1 Joint Significance

1.1 Subset of Regressors

Suppose we regress y on three different regressors, w, x and z, and both slope coefficients for x and z have high enough p-values that we conclude each one is statistically insignificant. It is still possible, however, that they may be *jointly* significant, even if they are individually insignificant. In other words, we want to test simultaneously that

$$H_0 : \beta_x = \beta_z = 0,$$

$$H_1 : \text{at least one of } \beta_x, \beta_z \neq 0.$$

Think of H_0 as being a *restriction* placed on β_x and β_z that we want to test. If the restriction leads to goofy results, then we conclude that the restriction is a bad one and we instead accept the alternative, namely, that β_x and β_z are jointly significant.

The test proceeds as follows. The first thing to do is take the model where β_x and β_z are unrestricted (that is, a regression where x and z are included and thus their coefficients are estimated) and find its sum of squared residuals, call it RSS_u. Then make the restrictions (by not even including them in the regression, which implicitly sets them equal to zero) and find that model's sum of squared residuals, call it RSS_r.

If β_x and β_z are jointly insignificant, i.e. if H_0 is true, then you would expect the difference between the two RSS terms to be small since the RSS represents unexplained variation in y. In other words, if β_x and β_z are jointly insignificant, then we shouldn't expect much difference in how well the model explain things (or how badly the model explains things, since we're using RSS) whether they're both simultaneously included or not. In other-other words, we're testing whether or not the regression is significantly (in the statistical sense) less bad when we include regressors x and z.

We just need to formalize what we mean by a "small" difference between the two. This is given by the F-statistic,

$$F \equiv \frac{(\text{RSS}_\text{r} - \text{RSS}_\text{u})/q}{\text{RSS}_\text{u}/(n-k)}, \tag{1}$$

which is drawn from the $F_{q,n-k}$ distribution, where

- n is the number of observations;
- k is the number of parameters being estimated in the unrestricted model, in this case $k = 4$ because we estimate the intercept plus slope coefficients for w, x, and z;

- $q = 2$ is the number of parameters being tested;
- $F_{q,n-k}$ is the F-distribution with $k - g$ parameters included in the restriction and $n - k$ is the unrestricted degrees of freedom.

1.2 Overall Significance

We've already seen this one, so I'll be brief. At the extreme end, we can test whether *all* regressors are jointly significant by comparing it to a regression with *no* regressors. There are k estimates, $k - 1$ of them are from regressors, therefore we are making $q = k - 1$ restrictions. Since we are imposing that all regressor coefficients are zero, it means the restricted model has zero explanatory power and therefore $\text{RSS}_\text{r} = \text{TSS}$. Using these two pieces of information, we can write the F-statistic as

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

where R^2 is from the unrestricted regression.

1.3 Individual Significance

We've also seen this ons already, so I'll be brief again. At the other extreme end, we can use the F-test to test just one restriction, for example $H_0 : \beta_x = 0$ against $H_1 : \beta_x \neq 0$. This looks like a simple, ordinary hypothesis test, and indeed, the F-statistic in this case will be the t-statistic squared.

1.4 Inference

The unrestricted model can never explain less than the restricted model, and it follows that $\text{RSS}_\text{u} \leq \text{RSS}_\text{r}$. This means that the F-statistic can never be negative; and furthermore it means that we only reject the null hypothesis when the F-statistic is too large in the positive direction. Therefore inference only looks at the right-tail, so there is no need to chop the significance in half when finding critical values or multiply by 2 when finding the p-value. Also note that the formulas for F as given ***are only valid with homoskedasticity!!!!111!*** Could be a multiple choice question. Typically we have to use a different version of the F-statistic to account for heteroskedasticity which is difficult to compute; we'll rely on R to find it.

2 Examples

2.1 Subset of Regressors

Consider the `sleep.csv` data from my website. The data has 706 observations and five variables including hours slept per week and hours worked per week, along with age and dummy variables for good health and for sex. We're going to explain how much sleep a person gets with how much they work, their age, their health, and their sex with the following regression:

$$sleep = \beta_1 + \beta_2 age + \beta_3 hoursworked + \beta_4 goodhealth + \beta_5 male + \epsilon.$$

The R code and regression results are shown in Figure 1. Notice that age and health are not significant at 5 percent. Because 5 percent is the de facto standard, we would therefore typically conclude that the effects of age and health are statistically indistinguishable from zero. That is, we cannot reject either $H_0 : \beta_2 = 0$ or $H_0 : \beta_4 = 0$ at 5 percent significance.

But even though we conclude age and health have zero explanatory power *in isolation*, it could still be the case that they have nonzero explanatory power *together*. In other words, they do not have individual significance, but they might have joint significance. The null hypothesis for joint significance is $H_0 : \beta_2 = \beta_4 = 0$ against the alternative of H_1 : at least one of $\beta_2, \beta_4 \neq 0$. If we can reject the null, then we conclude that age and health have joint explanatory power, even if they do not have explanatory power when considered in isolation.

We will need to run two regressions in order to test whether age and health are jointly significant or not. First we run the entire regression, which we've already done and have saved as `olsu`. This is the **unrestricted** regression because we are estimating coefficients for every single variable of interest. Of particular importance will be the residual sum of squares, which I will now calculate. Call this the unrestricted RSS, denoted RSS_u. R output indicates that $\text{RSS}_\text{u} = 34007$, approximately.

Next we run the **restricted** regression which assumes the null hypothesis is true. Because the null hypothesis says that age and health are irrelevant, this means running a regression that omits these regressors entirely, thereby implicitly setting $\beta_2 = \beta_4 = 0$ as per the null. We will take the residual sum of squares from this **auxiliary** regression and call is the restricted RSS, denoted RSS_r. R output indicates that $\text{RSS}_\text{r} = 34357$, approximately.

If the null hypothesis is true so that age and health don't explain anything—that is, if age and health are *irrelevant*—then the RSS should be the same whether they're included

or not. However, if the null hypothesis is false so that age and health actually do jointly explain something, then RSS should be *smaller* when we include them in the regression. We have $n = 706$ observations, $k = 5$ estimates in the unrestricted regression, and we're testing $q = 2$ restrictions. Accordingly the test statistic is given by

$$F \equiv \frac{(\text{RSS}_\text{r} - \text{RSS}_\text{u})/q}{\text{RSS}_\text{u}/(n-k)} = \frac{(34357 - 34007)/2}{34007/(706-5)} = 3.60.$$

The p-value for this F-statistic is 0.02777, so we reject the null in favor of the alternative. The conclusion is that age and health have joint explanatory power at 5 percent significance (even though none of them have individual explanatory power at 5 percent significance).

Note that we can also perform the test using the `anova()` function, shown in Figure 1.

2.2 Non-Zero Linear Hypothesis

Again consider the overall regression

$$sleep = \beta_1 + \beta_2 age + \beta_3 hoursworked + \beta_4 goodhealth + \beta_5 male + \epsilon.$$

Suppose we think that men on average sleep two hours more than women per week, but age is irrelevant. Our null hypothesis is then $H_0 : \beta_2 = 0$ and $\beta_5 = 2$ against alternative $H_1 : \beta_2 \neq 0$ or $\beta_5 \neq 2$.

```
> olsu = lm(sleep ~ age+hoursworked+goodhealth+male, data=sleep)
> stargazer(olsu, type = "text")

================================================
                        Dependent variable:
                    ---------------------------
                               sleep
------------------------------------------------
age                             0.041*
                                (0.023)

hoursworked                    -0.163***
                                (0.018)

goodhealth                      -1.553*
                                (0.850)

male                            1.439**
                                (0.572)

Constant                       59.170***
                                (1.376)
------------------------------------------------
Observations                      706
R2                                0.121
Adjusted R2                       0.116
Residual Std. Error       6.965 (df = 701)
F Statistic           24.069*** (df = 4; 701)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01

> RSSu = sum(olsu$residuals^2)
> RSSu
[1] 34007.12

> olsr = lm(sleep ~ hoursworked + male, data = sleep)
> RSSr = sum(olsr$residuals^2)
> RSSr
[1] 34356.62

> F = ((RSSr - RSSu)/2)/(RSSu/(706-5))
> F
[1] 3.602217

> pf(F, 2, 706-5, lower.tail = FALSE)
[1] 0.02776907

> anova(olsu, olsr)

Model 1: sleep ~ age + hoursworked + goodhealth + male
Model 2: sleep ~ hoursworked + male
  Res.Df   RSS Df Sum of Sq      F  Pr(>F)
1    701 34007
2    703 34357 -2    -349.5 3.6022 0.02777
```

FIGURE 1: We reject the joint irrelevance of age and health at 5 percent significance because the p-value is less than 0.05.

Imposing the null hypothesis leads to a restricted regression of form

$$sleep = \beta_1 + \beta_3 hoursworked + \beta_4 goodhealth + 2male + \epsilon,$$

or better yet,

$$sleep - 2male = \beta_1 + \beta_3 hoursworked + \beta_4 goodhealth + \epsilon.$$

The preceding regression is what we run to get the restricted RSS. As shown in Figure 2, doing so yields $\text{RSS}_\text{u} = 34198$, approximately. We are testing $q = 2$ restrictions, we have $n = 706$ observations and $k = 5$ estimates in the full regression. Ergo the test statistic is

$$F = \frac{(34198 - 34007)/2}{34007/(706 - 5)} \approx 1.97.$$

The p-value for this F-statistic is about 0.14, so we fail to reject the null and conclude that men sleep two hours more than women per week and age is irrelevant.

We can also test this using the `linearHypothesis()` function from the "car" package, as shown in R output.

```
> olsu = lm(sleep ~ age+hoursworked+goodhealth+male, data=sleep)
> stargazer(olsu, type = "text")

================================================
                        Dependent variable:
                    ---------------------------
                               sleep
------------------------------------------------
age                            0.041*
                               (0.023)

hoursworked                   -0.163***
                               (0.018)

goodhealth                     -1.553*
                               (0.850)

male                           1.439**
                               (0.572)

Constant                      59.170***
                               (1.376)
------------------------------------------------
Observations                     706
R2                              0.121
Adjusted R2                     0.116
Residual Std. Error       6.965 (df = 701)
F Statistic           24.069*** (df = 4; 701)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01

> RSSu = sum(olsu$residuals^2)
> RSSu
[1] 34007.12

> olsr =  lm(sleep - 2*male ~ hoursworked+goodhealth, data=sleep)
> RSSr = sum(olsr$residuals^2)
> RSSr
[1] 34198.23

> F = ((RSSr - RSSu)/2)/(RSSu/(706-5))
> F
[1] 1.969799

> pf(F, 2, 706-5, lower.tail = FALSE)
[1] 0.1402562

> linearHypothesis(olsu, c("age = 0","male = 2"))

Model 1: restricted model
Model 2: sleep ~ age + hoursworked + goodhealth + male

  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1    703 34198
2    701 34007  2    191.12 1.9698 0.1403
```

FIGURE 2: We fail to reject the claim that men on average sleep two hours more than women per week and age is irrelevant at 5 percent significance because the p-value is greater than 0.05.