In-Class Lab 8

ECON 4223 (Prof. Tyler Ransom, U of Oklahoma)

February 17, 2022

The purpose of this in-class lab is to practice conducting ***joint*** hypothesis tests of regression parameters in R. We will do this using t-tests and F-tests. The lab may be completed in your group, but each group member should submit their own copy. To get credit, upload your .R script to the appropriate place on Canvas.

For starters

Open up a new R script (named `ICL8_XYZ.R`, where `XYZ` are your initials) and add the usual "preamble" to the top:

```
# Add names of group members HERE
library(tidyverse)
library(broom)
library(wooldridge)
library(car)
library(magrittr)
library(modelsummary)
```

Load the data

We'll use a data set on earnings and ability, called `htv`. The data set contains a sample of 1,230 workers.

```
df <- as_tibble(htv)
```

Check out what's in the data by typing

```
datasummary_df(df)
datasummary_skim(df,histogram=FALSE)
```

The main variables we're interested in are: wages, education, ability, parental education, and region of residence (`ne`, `nc`, `west`, and `south`).

Create regional factor variable

Let's start by creating a factor variable from the four regional dummies. Borrowing code from lab 6, we have:

```
df %<>% mutate(region = case_when(ne==1 ~ "Northeast",
                                  nc==1 ~ "NorthCentral",
                                  west==1 ~ "West",
                                  south==1 ~ "South")) %>%
        mutate(region = factor(region))
```

Regression and Hypothesis Testing

Estimate the following regression model:

$$educ = \beta_0 + \beta_1 motheduc + \beta_2 fatheduc + \beta_3 abil + \beta_4 abil^2 + \beta_5 region + u$$

Note that *abil* is in standard deviation units. You will need to use a `mutate()` function to create $abil^2$ (not shown here). Call it `abilsq`. *region* represents the factor variable you created above.[1]

```
est <- lm(educ ~ motheduc + fatheduc + abil + abilsq + region, data=df)
tidy(est)
modelsummary(est)
```

t-test

1. Test the hypothesis that *abil* has a linear effect on *educ*.

F-test (single parameter)

2. Now test that *motheduc* and *fatheduc* have equal effects on *educ*. In other words, test $H_0 : \beta_1 = \beta_2; H_a : \beta_1 \neq \beta_2$. To do this, you will need to obtain $se(\beta_1 - \beta_2)$. Luckily, R will do this for you with the `linearHypothesis()` function in the `car` package:

```
linearHypothesis(est, "motheduc = fatheduc")
```

The resulting p-value is that of an F test, but one would get an identical result by using a t-test, since this is a simple hypothesis (see Wooldridge (2015), pp. 125-126).

F-test (multiple parameters)

The p-values from the previous regression might indicate that the three region dummies don't contribute to education.

3. Test the hypothesis that they don't; i.e. test

$$H_0 : \text{all region dummies} = 0; H_a : \text{any region dummy} \neq 0$$

The code to do this again comes from the `linearHypothesis()` function. The syntax is to encolose each component hypothesis in quotes and then surround them with `c()`, which is how R creates vectors.

```
linearHypothesis(est, c("regionNortheast=0", "regionSouth=0", "regionWest =0"))
```

or, more simply,

```
linearHypothesis(est, matchCoefs(est,"region"))
```

Alternatively, you can perform the F-test as follows (no need to put this in your R-script; I'm just showing you how to do it "by hand"):

```
est.restrict <- lm(educ ~ motheduc + fatheduc + abil + abilsq, data=df)
Fstat.numerator   <- (deviance(est.restrict)-deviance(est))/3
Fstat.denominator <- deviance(est)/1222
Fstat <- Fstat.numerator/Fstat.denominator
p.value <- 1-pf(Fstat,3,1222)
```

This gives the exact same answer as the `linearHypothesis()` code.

References

Wooldridge, Jeffrey M. 2015. *Introductory Econometrics: A Modern Approach*. 6th ed. Cengage Learning.

[1] Here the notation of $\beta_5 region$ is not quite right. It more technically should be written $\beta_5 region.NE + \beta_6 region.S + \beta_7 region.W$, where each of the $region.X$ variables is a dummy. The way it is written above, $\beta_5 region$ implies that β_5 is a vector, not a scalar.