution are essentially identical.
- Many regression packages (including STATA) compute p-values of F-statistics using the F distribution

You will encounter the F distribution in published empirical work.

Summary

$$F = \frac{(R^2_{unrestricted} - R^2_{restricted})/q}{(1 - R^2_{unrestricted})/(n - k_{unrestricted} - 1)}$$

- The homoskedasticity-only F-statistic rejects when adding the two variables increased the R^2 by "enough" – that is, when adding the two variables improves the fit of the regression by "enough"
- If the errors are homoskedastic, then the homoskedasticity-only F-statistic has a large-sample distribution that is χ^2_q/q.
- But if the errors are heteroskedastic, the large-sample distribution is a mess and is not χ^2_q/q

- The "one at a time" approach of rejecting if either of the *t*-statistics exceeds 1.96 rejects more than 5% of the time under the null (the size exceeds the desired significance level)
- For *n* large, the *F*-statistic is distributed χ_q^2/q (= $F_{q,\infty}$)
- The homoskedasticity-only *F*-statistic is important historically (and thus in practice), and can help intuition, but isn't valid when there is heteroskedasticity