HW4 Group 1, Austin Halvorsen

Pedram Jahangiry

Oct 15 2020

# Problems

## Question 1

### (i)

If they trade sleep for work, then would have a negative sign ()

### (ii)

I would think that and will have positive signs. However, I think it is hard to tell if someone with more education would more inclined to get a healthy amount of sleep, or be pressed to pursue their desires in academic or research fields causing them to lose out on sleep and pursue their goals.

### (iii)

Since we are measuring in minutes we first calculate 5 hours in minutes: . Since sleep is predicted to fall by , that gives us a total of 44.4, or approximately 45 minutes of lost sleep. This really is not a large trade off considering that this would be 45 minutes per week of lost sleep or about 6 minutes per night.

### (iv)

The negative sign of our our coefficient means that the more education we get, the less sleep we are likely to get. The magnitude does not have a large impact however. If we assume someone with a high school degree versus a Bachelors degree is a four year difference, that means a college graduate loses about 45 minutes of sleep a week.

### (v)

The model returns an value of only 11.3%, so the explanatory variables do not explain a lot of variation in this model. We are overlooking many other things like marital status, kids, income, or health. These are would be correlated with because someone with kids, or married, will have those factors affecting their sleep.

## Question 2

### (i)

We would expect to have because the lower the rank (higher the number) the *less* prestigious that law school is. In other words, for each increase rank (i.e. 2 to 3 or 30 to 31) you are thought to be less prestigious and therefore have a lower salary.

### (ii)

I would expect the other variable to have the following signs:

* and would most likely be positive as these are measures that don’t depend on what school you attend, so someone who scored higher and had a higher GPA is more likely to have a higher salary.
* and are also more likely to have positive signs as well as they indicate a better rank of the school and would therefore lead to higher salaries.

### (iii)

That would be the coefficient expressed as a percent, so **24.8%**

### (iv)

The coefficient on implies that a **1%** increase in library volume translates to a **0.095%** increase in median starting salary, ceteris paribus.

### (v)

I would say you are better off going to a school with a lower rank. The difference between schools ranked 20 apart would be or about 6.6% less in starting median salary.

## Question 3

### (i)

could have decreased because we have reduced the number in the sample from 142 -> 99. Additionally, can only increase (or decrease) by adding or removing variables. So the only way that it could decrease is if it were to have a reduced sample size or an irrelevant variable included that has no correlation with y or any x.

## Question 4

### (i)

house\_mrm <- lm(price~sqrft+bdrms, data = house)  
stargazer(house\_mrm, 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

From our regression results, our estimated model would be:

This give and a sample of

### (ii)

Holding square footage constant, if we add one more bedroom the price would increase by 15.2 thousand ($15,200).

### (iii)

If we add a bedroom and thus creating an additional 140 square feet in size to the house, the estimated price would be: or approximately $33.1 thousand ($33,100). This is more than double the price of just adding a bedroom.

### (iv)

Using our value, the model explains about 63% of variation from square footage and number of bedrooms.

### (v)

If a home has 2,438sqft and 4 bedrooms, it will be predicted to sale for:

Or $353,540

### (vi)

If the actual sale price for this home was $300,000, but the predicted price was $353,540, then the residual would be . This would mean that the buyer underpaid for this house.

## Question 5

### (i)

discrim\_summary <- price %>%   
 select(prpblck, income) %>%   
 filter(!is.na(prpblck),  
 !is.na(income)) %>%   
 summarise\_all(list(mean, sd)) %>%   
 rename('Mean (prpblck)' = prpblck\_fn1,  
 'Mean (Income)' = income\_fn1,  
 'SD (prpblck)' = prpblck\_fn2,  
 'SD (income)' = income\_fn2)  
  
pander(discrim\_summary)

|  |  |  |  |
| --- | --- | --- | --- |
| Mean (prpblck) | Mean (Income) | SD (prpblck) | SD (income) |
| 0.1135 | 47054 | 0.1824 | 13179 |

The units for these measures are dollars for income and prpblck represents the proportion of the black population.

### (ii)

discrim\_mrm <- lm(psoda~prpblck+income, data = price)  
stargazer(discrim\_mrm, type='text', digits = 10)

==================================================  
 Dependent variable:   
 ------------------------------  
 psoda   
--------------------------------------------------  
prpblck 0.1149882000\*\*\*   
 (0.0260006400)   
   
income 0.0000016027\*\*\*   
 (0.0000003618)   
   
Constant 0.9563196000\*\*\*   
 (0.0189920100)   
   
--------------------------------------------------  
Observations 401   
R2 0.0642203900   
Adjusted R2 0.0595179800   
Residual Std. Error 0.0861147000 (df = 398)   
F Statistic 13.6569100000\*\*\* (df = 2; 398)  
==================================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Basing our model of the above regression output, we would get the following model:

The resulting with a sample size of . Since already represents proportion, a 1 increase would represent a 100% increase, not a single unit. So to make this interpretation more meaningful, we will consider a 0.01 increase (or a 1% increase) in to interpret our results. If we take our coefficient of we get an increase of less than 1 cent, or essentially no effect. On an individual transaction level, this effect seems minuscule. However, if populations increase more dramatically, say 10%-20% we would observe large increases in the the price of soda.

### (iii)

discrim\_srm <- lm(psoda~prpblck, data = price)  
stargazer(discrim\_srm, type='text', digits = 10)

=================================================  
 Dependent variable:   
 -----------------------------  
 psoda   
-------------------------------------------------  
prpblck 0.0649268700\*\*\*   
 (0.0239569700)   
   
Constant 1.0373990000\*\*\*   
 (0.0051904520)   
   
-------------------------------------------------  
Observations 401   
R2 0.0180755100   
Adjusted R2 0.0156145500   
Residual Std. Error 0.0881017700 (df = 399)   
F Statistic 7.3448910000\*\*\* (df = 1; 399)  
=================================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

When we control for income, the estimated coefficient for prpblck is 0.065, which is much lower than the estimate that did not control for income. This means that the discrimination effect goes down when we do not include income in our model.

### (iv)

discrim\_log <- lm(log(psoda)~prpblck+log(income), data = price)  
stargazer(discrim\_log, type='text')

===============================================  
 Dependent variable:   
 ---------------------------  
 log(psoda)   
-----------------------------------------------  
prpblck 0.122\*\*\*   
 (0.026)   
   
log(income) 0.077\*\*\*   
 (0.017)   
   
Constant -0.794\*\*\*   
 (0.179)   
   
-----------------------------------------------  
Observations 401   
R2 0.068   
Adjusted R2 0.063   
Residual Std. Error 0.082 (df = 398)   
F Statistic 14.540\*\*\* (df = 2; 398)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

The estimated model is:

Where and

If increases by 0.20, then we predict the price of soda to increase by or **2.44%**

### (v)

discrim\_pov <- lm(log(psoda)~prpblck+log(income)+prppov, data = price)  
stargazer(discrim\_pov, type='text')

===============================================  
 Dependent variable:   
 ---------------------------  
 log(psoda)   
-----------------------------------------------  
prpblck 0.073\*\*   
 (0.031)   
   
log(income) 0.137\*\*\*   
 (0.027)   
   
prppov 0.380\*\*\*   
 (0.133)   
   
Constant -1.463\*\*\*   
 (0.294)   
   
-----------------------------------------------  
Observations 401   
R2 0.087   
Adjusted R2 0.080   
Residual Std. Error 0.081 (df = 397)   
F Statistic 12.604\*\*\* (df = 3; 397)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

The estimated model is:

Where and

When we add to the model, our coefficient falls from 0.122 to 0.073.

### (vi)

cor(log(price$income), price$prppov, method = 'pearson', use = 'complete.obs')

[1] -0.838467

The correlation is . I would expect this because you would expect income to decrease with higher poverty rates.

### (vii)

Because they are not perfectly correlated and only highly correlated, they should remain in the analysis as they add an additional variable to help explain the effect of discrimination in the data.