

More Hypothesis Tests with Multiple Regressions

Pull in mosaic

In R:
`library(mosaic)`

Second: call up data set called Galton (built into R)

In R:
`data(Galton)`

Look at the first few rows to get a feel for the data:

In R:
`head(Galton)`

There are 898 observations of 6 variables:

- family # (which is a number assigned to each family interviewed and won't be used in our analysis)
- father (which is father's height)
- mother (which is mother's height)
- sex (which is the gender of the child)
- height (of the child)
- nkids (# of children in the family; also won't be used in our analysis)

Let's run our first model:

$$Height = \beta_0 + \beta_1 \text{Father} + \beta_2 \text{Mother} + \beta_3 \text{Sex} + e$$

In R:
`model1 = lm(height~father + mother + sex, data=Galton)`
`summary(model1)`

To get the RSS, we need to do the following

In R:
`rssmod1 = sum(residuals(model1)^2)`
`rssmod1`

Note that, in our data, males and females are listed as M and F, respectively. But, data need to be numerical for regressions to work. Therefore, R is smart enough to convert the data to zeros and ones. Notice in your results that “sex” becomes “sexM.” This means that R has set this variable to be equal to 1 if the person is a male, and zero if the person is a female.

1. We would expect that taller parents would have taller children. Test whether β_1 is positive (at the 5% level of significance). Do the same for β_2 . For each you must state your null and alternative hypotheses, your decision rule, and your final decision based on the test.
2. We would expect males to be taller than females. Suppose we believe that, all else equal, males are 5 inches taller than females. In other words, we believe that β_3 is equal to 5. Test this hypothesis at the 5% level. (Remember the formula for the calculated t-statistic is different in this case! You can't use the one provided by R.)
3. Test whether $\beta_1 = \beta_2 = \beta_3 = 0$ simultaneously at the 5% level of significance.

Suppose that we believe that each parent's height contributes equally to the child's height. This means that it would be okay to restrict the model to just an average measure of the parents' height.

We need to first create a variable that measures the combined height.

In R:

```
Galton$averageht = (Galton$father + Galton$mother)/2
```

Then, we need to create a second model with this variable in place of the parents' height.

$$Height = \beta_4 + \beta_5 Averageht + \beta_6 Sex + e$$

In R:

```
restrictedmodel1 = lm(height~averageht + sex, data=Galton)  
summary(restrictedmodel1)
```

Again, we need to get the RSS:

In R:

```
rssrestmod1 = sum(residuals(restrictedmodel1)^2)  
rssrestmod1
```

4. Test whether this restriction is valid. Note that your first model is the unrestricted model, and the second model is the restricted model. Also, the number of restrictions is one. (You are basically testing whether $\beta_1 = \beta_2$.)

Let's try something else. Let's estimate the original model separately for males and females.

In R:

```
modelfem = lm(height~father + mother, data=subset(Galton, sex == 'F'))  
summary(modelfem)  
rssfem = sum(residuals(modelfem)^2)  
rssfem
```

Note: The quote marks around the F are because this is a string (non-numeric) variable. If sex had been 1 for males and 0 for females, you would just use the numbers without the quote marks.

```
modelmale = lm(height~father + mother, data=subset(Galton, sex == 'M'))
summary(modelmale)
rssmale = sum(residuals(modelmale)^2)
rssmale
```

We also need to run a model with the full data, but excluding sex as a variable:

In R:

```
modelfull = lm(height~father + mother, data=Galton)
rssfull = sum(residuals(modelfull)^2)
rssfull
```

5. Let's test whether the coefficients on father and mother are the same for male and female children. In this case, modelfull is the restricted model (it restricts the coefficients to be the same for both genders of the children). Modelfem and modelmale are the unrestricted models. Note that the number of restrictions is 2, because that is the number of coefficients we are requiring to be equal across genders: the coefficient on father and the coefficient on mother.