International School of Economics at TSU

Econometrics 2

Lasha Chochua

Problem Set 4

Instructions: You are encouraged to solve the problems before the recitation. Additionally, you are encouraged to work in groups. It is **not mandatory** to submit solutions unless stated otherwise. However, if you would like to share your solution, I would be happy to review it.

Problem 1: Let X denote a 5×2 matrix and y a 5×1 vector:

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 3 \\ 1 & 5 \\ 1 & 2 \end{bmatrix}, \quad y = \begin{bmatrix} 14 \\ 17 \\ 8 \\ 16 \\ 3 \end{bmatrix}$$

a. Compute $Q = X'X$, $\det(Q)$, and Q^{-1}

b. Compute $A = (X'X)^{-1}X'$

Problem 2 Consider the regression model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i$$

Let $\hat{\beta}_0$ and $\hat{\beta}_1$ denote OLS estimators. Let ω_x and ω_y denote scaling factors for X and Y, respectively. Define:

- $Y^* = \omega_y Y$
- $X^* = \omega_x X$

What is the relationship between:

a. $\hat{\beta}_1$ and $\hat{\beta}_1^*$

b. $\hat{\beta}_0$ and $\hat{\beta}_0^*$

c. $\text{var}(\hat{\beta}_0)$ and $\text{var}(\hat{\beta}_0^*)$

d. $\text{var}(\hat{\beta}_1)$ and $\text{var}(\hat{\beta}_1^*)$

e. $\hat{\sigma}^2$ and $(\hat{\sigma}^*)^2$

f. R^2_{xy} and $R^2_{x^*y^*}$

Problem 3 Given Regression Output

$$\begin{array}{llll} \widehat{Price} = 119.2 & + 0.485\,BDR & + 23.4\,Bath & + 0.156\,Hsize \\ \quad (23.9) & \quad (2.61) & \quad (8.94) & \quad (0.011) \\ & + 0.002\,Lsize & + 0.090\,Age & - 48.8\,Poor \\ & \quad (0.00048) & \quad (0.311) & \quad (10.5) \end{array}$$

$$\bar{R}^2 = 0.72, \quad SER = 41.5$$

a. Is the coefficient on BDR statistically significantly different from 0?

b. Typically, five-bedroom houses sell for much more than two-bedroom houses. Is this consistent with your answer to (a) and with the regression more generally?

c. A homeowner purchases 2000 square feet from an adjacent lot. Construct a 99% confidence interval for the change in the value of her house.

d. Lot size is measured in square feet. Do you think that another scale might be more appropriate? Why or why not?

e. The F-statistic for omitting BDR and Age from the regression is $F = 0.08$. Are the coefficients on BDR and Age statistically different from 0 at the 10% level?

Problem 4 You are given the following data for $n = 5$ observations with Y_i representing log(wage) and $Educ_i$ representing years of education:

i	Y_i	$Educ_i$
1	2.5	12
2	3.0	16
3	2.2	10
4	2.8	14
5	2.6	13

a. Construct the $\mathbf{X}$ matrix including constants and the $\mathbf{Y}$ vector. Calculate $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{Y}$.

b. Compute the OLS estimator. Interpret the coefficient on education.

c. Calculate the fitted values and the residuals for each observation.

d. Verify that the orthogonality condition holds by directly computing $\mathbf{X}'\hat{\mathbf{e}}$. What does this result tell us about the relationship between the residuals and the regressors?

Empirical Exercise: Birthweight and Smoking

Use the `Birthweight_Smoking` dataset to answer the following questions.

a. Regress *Birthweight* on *Smoker*. What is the estimated effect of smoking on birth weight?

b. Regress *Birthweight* on *Smoker*, *Alcohol*, and *Nprevist*.

i. Explain why the exclusion of *Alcohol* and *Nprevist* could lead to omitted variable bias in the regression estimated in (a).
ii. Is the estimated effect of smoking on birth weight substantially different from the regression that excludes *Alcohol* and *Nprevist*? Does the regression in (a) seem to suffer from omitted variable bias?
iii. Jane smoked during her pregnancy, did not drink alcohol, and had 8 prenatal care visits. Use the regression to predict the birth weight of Jane's child.
iv. Compute R^2 and $\bar{R}^2$. Why are they so similar?
v. How should you interpret the coefficient on *Nprevist*? Does the coefficient measure a causal effect of prenatal visits on birth weight? If not, what does it measure?

c. Estimate the coefficient on *Smoking* for the multiple regression model in (b), using the three-step process following the Frisch–Waugh theorem (See Appendix 6.3 in Stock and Whatson). Verify that the three-step process yields the same estimated coefficient for *Smoking* as that obtained in (b).

d. An alternative way to control for prenatal visits is to use the binary variables *Tripre0* through *Tripre3*. Regress *Birthweight* on *Smoker*, *Alcohol*, *Tripre0*, *Tripre2*, and *Tripre3*.

i. Why is *Tripre1* excluded from the regression? What would happen if you included it in the regression?
ii. The estimated coefficient on *Tripre0* is large and negative. What does this coefficient measure? Interpret its value.
iii. Interpret the value of the estimated coefficients on *Tripre2* and *Tripre3*.
iv. Does the regression in (d) explain a larger fraction of the variance in birth weight than the regression in (b)?