Logan Albiani

11/29/18

ECN 102

Prof. Siegler

Problem Set #3

1. *Types of Colleges, Parental Income, and their Children’s Incomes in Adulthood*
   1. ***Use R and stargazer package to estimate and report the OLS regression model results using the data set income.csv, which is attached as part of this problem set.***

===============================================

Dependent variable:

---------------------------

kmedian

-----------------------------------------------

hselective 14,268.320\*\*\*

(1,413.672)

midwest 1,383.552\*\*

(638.902)

pctfemale -178.044\*\*\*

(17.449)

pctkmarried 80.873\*\*\*

(22.201)

parmedian 0.168\*\*\*

(0.012)

public -457.855

(1,489.233)

private 128.198

(1,419.041)

ivyplus 27,754.520\*\*\*

(2,897.113)

otherelite 16,949.220\*\*\*

(1,699.537)

selective 7,972.445\*\*\*

(1,053.013)

northeast 5,965.956\*\*\*

(624.987)

west 1,335.548\*

(738.892)

Constant 22,970.280\*\*\*

(1,711.888)

-----------------------------------------------

Observations 1,350

R2 0.565

Adjusted R2 0.561

Residual Std. Error 8,551.116 (df = 1337)

F Statistic 144.925\*\*\* (df = 12; 1337)

===============================================

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

* 1. ***Suppose someone claims that attending a selective college or university increases the median earnings at age 34 by more than $6,000 per year. What are the null and alternative hypotheses? What is the value of the t-test statisc in this case? Can you reject the null hypothesis at the 5 percent significance level? What can you conclude based on this t-test? Explain***

Null Hypothesis: H0: b06000

Alternative Hypothesis: HA: b1 > 6000

Tn-k = (b1-b0)/se(b1) => (14,268.32-6,000)/1,413.672 = **5.846**

CI: 14268 + (1.96\*1413.672) = 17041 and 14268– (1.96\*1413.672) = 11495

CI: [11495, 17041]

$6,000 lies outside of the confidence interval. In fact, it is far below the interval. Therefore, the null hypothesis must be incorrect, and we can say, with 95% confidence, that the alternative hypothesis is correct. Implicitly, this means that due to the regression model producing an expected value of $14,268 ‘extra’ dollars for those who attend highly selective universities, we would need to see $6000 fall within its corresponding 95% confidence interval in order to accept the null hypothesis. Because we don’t see this, it means it is highly unlikely that graduates from highly selective schools make $6000 or less than do those don’t attend highly selective schools.

* 1. ***What proportion of the variance in median earnings at age 34 is explained by the regression model?***

In the report generated the stargazer package, we can see that adjusted R2 = .561. Because the only regressor used is hselective, the reported coefficient of determination is solely referring to it. This means 56.1% of the variation in the data run through this particular regression model is explained by the various regressors present in the model.

* 1. ***All else equal, suppose that two random individuals are equivalent in every way except one student attends a college where the median parents’ income is $50,000 and the other’s is $200,000. What is the predicted difference in income at the age of 34?***

Ŷ = B0 + b1x1 => $22970 + .168x

$22970 + .168($50000) = $31370

$22970 + .168($200000) = $56570

In this particular model, where the only regressor in use is parental income, we can see that 16.8% of a parent’s income is predicted to be added onto the constant. This means that a high income, such as $200000 will greatly contribute to the success of the child in terms of wealth.

* 1. ***What is the predicted medium earning at age 34 for an individual who attended a public, highly-selective university in the Midwest where 54% of the student body was female, 60% of the graduates were married at age 34, and who attended a college or university where the median parents’ income is $100,000 per year? Show your work and explain.***

Ŷ = B0 + b1x1 + b2x2 + b3x3 + b4x4 + b5x5 + b6x6

Ŷ = 22970 – 1(458) + 1(14268) – .54(178) + .6(81) + .168(100000)

Ŷ = $53532

In this model, we use five regressors: midwest, parmedian, pctfemale, pctkmarried, and hselective. Due to the fact that 60% of individuals are married by 34 and 54% are female, we must multiply each regressor by that corresponding percentage. Alongside this, similar to the last problem, 16.8% of a parent’s income is predicted to linearly add to a graduate’s income. Furthermore, being a student at a highly selective school has a positive effect on income while attending school in the midwest has a negative impact. In all, once all regressors are accounted for, the model predicts that an individual with these characteristics will have an income of $53,532 by the age of 34.

* 1. ***UC Davis is observation 1092. Use the regression results to find the size of UC Davis’s residual. Do Davis graduates do better or worse than the model predicts? Is this a good thing or a bad thing for you in the future? Explain***

Using the linear regression model which accounts for all regressors, UC Davis’s residual $17140. This means that the model predicts UC Davis graduates to make $17140 less than what they actually do. If the model is to be taken as the golden standard of predicting my future income, then I would believe this is good news for me, as I’ll be making more money than expected as per the model. However, considering

𝑒1 = Y1 - Ŷ1

**17140** = 61600 –44460

* 1. ***Suppose the goal is to earn a high income by 34. In plain English, use the regression results above to explain which type of university is the best to attend in order to maximize the possibility of achieving this goal.***

We have a few possibilities in terms of schools of which we can attend: public, private, selective, highly selective, and Ivy League/elite. Because we see the highest value associated with private universities and Ivy Leagues in their respective categories, then we can confidently say the best school to attend in order to increase the likelihood of achieving this goal would be to attend an elite private school in the northeast.

1. *Discrimination in the Market for Basketball Trading Cards*
   1. ***Which explanatory variables in the regression model above are NOT dummy variables? Briefly explain.***

Dummy variables are present in a regression model when they are simply regressors that a particular piece of data do or do not have. I short, they are bivariate. This means that those variables which are not dummy variables are those which can have a range of possible values. These are availability and points contributed.

* 1. ***Which of the estimated coefficients are NOT statistically different from zero at the 5 percent level of significance? Briefly explain.***

Tn-k = (b1)/se(b1) ~ T.95 = 1.645

|  |  |  |  |
| --- | --- | --- | --- |
| **Explanatory Variable** | **Coefficients** | **Standard Error** | **T-value** |
| *Availability* | -5.723 | .363 | -15.77 |
| *Pointers Contributed Per Game* | .109 | .016 | 6.81 |
| *Point Guard* | -.225 | .181 | -1.24 |
| *Small Forward* | -.329 | .168 | -1.96 |
| *Power Forward* | -.249 | .179 | -1.39 |
| *Center* | -.562 | .184 | -3.05 |
| *Black non-BHOF* | -.991 | .187 | -5.30 |
| *White non-BHOF* | -.782 | .203 | -3.85 |
| *White BHOF* | -.609 | .218 | -2.79 |

The only two variables which have t-values which are less than |T.95| (1.645) are power forward and point guard. This means that these two variables have no statistical evidence to say that they have a significant effect on the availability of a particular basketball card.

* 1. ***Suppose that a player contributes one additional point per game, all else equal. How much are card prices predicted to change as a result, all else equal? Briefly explain.***

BPoints/game = .109

.109 \* 1 = .109 or approximately 11 cents.

Therefore, every additional point contributed per game will result, as per the model, in approximately an 11-cent increase in the value of that particular player’s card.

* 1. ***Based on the regression results, what can you conclude about the impact of a player’s race on the value of their trading card prices? Is there evidence of customer discrimination against black players? Explain as specifically as possible using the regression results.***

Using the two comparable results that we have on players’ race (black non-hall of fame and white non-hall of fame) we can compare their T-values. At the ninety-five percent level of significance, it seems as though race has an effect on both races’ card values. However, with black non-hall of fame players, the effect is more pronounced and evinced by the fact that the T-value for this is significantly larger (approximately two standard deviations more) for blacks than is for whites. On top of this, the amount by which race has an effect on basketball cards is around twenty cents larger (in the negative direction) than it is for whites. If we assume that the market for basketball cards follows a normal competitive market, and that equilibrium in supply and demand create price, then it is well within the realm of possibilities that consumers have less demand for black players’ cards than they do for white players’ cards. Yes, there is evidence for consumer discrimination.

***R-Code***

## Problem 1 ##

olsall <- lm(kmedian ~ hselective + midwest + pctfemale + pctkmarried + parmedian + public + private+ ivyplus + otherelite + selective + northeast + west, data=income)

library(stargazer)

summary(olsall)

stargazer(olsall, type="text")

confint(olsall, level=.95)

residuals <- residuals(olsall)

options(max.print=1500)

residuals(olsall)