HW5-Patrick Neyland

Patrick Neyland

2022-10-04

## Question 1

### Part i

The sign will be negative—as an adult works more, they will get less sleep.

### Part ii

I think will be positive, however, I do not think it will be statistically significant. I think will be negative—the older someone gets, the less sleep they need.

### Part iii

The number of minutes is predicted to fall by 44.4 minutes. This doesn’t seem to be too big of a tradeoff. Five additional hours while only losing an average of 6.3428571 minutes of sleep each night seems worth it—if you are only considering these two factors in the discussion of total utility of ones time.

### Part iv

Getting one more year of education, on average, holding all else constant, will result in an 11.13 minute decrease in the amount of sleep each week. 11.13 minutes and a negative sign were both unexpected to me.

### Part v

Definitely not. With an of just 0.113, only 11.3 percent of the variation of sleep can be explained by these variables. The number of dependents, base salary, and level of community involvement would be other factors that could affect the amount of sleep someone is getting. I think each of these other factors would be correlated with .

## Question 2

### Part i

Because the order of rank is in reverse numerical order. The higher the number of the rank for a particular school, the lower its rank. It’s a lot like golf, the lower yours score the better.

### Part ii

I expect to be positive. A high LSAT score is supposed to indicate aptitude in the study of law. Therefore, someone with a high LSAT score would presumably provide more value than someone with a low score.  
I expect to be positive. For the same reasons as my prediction for but with a more general interpretation. GPA may indicate aptitude in the study of any subject.  
I expect to be positive. Having more volumes in a library meas that the students have more resources to learn. However, I don’t think this one will have a significant impact on the model.  
I expect to be positive. Cost of attendance is usually correlated with prestige and sometimes with quality of education. If a school is both prestigious and offers a quality education, I would expect its graduates to be offered high salaries.

### Part iii

If the median increases by one point, on average, the ceteris paribus difference in median salary will be an increase of 24.8 percent.

### Part iv

As the number of volumes in the library increases by 1 percent, on average, holding all else constant, the median starting salary will increase by 0.095 percent.

### Part v

I think ranking does matter to some extent. In this model, a 20 ranking difference would result in a 6.6 percent change in predicted starting salary. The difference does seem significant. I think it is best to attend a school that fits your personal goals and skills. The school where you can be the best student you can be will yield the best salary for you.

## Question 3

The sample size is smaller so the variance of the model will be larger, leading to a smaller .

FACT: As k increases, is going to increase as long as correlation between and is not equal to zero.

However, models with different sample sizes cannot be compared against eachother. These two models are working with different data, the model with

### Question 4

model4\_1 <- lm(price ~ sqrft + bdrms, data = hprice1)  
stargazer(model4\_1, type = "text")

===============================================  
 Dependent variable:   
 ---------------------------  
 price   
-----------------------------------------------  
sqrft 0.128\*\*\*   
 (0.014)   
   
bdrms 15.198   
 (9.484)   
   
Constant -19.315   
 (31.047)   
   
-----------------------------------------------  
Observations 88   
R2 0.632   
Adjusted R2 0.623   
Residual Std. Error 63.045 (df = 85)   
F Statistic 72.964\*\*\* (df = 2; 85)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

### Part ii

The estimated increase in price will be $15,198.

### Part iii

33118

The new increase with the sqrft specified for the new bedroom is $33,118 The result from this specification is much larger than in part ii because in part 2, it is assumed that the of the house remains the same despite the additional bedroom.

### Part iv

63.2 percent

### Part v

new\_df <- data.frame(sqrft = c(2438), bdrms = c(4))  
predict(model4\_1, new\_df)

1   
354.6052

354.6052\*1000

[1] 354605.2

The predicted selling price for the OLS model with the given parameters is $354,605.20

### Part vi

(300-354.6052)\*1000

[1] -54605.2

The residual is -54,605.2 dollars. The model suggests that the buyer underpaid $54,605.20 for the house.

## Question 5

### Part i

[1] 0.1134864

[1] 47053.78

[1] 0.1824165

[1] 13179.29

The average is 0.11.  
The average is 47053.78.  
The standard deviation of is 0.18.  
The standard deviation of is 13179.29.

The measure of units for is proportion on a scale of 0 to 1. The measure of units for is median income in dollars.

### Part ii

stargazer(lm(psoda~prpblck+income, discrim), type = "text", digits = 7)

===============================================  
 Dependent variable:   
 ---------------------------  
 psoda   
-----------------------------------------------  
prpblck 0.1149882\*\*\*   
 (0.0260006)   
   
income 0.0000016\*\*\*   
 (0.0000004)   
   
Constant 0.9563196\*\*\*   
 (0.0189920)   
   
-----------------------------------------------  
Observations 401   
R2 0.0642204   
Adjusted R2 0.0595180   
Residual Std. Error 0.0861147 (df = 398)   
F Statistic 13.6569100\*\*\* (df = 2; 398)  
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

If the proportion of black people increases by 100 percent, or double the population, then will increase by 11 cents. This does not seem economically large.

### Part iii

stargazer(lm(psoda~prpblck, discrim), type = "text", digits = 7)

===============================================  
 Dependent variable:   
 ---------------------------  
 psoda   
-----------------------------------------------  
prpblck 0.0649269\*\*\*   
 (0.0239570)   
   
Constant 1.0373990\*\*\*   
 (0.0051905)   
   
-----------------------------------------------  
Observations 401   
R2 0.0180755   
Adjusted R2 0.0156145   
Residual Std. Error 0.0881018 (df = 399)   
F Statistic 7.3448910\*\*\* (df = 1; 399)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

The model from part ii showed an 11.5 cent increase while the model from part iii showed 6.5 cent increase when the black population increases by 100 percent. The discrimination effect was larger when controlled for income.

### Part iv

stargazer(lm(log(psoda)~prpblck+log(income), discrim), type = "text", digits = 7)

===============================================  
 Dependent variable:   
 ---------------------------  
 log(psoda)   
-----------------------------------------------  
prpblck 0.1215803\*\*\*   
 (0.0257457)   
   
log(income) 0.0765114\*\*\*   
 (0.0165969)   
   
Constant -0.7937680\*\*\*   
 (0.1794337)   
   
-----------------------------------------------  
Observations 401   
R2 0.0680923   
Adjusted R2 0.0634093   
Residual Std. Error 0.0821007 (df = 398)   
F Statistic 14.5404600\*\*\* (df = 2; 398)  
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

The estimated percentage change in would be 2.43.

### Part v

stargazer(lm(log(psoda)~prpblck+log(income)+prppov, discrim), type = "text", digits = 7)

===============================================  
 Dependent variable:   
 ---------------------------  
 log(psoda)   
-----------------------------------------------  
prpblck 0.0728073\*\*   
 (0.0306756)   
   
log(income) 0.1369552\*\*\*   
 (0.0267554)   
   
prppov 0.3803597\*\*\*   
 (0.1327903)   
   
Constant -1.4633320\*\*\*   
 (0.2937110)   
   
-----------------------------------------------  
Observations 401   
R2 0.0869615   
Adjusted R2 0.0800620   
Residual Std. Error 0.0813676 (df = 397)   
F Statistic 12.6039700\*\*\* (df = 3; 397)  
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

drops by 0.05

### Part vi

frame <- data.frame("log\_income" = log(discrim$income), "prpov" = discrim$prppov)  
frame <- na.omit(frame)  
cor(frame$log\_income, frame$prpov)

[1] -0.838467

The correlation is -0.839. This is roughly what I expected.

### Part vii

The multicollinearity of the two variables has no impact on bias and with just two independent variables having that level of correlation, it should also have very little impact on variance. The given statement is not true.