Homework 9

Multiple Regression Model - Further Issues (15 points)

Due Date: April 5 at 11:59 pm

Instruction:

- This HW must be done in Rmarkdown!
- Please submit both the .rmd and the Microsoft word files. (Do not submit a PDF or any other image files as the TAs are going to give you feedback in your word document)
- Name your files as: HW9-groupnumber-name
- All the HW assignments are individual work. However, I highly encourage you to discuss it with your group members.
- Late homework assignments will not be accepted under any circumstances.

Problems

Question 1 Using the data in RDCHEM, the following equation was obtained by OLS:

$$\widehat{rdintens} = \underset{(.429)}{2.613} + \underset{(.00014)}{.00030}sales - \underset{(.0000000037)}{.0000000070}sales^2$$

$$n = 32 \quad R^2 = .1484.$$

(i) At what point does the marginal effect of sales on rdintens become negative?

(ii) Would you keep the quadratic term in the model? Explain.

(iii) Define salesbil as sales measured in billions of dollars: salesbil = sales/1,000. Rewrite the estimated equation with salesbil and $salesbil^2$ as the independent variables. Be sure to report standard errors and the R-squared. [Hint: Note that $salesbil^2 = sales^2/(1,000)^2$.]

(iv) For the purpose of reporting the results, which equation do you prefer?

Question 2 The following three equations were estimated using the 1,534 observations in 401K:

$$\widehat{prate} = \underset{(.78)}{80.29} + \underset{(.52)}{5.44}mrate + \underset{(.045)}{.269}age - \underset{(.00004)}{.00013}totemp$$
$$R^2 = .100 \quad \bar{R}^2 = .098.$$

$$\widehat{prate} = \underset{(1.95)}{97.32} + \underset{(0.51)}{5.02}mrate + \underset{(.044)}{.314}age - \underset{(.28)}{2.66}log(totemp)$$
$$R^2 = .144 \quad \bar{R}^2 = .142.$$

$$\widehat{prate} = \underset{(.78)}{80.62} + \underset{(.52)}{5.34}mrate + \underset{(.045)}{.290}age - \underset{(.00009)}{.00043}totemp + \underset{(.00000000010)}{.0000000039}totemp^2$$
$$R^2 = .108 \quad \bar{R}^2 = .106.$$

Which of these three models do you prefer? Why?

Computer Exercises

Question 3 Use the data in KIELMC, only for the year 1981, to answer the following questions. The data are for houses that sold during 1981 in North Andover, Massachusetts; 1981 was the year construction began on a local garbage incinerator.

(i) To study the effects of the incinerator location on housing price, consider the simple regression model

$$log(price) = \beta_0 + \beta_1 log(dist) + u,$$

where price is housing price in dollars and dist is distance from the house to the incinerator measured in feet. Interpreting this equation causally, what sign do you expect for β_1 if the presence of the incinerator depresses housing prices? Estimate this equation and interpret the results.

(ii) To the simple regression model in part (i), add the variables log(intst), log(area), log(land), rooms, baths, and age, where intst is distance from the home to the interstate, area is square footage of the house, land is the lot size in square feet, rooms is total number of rooms, baths is number of bathrooms, and age is age of the house in years. Now, what do you conclude about the effects of the incinerator? Explain why (i) and (ii) give conflicting results.

(iii) Add $[log(intst)]^2$ to the model from part (ii). Now what happens? What do you conclude about the importance of functional form?

(iv) Is the square of log(dist) significant when you add it to the model from part (iii)?

Question 4 Consider a model where the return to education depends upon the amount of work experience (and vice versa):

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 educ * exper + u$$

(i) Show that the return to another year of education (in decimal form), holding exper fixed, is $\beta_1 + \beta_3 exper$

(ii) State the null hypothesis that the return to education does not depend on the level of exper. What do you think is the appropriate alternative?

(iii) Use the data in WAGE2 to test the null hypothesis in (ii) against your stated alternative. Assume that significance level $\alpha = 0.05$

(iv) Let θ_1 denote the return to education (in decimal form), when $exper = 10$: $\theta_1 = \beta_1 + 10\beta_3$ Obtain $\hat{\theta}_1$ and a 95% confidence interval for θ_1. (Hint: Write $\beta_1 = \theta_1 - 10\beta_3$ and plug this into the equation; then rearrange. This gives the regression for obtaining the confidence interval for θ_1.)