Problem Set 2 Solutions

BUAN 6356

Due: Monday, 2018-02-19-11:59pm

Deliverable: an R source-code file named ps2.r

Question 1

Data The data attend.csv contains 680 observations on students in a course on microeconomic principles.

Analysis

- Read the data attend.csv into a new variable 'context1' and familiarize yourself with the data.
- Create a new variable 'attendrt' for the attendance rate (classes attended out of the total number of classes). You can find information about the total number of classes in the labels .txt file.
- Create a new variable 'hwrt' for the homework completion rate (homework turned in of out the total number of homework assignments). You can find information about the total number of homework assignments in the labels .txt file.
- Run the following linear model using the 'lm' function. Store the result in: model1

$$\text{termGPA}_i = \beta_0 + \beta_1 \text{priGPA}_i + \beta_2 \text{ACT}_i + \beta_3 \text{attendrt}_i + \beta_4 \text{hwrt}_i + e_i \quad (1)$$

Interpretations a. Using model1, how does termGPA change with respect to attendance?

Every 10% increase in attendance rate is associated with a .105 point increase in term GPA controlling for prior GPA, ACT score, and homework completion.

b. How does termGPA change with respect to homework?

Every 10% increase in the homework completion rate is associated with a .091 point increase in term GPA controlling for prior GPA, ACT score, and attendance.

c. Predict the termGPA for a student with a 32 ACT and a 2.2 priGPA who attended 28 lectures and turned-in 8 homework assignments.

2.91 term GPA out of a possible 4.0

d. Predict the termGPA for a student with a similar attendance and homework pattern who had a 20 ACT and a 3.9 priGPA.

3.41 term GPA out of a possible 4.0

e. Intuitively, which variable is more important to the termGPA, ACT or priGPA?

priGPA seems to be more important

f. Predict the termGPA for a student with a 25 ACT and a 3.0 priGPA who attends all the classes, but only finishes half the homework assignments.

2.77 term GPA out of a possible 4.0

g. Predict the termGPA for a similarly qualified student who turns in all the homework assignments, but only attends half the classes.

2.70 term GPA out of a possible 4.0

h. Intuitively, which variable is more important to the termGPA, attendance or homework completion?

Attendance

i. Why is it easier to compare attendrt and hwrt than it is to compare priGPA and ACT score?

The units are the same.

Question 2

Data The data set in CEOSAL2.csv contains information on chief executive officers for U.S. corporations. The variable salary is annual compensation, in thousands of dollars.

Analysis

- Read the data CEOSAL2.csv into a new variable 'context2' and familiarize yourself with the data.
- Run the following linear model using the 'lm' function. Store the result in: model2 (Remember, you can use the natural log of a variable in the model just by adding 'log( )' around the variable)

$$\ln[\text{salary}_i] = \beta_0 + \beta_1 \ln[\text{mktval}_i] + \beta_2 \text{profits}_i + \beta_3 \text{ceoten}_i + e_i \tag{2}$$

- Run the following linear model using the 'lm' function. Store the result in: model3

$$\ln[\text{salary}_i] = \beta_0 + \beta_1 \ln[\text{mktval}_i] + \beta_2 \text{profits}_i + \beta_3 \text{ceoten}_i + \beta_4 \ln[\text{sales}_i] + e_i \tag{3}$$

Interpretations a. We used natural logs on all the dollar-valued quantities except profits in models 2 \& 3 (eq. 2 \& 3). Why did we not take the log of profits?

Profits are sometimes negative.

b. Interpret the estimated coefficient on log mktval in model 2 (eq 2).

Every 10% increase in market value is associated with a 2.39% increase in CEO salary controlling for profits and CEO tenure.

c. Interpret the estimated coefficient on log mktval in model 3 (eq 3).

Every 10% increase in market value is associated with a 1.02% increase in CEO salary controlling for profits, CEO tenure, and sales.

d. Compare the test statistics on log mktval between model 2 and model 3. Please explain the differences you find in terms of the biases we discussed in class.

In model2 the t-stat on market value is 4.267 (significant at the 0.1% level) whereas in model3 the t-stat is 1.614 (insignificant at the 10% level). Including sales eliminates the effect of market value on CEO salary. Because of omitted variable bias, market value was a proxy-variable entirely capturing the effect of sales on CEO salary.

e. Is the coefficient on profits significant in model 3 (eq 3)?

No, $t = 0.193$ and $p = 0.8470$

f. Interpret the estimated coefficient on log sales in model 3 (eq 3).

Every 10% increase in sales is associated with a 1.62% increase in CEO salary controlling for profits, CEO tenure, and market value of the firm.

Question 3

Data hprice1.csv contains data on 88 U.S. houses, their characteristics, and their prices at the time of sale.

Analysis

- Read the data hprice1.csv into a new variable 'context3' and familiarize yourself with the data.
- Run the following linear model using the 'lm' function. Store the result in: model4

$$\text{price}_i = \beta_0 + \beta_1 \text{bdrms}_i + \beta_2 \ln[\text{lotsize}_i] + \beta_3 \ln[\text{sqrft}_i] + \beta_4 \text{colonial}_i + e_i \tag{4}$$

- Run the following linear model using the 'lm' function. Store the result in: model5

$$\ln[\text{price}_i] = \beta_0 + \beta_1 \text{bdrms}_i + \beta_2 \ln[\text{lotsize}_i] + \beta_3 \ln[\text{sqrft}_i] + \beta_4 \text{colonial}_i + e_i \tag{5}$$

Interpretations a. Using model4, how does price change with repect to lotsize?

A 100% change in lot size is associated with a \$61,445.71 increase in the value of the home controlling for number of bedroom, square footage, and whether or not the home was in the colonial style.

b. Using model5, how does price change with repect to lotsize?

A 1% change in lot size is associated with a 0.16781898% change in price controlling for number of bedroom, square footage, and whether or not the home was in the colonial style.

c. Using model4, interpret the estimated coefficient on colonial (please ignore significance here).

Having a colonial house adds \$4,133.61 to your home value on average controlling for number of bedroom, lot size, and square footage.

d. Which model (4 or 5) better fits the data for this data set? On what criterion/criteria are you basing your judgement?

The R-squared in model4 is 0.6781 whereas the R-squared in model5 is 0.6491. Therefore, we choose model4 because the R-squared is higher.

e. Suppose your house is worth \$300k. You are considering an expansion of your home to add a master suite (+1 bedroom to your home). This expansion would increase your square-footage by 10% and would cost \$50k. You have valued your enjoyment of the additional space at \$20k, so you would only be willing to consider the build if it were to also increase your property value accordingly. Does the appropriate model indicate that you should pursue the expansion?

Using model4, we find that the value of the expansion is \$41,123.17. This with the \$20,000 value to your family has a \$61,000 value to your family total, greatly exceeds the cost of the expansion. Therefore, it is optimal for you to purchase the expansion.

Question 4

Data The data in JTRAIN2.csv come from a job training experiment conducted for low income men in the U.S. during 1976-1977; see Lalonde (1986).

Analysis

- Read the data JTRAIN2.csv into a new variable 'context4' and familiarize yourself with the data.
- Run the following linear model using the 'lm' function. Store the result in: model6

$$\text{re78}_i = \beta_0 + \beta_1 \text{re75}_i + \beta_2 \text{train}_i + \beta_3 \text{educ}_i + \beta_4 \text{black}_i + e_i \tag{6}$$

Interpretations a. Using model6, how does real earnings in '78 change with respect to real earnings in '75?

Every dollar of real earnings in 1975 is related to a \$0.14697 increase in real earnings in 1978 controlling for job training, race, and education.

b. Using model6, how does real earnings in '78 change with respect to training? Is the coefficient significant?

The value of job training is an extra \$1,684.22 increase in real earnings in 1978 controlling for real earnings in '75, race, and education. The coefficient is significant at the 1% level.

c. Interpret the estimated coefficient on black from model6.

Being black has an associated disadvantage of $-\$2,112.77$ controlling for education, job training, and real earnings in '75.