Problem Set 3 Solutions

Minzhi Meng

October 25, 2019

Question 1

To capture the life-cycle pattern of wages, consider the following multiple linear regression model

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \epsilon$$

where $wage$ is the wage rate measured in dollar per hour, $educ$ is years of education, $exper$ is years of work experience, $exper^2$ is squared years of experience (in the dataset is called $expersq$).

(1) Estimate the equation and report the results. (For full credit give your answer as a formula and a word explanation)

Solution:
$wage = -3.9649 + 0.5953educ + 0.2683exper – 0.0046exper^2$
Standard interpretation. For example take the coefficient on $educ$. The estimated coefficient for $educ$ (β_1) is 0.5953. $pvalue < 0.05$, thus at 0.05 level of significance, the coefficient β_1 is statistically significant. It says that on average, one extra year of education increases the wage of a person by 0.5963 dollar per hour, holding other variables constant.

(2) What is the marginal effect of a year increase in the work experience for a person with 18 years of work experience?

Solution:
$\frac{\partial wage}{\partial exper}|_{educ} = 0.2683 - 2 * 0.0046 * exper = 0.2683 - 2 * 0.0046 * 18 = 0.1027$

Question 2

Load data via link below and store in dataframe $fare$, where $year$ is 1997, 1998, 1999, 2000, id is route identifier, $dist$ is the distance of the route measured in miles, $passen$ is the average number of passengers per day, $fare$ is the average one-way fare of the route measured in dollar per day, $conc$ is the percent of market controlled by the biggest carrier on route

(1) Generate a new variable called $yr00$ which takes on values 1 if the observation is in 2000 and 0 otherwise (a dummy variable).

Solution:
skipped...

(2) Estimate the model 1 using OLS and print the summary results.

Solution:
Model 1): $fare = 46.1894 + 0.0888dist + 0.7336conc + \epsilon$

(3) Consider the null hypothesis that $\beta_2 = 0$ and $\beta_3 = 0$ in Model 1. What is the alternative hypothesis? Conduct the test and state your decision.

Solution:
$F(2, 4593) = 1633$
$Prob > F = 0.0000$
The alternative hypothesis is that either of the coefficients is nonzero. The p value is very low so we reject the null in favor of the alternative implying that these coefficients are nonzero.

(4) State in words and numbers how we should interpret the estimated coefficients in Model 1. (for full credit explain how you know if it is statistically signicant)

Solution:
Standard interpretation. For example take the coefficient on *conc*. $\beta_3 = 0.7336$, $pvalue < 0.05$, thus at 0.05 level of significance, the coefficient for *conc* (β_3) is statistically significant. It refers to market concentration and thus on average, an 1% increase in the market concentration increases the fare by 0.7336 dollar, holding other variables constant.

(5) Generate an interaction term called $yr00dist$ by multiplying the variable $dist$ by the variable you created in part (1), $yr00dist = yr00 * dist$

Solution:
skipped...

(6) Estimate the model 2 using OLS and print the summary results.

Solution:
Model 2): $fare = 41.6998 + 0.0893dist + 0.7438conc + 14.7778yr00 - 0.0016yr00dist + \epsilon$

(7) What is the base group for Model 2?

Solution:
Where the dummy variables take the 0 values - in this case, where yr00 = 0, that is in all other years than 2000

(8) State in words and numbers how we should interpret the estimated coecients respectively in Model 2. (for full credit explain how you know if it is statistically signicant)

Solution:
The coefficient of yr00dist is -0.0016. We can interpret that on average fare in 2000 decreases by 0.1% less than it does in other years than 2000 when distance increases by

one mile, holding other variables constant. However, the p value is 0.604 which is larger than 0.05, thus at the level of 0.05 significance, the interaction term is not statistically significant.