HW6 Group 1, Austin Halvorsen

Pedram Jahangiry

Oct 22 2020

# Problems

## Question 1

### (i)

### (ii)

## Question 2

We can know that where is just the regression of . We know that and we can assume that the .

## Question 3

### (i)

df1 = 351 df2 = 352

The SER in the second is smaller because the number of explantory variables increased.

### (ii)

### (iii)

The SER

## Question 4

### (i)

No, mathematically it does not make sense. The problem states that the sum of all activities must be 168, so if we change one, then at least one other category will need to change so that the sum remains 168.

### (ii)

Because we can write one variable from this function as a perfect linear function of the other independent variables, as well as for any of the other variables, this violates MLR.3.

### (iii)

I would drop one of the independent variables, like *leisure* and replace it with , thus creating the new model:

Now, this will not violate assumption MLR.3 because if we hold sleep and work fixed, but increase study by an hour, then leisure must be falling by an hour. This would apply to the other parameters interpretations as well.

# Computer Exercises

## Question 5

### (i)

d1 <- wooldridge::meapsingle  
  
srm1 <- lm(math4~pctsgle, data = d1)  
stargazer(srm1, type='text')

===============================================  
 Dependent variable:   
 ---------------------------  
 math4   
-----------------------------------------------  
pctsgle -0.833\*\*\*   
 (0.071)   
   
Constant 96.770\*\*\*   
 (1.597)   
   
-----------------------------------------------  
Observations 229   
R2 0.380   
Adjusted R2 0.377   
Residual Std. Error 12.480 (df = 227)   
F Statistic 138.853\*\*\* (df = 1; 227)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

The model from our output:

Our intercept is

### (ii)

srm2 <- lm(math4~pctsgle+lmedinc+free, data = d1)  
stargazer(srm2, type='text')

===============================================  
 Dependent variable:   
 ---------------------------  
 math4   
-----------------------------------------------  
pctsgle -0.200   
 (0.159)   
   
lmedinc 3.560   
 (5.042)   
   
free -0.396\*\*\*   
 (0.070)   
   
Constant 51.723   
 (58.478)   
   
-----------------------------------------------  
Observations 229   
R2 0.460   
Adjusted R2 0.453   
Residual Std. Error 11.696 (df = 225)   
F Statistic 63.848\*\*\* (df = 3; 225)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Adding in the variables and drops the slope coefficient of to .

### (iii)

cor(d1$lmedinc, d1$free)

[1] -0.7469703

### (iv)

### (v)

## Question 6

### (i)

d2 <- wooldridge::htv  
  
educ <- range(d2$educ)  
print(educ[2]-educ[1])

[1] 14

The Range of the educ variable is **14**

sum(d2$educ<13)/sum(d2$educ>0)

[1] 0.5674797

The proportion of men who have completed 12 grade, and no higher, is **56.7%**

## Average Mens education  
mean(d2$educ)

[1] 13.0374

## Average Parents education   
(mean(d2$motheduc) + mean(d2$fatheduc))/2

[1] 12.3126

The average men had a mean education of **13.037** while the parents had an average of **12.313**. So the parents had less, on average, education.

### (ii)

mrm1 <- lm(educ~motheduc+fatheduc, d2)  
stargazer(mrm1, type='text')

===============================================  
 Dependent variable:   
 ---------------------------  
 educ   
-----------------------------------------------  
motheduc 0.304\*\*\*   
 (0.032)   
   
fatheduc 0.190\*\*\*   
 (0.022)   
   
Constant 6.964\*\*\*   
 (0.320)   
   
-----------------------------------------------  
Observations 1,230   
R2 0.249   
Adjusted R2 0.248   
Residual Std. Error 2.042 (df = 1227)   
F Statistic 203.684\*\*\* (df = 2; 1227)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

The resulting model is:

With an and . This means that 24.8% of the variation in education is explained by parents education. We can interpret our coefficients to mean that for every increase in a mothers education, the son will increase by 0.304 years. And simililary with the fathers education.

### (iii)

mrm2 <- lm(educ~motheduc+fatheduc+abil, d2)  
stargazer(mrm1, mrm2, type='text')

=======================================================================  
 Dependent variable:   
 ---------------------------------------------------  
 educ   
 (1) (2)   
-----------------------------------------------------------------------  
motheduc 0.304\*\*\* 0.189\*\*\*   
 (0.032) (0.029)   
   
fatheduc 0.190\*\*\* 0.111\*\*\*   
 (0.022) (0.020)   
   
abil 0.502\*\*\*   
 (0.026)   
   
Constant 6.964\*\*\* 8.449\*\*\*   
 (0.320) (0.290)   
   
-----------------------------------------------------------------------  
Observations 1,230 1,230   
R2 0.249 0.428   
Adjusted R2 0.248 0.426   
Residual Std. Error 2.042 (df = 1227) 1.784 (df = 1226)   
F Statistic 203.684\*\*\* (df = 2; 1227) 305.172\*\*\* (df = 3; 1226)  
=======================================================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Our resulting model is now:

With an and

By adding the variable we were able to explain 42.8% of the variation of . Controlling for parent’s education, we can see the increase from our first model to this model.

### (iv)

sqabil <- d2$abil^2  
mrm3 <- lm(educ~motheduc+fatheduc+abil+sqabil, d2)  
stargazer(mrm2, mrm3, type='text')

=======================================================================  
 Dependent variable:   
 ---------------------------------------------------  
 educ   
 (1) (2)   
-----------------------------------------------------------------------  
motheduc 0.189\*\*\* 0.190\*\*\*   
 (0.029) (0.028)   
   
fatheduc 0.111\*\*\* 0.109\*\*\*   
 (0.020) (0.020)   
   
abil 0.502\*\*\* 0.401\*\*\*   
 (0.026) (0.030)   
   
sqabil 0.051\*\*\*   
 (0.008)   
   
Constant 8.449\*\*\* 8.240\*\*\*   
 (0.290) (0.287)   
   
-----------------------------------------------------------------------  
Observations 1,230 1,230   
R2 0.428 0.444   
Adjusted R2 0.426 0.443   
Residual Std. Error 1.784 (df = 1226) 1.758 (df = 1225)   
F Statistic 305.172\*\*\* (df = 3; 1226) 244.906\*\*\* (df = 4; 1225)  
=======================================================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

mrm3\_df <- as.data.frame(mrm3[1])  
abil\_coef <- mrm3\_df[4,]  
abil\_coef\_sq <- mrm3\_df[5,]  
  
min\_abil <- -abil\_coef / (abil\_coef\_sq\*2)  
print(min\_abil)

[1] -3.967098

Our model would predict education at a minimized ability level of **-3.9671.**

### (v)

table(d2$abil <= min\_abil)

FALSE TRUE   
 1215 15

Based on our data, only 15 men would have an “ability” less than the predicted calculated level. This is important because this tells us that there are 15 men that have an ability less than -3.967, yet have an education that is .

### (vi)

estimate <- mrm3\_df[1,] + (mrm3\_df[2,]\*mean(d2$motheduc)) + (mrm3\_df[3,]\*mean(d2$fatheduc)) + (mrm3\_df[4,]\*d2$abil) + mrm3\_df[5,]\*sqabil  
   
d2 %>%   
 ggplot(aes(x=abil,y=estimate))+  
 geom\_point(size=1, color="green")+  
 ylab("Predicted Education")+  
 xlab ("Ability")+  
 theme\_linedraw()

