HW6 Patrick Neyland

Patrick Neyland

2022-10-11

## Question 1

### Part i

If is unbiased, then the expected value of theta hat is equal to theta.

### Part ii

going from cov to corr Define unbiasedness

## Question 2

pg. 79 Wooldridge 6e The bias will likely be negative because I think avg ability() will be positive and, as stated in the problem, correlation between variables is negative.

## Question 3

### Part i

Model 1 has df of 351 Model 2 has df of 350 Because SER is controlling for RBIs. less variance in sigma squared

### Part ii

Yeah, it makes sense. Having a little more experience can help someone have more RBIs, However

Variance inflation factor

I would say there is moderate collinearity between and .

1/(1-0.237)

[1] 1.310616

0.487^2

[1] 0.237169

model3\_3 <- lm(lsalary ~ years + rbisyr, data = mlb1)  
vif(model3\_3)

years rbisyr   
1.311086 1.311086

### Part iii

Because the value of in model 2 is much smaller in model 1. This is a rare occurrence and will decrease in likelihood as more variables are added.

## Question 4

### Part i

Yes, I think holding sleep, work, and leisure fixed while changing study makes sense.

### Part ii

Is it because they all sum to 168. They are all part of the same variable—time. This leads to perfect collinearity among the independent variables.

### Part iii

Dropping one of the variables would satisfy MLR.3

## Question 5

### Part i

model5\_1 <- lm(math4 ~ pctsgle, data = meapsingle)  
stargazer(model5\_1, 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 effect of single parenthood seems large.

### Part ii

model5\_2 <- lm(math4 ~ pctsgle +lmedinc + free, data = meapsingle)  
stargazer(model5\_2, 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

It decreases significantly because other important variables are being controlled for. It seems that free lunch is a much better indicator of math scores.

### Part iii

cor(meapsingle$lmedinc,meapsingle$free)

[1] -0.7469703

The sample correlation between and is -0.747 Yes, it has the sign I expected. The higher a family’s income, the less likely they will be to receive free lunch at school.

### Part iv

I don’t think the correlation between and is enough to justify dropping one of them on its own. The variable should also be considered. If we remove it without deeper analysis, we may be underspecifying the model and introducing unnecessary bias.

### Part v

Watch class from 10/13 and look at the lecture R scripts

vif(model5\_2)

pctsgle lmedinc free   
5.740981 4.118812 3.188079

has the largest VIF, which is 5.741. No, it does not change which model I would use.

## Question 6

### Part i

x <- range(htv$educ)  
diff(x)

[1] 14

tb1 <- table(htv$educ==12)  
mean(htv$educ==12)

[1] 0.4162602

ptb1 <- prop.table(tb1)\*100  
ptb1[2]

TRUE   
41.62602

avg\_educ <- round(mean(htv$educ),2)  
avg\_peduc <- round(mean(htv$motheduc + htv$fatheduc)/2,2)

The range of the variable in the sample is 14.  
41.63 percent of men completed the twelfth grade but no higher.  
Average is 13.04 and the average parent education is 12.31—on average, men have higher levels of education than their parents.

### Part ii

model6\_2 <- lm(educ ~ motheduc + fatheduc, data = htv)  
stargazer(model6\_2, 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

24.9 percent of the sample variation in is explained by parents’ education. Ceteris paribus, if increases by 1 year, will increase by 0.304 years on average.  
### Part iii

model6\_3 <- lm(educ ~ motheduc + fatheduc + abil, data = htv)  
stargazer(model6\_3, type = "text")

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

Yes, “ability” helps explain variations in education because it is relevant and not perfectly correlated with parents education. This is evidenced by the improve in part iii over the in part ii.

### Part iv

abil\_squared <- htv$abil^2  
model6\_4 <- lm(educ ~ motheduc + fatheduc + abil + abil\_squared, data = htv)  
stargazer(model6\_4, type = "text")

===============================================  
 Dependent variable:   
 ---------------------------  
 educ   
-----------------------------------------------  
motheduc 0.190\*\*\*   
 (0.028)   
   
fatheduc 0.109\*\*\*   
 (0.020)   
   
abil 0.401\*\*\*   
 (0.030)   
   
abil\_squared 0.051\*\*\*   
 (0.008)   
   
Constant 8.240\*\*\*   
 (0.287)   
   
-----------------------------------------------  
Observations 1,230   
R2 0.444   
Adjusted R2 0.443   
Residual Std. Error 1.758 (df = 1225)   
F Statistic 244.906\*\*\* (df = 4; 1225)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Leaving out other independent variables because they have no effect.

Set equal to zero.

is equal to -3.93. is minimized when is -3.93.

Because the second derivative is positive, the critical value (-3.93) is a minimum.

### Part v

-0.401/0.102

[1] -3.931373

tb2 <- table(htv$abil<(-0.401/0.102))  
ptb2 <- prop.table(tb2)\*100  
ptb2[2]

TRUE   
1.219512

From the data, only 1.22 percent of men have an less than -3.931.

### Part vi

peduc\_tot <- 12.18\*.190 + .109\*12.45  
curve( 8.240 + peduc\_tot + 0.401\*x + 0.051\*x^2, from=-4, to=7, n=500,   
 xlab="Ability", ylab="Years of Education",   
 col="blue", lwd=2,   
 main="Predicted Education based on Ability" )

