Waseda University
School of Political Science and Economics

Homework 2

Daniel Fabio Groth

Econometrics, Fall 2024

Table of contents

Problem 1 **3**
- **Excercise 1** 3
- **Excercise 3** 3
- **Excercise 5** 3

Problem 2 **4**

Problem 3 **5**

Problem 1

Solve excercise 1,3 and 5 in Problem set 2.

Excercise 1

Show the following equalities hold:

1) $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X_n})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i(X_i - \bar{X}_n)$

Proof:

2) $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X_n})(Y_i - \bar{Y_n}) = \frac{1}{n}\sum_{i=1}^{n} X_i(Y_i - \bar{Y_n}) = \frac{1}{n}\sum_{i=1}^{n} Y_i(X_i - \bar{X_n})$

Proof:

Excercise 3

Consider a regression model that has no intercept term:

$$Y_i = X_i\beta_1 + \epsilon_i = 1, ..., n.$$

Derive the least squares estimator for β_1.

Excercise 5

Let $(\hat{\beta}_0, \hat{\beta}_1)$ be the ordinary least squares estimator of

$$Y_i = \beta_0 + X_i\beta_1 + \epsilon_i, i = 1, ..., n.$$

The prediction error (i.e, residual) for each i is given by $\hat{e_i} = Y_i - \hat{\beta}_0 n - X_i\hat{\beta}_1 n$. Show that the sum of the residuals is zero, i.e, $\sum_{i=1}^{n}\hat{e_i} = 0$.

Problem 2

Show that under Assumptions 1-3 in the L.6 slides, the variance of $\hat{\beta}_{n1}$ is given by $X_1, ..., X_n$ is given by:

$$\frac{\sigma^2}{n} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$$

Proof:

Problem 3

In this problem, you calculate the OLS estimators using R. Please obtain your own data by using the following code:

```
set.seed(34)

data <- as.data.frame(state.x77)
data <- data[sample(1:50, 40),]
```

where you need to input the last two digits of your student number for A. Here we use the information of the life expectancy as Y and the illiteracy rate as X. Then answer the following problems.

1) We consider the following two models.

Model 1: $Y_i = \beta_0 + X_i\beta_1 + \epsilon_i$

Model 2: $Y_i = X_i\beta_1 + \epsilon_i$

Obtain the OLS estimators for these two models **without using the lm() function** and compare the results with those given by the lm function.

2) For the two models, visually compare the distribution of the data and the lines obtained by OLS as we did in p.16 in the Lecture 6 slides. Discuss which results look more reasonable

3) Based on the "more reasonable" model you chose, explain what the estimated value of β_1 implies about the relationship between the illiteracy rate and the life expectancy.