Niclas Knecht (niclas.knecht@u-bordeaux.fr)
Econometrics
Spring 2024

Problem Set 3

Problem 1

Consider a linear model to explain monthly beer consumption:

$$beer = \beta_0 + \beta_1\, inc + \beta_2\, price + \beta_3\, educ + \beta_4\, female + u,$$

with $\mathbb{E}[u|inc, price, educ, female] = 0$ and $Var[u|inc, price, educ, female] = \sigma^2\, inc^2$. Write the transformed equation that has a homoskedastic error term.

Problem 2 (Stata)

Use the data in **td3_houseprices.dta** to estimate the model

$$price = \beta_0 + \beta_1\, sqrft + \beta_2\, bdrms + u,$$

where *price* is the house price measured in thousands of dollars.

a) Write out the results in equation form.

b) What is the estimated increase in price for a house with one more bedroom, holding square footage constant?

c) What is the estimated increase in price for a house with an additional bedroom that is 140 square feet in size? Compare this to your answer in b).

d) What percentage of the variation in price is explained by square footage and number of bedrooms?

e) The first house in the sample has $sqrft = 2,438$ and $bdrms = 4$. What is the predicted selling price for this house from the OLS regression?

f) The actual selling price of the first house in the sample was USD 300,000 (so $price = 300$). Find the residual for this house. Does it suggest that the buyer underpaid or overpaid for the house?

Problem 3 (Stata)

The file **td3_ceo.dta** contains data on 177 chief executive officers, which can be used to examine the effects of firm performance on CEO salary.

a) Estimate a model relating annual salary to firm sales and market values. Make the model of the constant elasticity variety for both independent variables. Write the results out in equation form.

$$log(salary) = \beta_0 + \beta_1\, log(sales) + \beta_2\, log(mktval) + u.$$

b) Add profits to the model from a). Why can this variable not be included in logarithmitic form? Would you say that these firm performance variables explain most of the variation in CEO salaries?

c) Add the variable *ceoten* (years as CEO with this company) to the model from b). What is the estimated percentage return for another year of CEO tenure, holding other factors fixed?

d) Find the sample correlation coefficient between the variables $log(mktval)$ and $profits$. Are these variables highly correlated? What does this say about the OLS estimators?

Problem 4 (Stata)

Use the dataset **td3_gpa.dta** on 4,137 college students and estimate the following equation by OLS:

$$colgpa = \beta_0 + \beta_1\, hsperc + \beta_2\, sat + u,$$

where *colgpa* is the grade point average (GPA) after the fall semester measured on a four-point scale, *hsperc* is the percentile in the high school graduating class (e.g. if a student is in the top-5% of their class, $hsperc = 5$), and *sat* is the combined maths and verbal score on the Student Achievement Test.

a) Why does it make sense for the coefficient on *hsperc* to be negative?

b) What is the predicted college GPA when $hsperc = 20$ and $sat = 1050$?

c) Suppose that two high school graduates, A and B, graduated in the same percentile from high school, but student A's SAT schore was 140 points higher (about one standard deviation in the sample). What is the predicted difference in college GPA for these two students? Is the difference large?

d) Holding *hsperc* fixed, what difference in SAT scores leads to a predicted *colgpa* difference of 0.50, or one half of a grade point? Comment on your answer.

Problem 5 (Stata)

Consider a model where the return to education depends upon the amount of work experience:

$$log(wage) = \beta_0 + \beta_1\, educ + \beta_2\, exper + \beta_3\, educ \cdot exper + u.$$

a) Show that the return to another year of education, holding *exper* fixed, is $\beta_1 + \beta_3\, exper$.

b) 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? Use the data in **td3_wage.dta** to test the null hypothesis against your stated alternative.

c) Let θ_1 denote the return to education, 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 into the equation; then, rearrange. This gives the regression for obtaining the confidence interval for θ_1.)

Problem 6 (Stata)

Consider the data in **td3_sleep.dta**. The variable *sleep* is total minutes per week spent sleeping at work, *totwrk* is total weekly minutes spent working, *educ* and *age* are measured in years, and *male* is a gender dummy. Estimate the following equation with OLS:

$$sleep = \beta_0 + \beta_1\, totwrk + \beta_2\, educ + \beta_3\, age + \beta_4\, age^2 + \beta_5\, male + u.$$

a) All other factors being equal, is there evidence that men sleep more than women? How strong is the evidence?

b) Is there a statistically significant trade-off between working and sleeping? What is the estimated trade-off?

Problem 7 (Stata)

Using data from **td3_sat.dta** consider the following equation:

$$sat = \beta_0 + \beta_1\, hsize + \beta_2\, hsize^2 + \beta_3\, female + \beta_4\, black + \beta_5\, female \cdot black + u,$$

where the variable *sat* is the combined SAT score, *hsize* is the size of the student's high school graduating class (in hundreds), *female* is a gender dummy variable equal to one for women, and *black* is a race dummy variable equal to one for black people, and zero otherwise.

a) Estimate the equation. Is there strong evidence that $hsize^2$ should be included in the model? From this equation, what is the optimal (for student's SAT scores) high school size?

b) Holding *hsize* fixed, what is the estimated difference in SAT score between non-black women and non-black men? Is this estimated difference statistically significant?

c) What is the estimated difference in SAT score between non-black men and black men? Test the null-hypothesis that there is no difference between their scores.

d) What is the estimated difference in SAT score between black women and non-black women? What would you need to do to test whether the difference is statistically significant?

Problem 8

Let y_i be independently and identically distributed with mean μ and variance σ^2, and consider linear estimators of the mean μ of the form $\hat{\mu} = \sum_{i=1}^{n} a_i y_i$.

a) Derive the restriction on a_i needed to guarantee that the estimator $\hat{\mu}$ is unbiased.

b) Derive the expression for the variance of the estimator $\hat{\mu}$.

c) Derive the linear unbiased estimator that has the minimal variance in this class of estimators.