

Download the data file “colldist.xls” from D2L. Import it into R. This data has six variables:

female = 1 if student is female; =0 otherwise

test = student’s score on standardized test

dadcoll = 1 if student’s father is a college graduate; =0 otherwise

momcoll = 1 if student’s mother is a college graduate; =0 otherwise

dist = distance from 4-year college in tens of miles to student’s home

ed = years of education completed by the student

Use R to answer the following questions. You will need to turn in a printout of your work in R and answers for each of the questions listed above. For all hypothesis tests, lists the null and alternative hypotheses, the decision rule, the calculated test statistic, the critical value of the test statistic, and the decision.

1. Run a regression with ed as the dependent variable and the other five variables as the independent variables. Write out the estimated regression equation.
2. Provide an interpretation of the coefficient on dadcoll.
3. Provide an interpretation of the coefficient on dist.
4. Test the significance of the estimated coefficient on dist. (Use a 5% significance level.)

5. Separate the data file by gender. Estimate three equations (full data, males, females) with `ed` as the dependent variable and `dist`, `test`, `dadcoll`, and `momcoll` as the independent variables. Test whether the model should be estimated separately by gender or pooled together (at the 5% level).

6. Perhaps it only matters if both of your parents go to college. Therefore, we are going to create a variable (`bothcoll`) that is equal to 1 if both parents went to college, and equal to 0 otherwise. To do this, follow these steps in R:

In R:

```
colldist$bothcoll = colldist$momcoll*colldist$dadcoll
```

Now, you can run a regression with `bothcoll` as an independent variable with `dist`, `test`, `female`. (`Ed` is still the dependent variable.) This model will be the restricted model. The very first model you ran (in problem 1) is the unrestricted model. Test whether this restriction is valid at the 5% significance level.