Econometrics 1.1

M2 PPD-APE

Exercise 1 Lets *kids* denote the number of children ever born to a woman, and let *educ* denote years of education for the woman. A simple model relating fertility to years of education is:

$$kids = \beta_0 + \beta_1 educ + u$$

where u is the unobserved error.

1. What kind of factors are contained in u?
2. Are these likely to be correlated with level of education?

Exercise 2 In the simple linear regression model $y = \beta_0 + \beta_1 x + u$, suppose that $E(u) \neq 0$. Letting $\alpha = E(u)$, show that the model can always be rewritten with the same slope, but a new intercept and error, where the new error has a zero expected value.

Exercise 3 - Ordinary Least Squares Show that the ordinary least squares estimators are effectively those who minimize the expression

$$\sum_i \widehat{u_i}^2 = \sum_i (y_i - \widehat{\beta_0} - \widehat{\beta_1} x_i)^2.$$

1. Derive the first order conditions for the minimization
2. Check it is a minimum
3. Compare them to the normal conditions and conclude.

Exercise 4 The data Wages.xls provide some observations relating the number of years of education after baccalauréat (X) and the monthly wage (Y). We assume the model is

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

where $E(u|X) = 0$ and would like to estimate the coefficients using the sample.

1. Draw the scatterplot of the observations (using excel).

2. What is the formula for the OLS estimators of β_0 and β_1?
3. Compute them in the excel sheet [hint: fill in the columns].
4. Compute the predictions and the residuals for each observation.
5. Check the residuals approximately sum to 0.
6. Draw the regression line.

Exercise 5 We are interested in the relationship between infant birth weight, in kilos ($bwght$) and an explanatory variable, average number of cigarettes the mother smoked per day during the pregnancy ($cigs$). Based on data available on 1388 births in the US, the estimated relationship is:

$$\widehat{bwght} = 3.395 - 0.0145 cigs$$

1. What is the predicted birth weight when $cigs = 0$? What about when $cigs = 20$ (one pack per day)? Comment on the difference.
2. Does this simple regression necessarily capture a causal relationship between the child's birth weight and the mother's smoking habits? Explain.

Exercise 6 - Change in scale Assume a first model relating 2 variables y and x: $y_i = \beta_0 + \beta_1 x_i + u_i$.

1. We decide to change the scale in y. The new variable $\tilde{y} = y/10$ and the corresponding model is $\tilde{y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 x_i + \tilde{u}_i$. Express $\tilde{\beta}_0$ and $\tilde{\beta}_1$ as functions of β_0 and β_1.
2. Now, we decide to change the scale in x. The new variable $\tilde{x}^* = 10 \cdot x$ and the corresponding model is $y_i = \beta_0^* + \beta_1 x_i^* + u_i^*$. Express β_0^* and β_1^* as functions of β_0 and β_1.

Exercise 7 We assume a model where $Y_i = \beta_0 + \beta_1 X_i + u_i$. The data are the following:

Y_i	55	17	36	85	62	18	33	41	63	87
X_i	18	7	14	31	21	5	11	16	26	29

1. Determine the regression line.
2. What is the estimate of $E(Y_i|X_i = 21)$?
3. What is the difference between the observed value of Y when $X_i = 21$ and the estimated value with the regression line? How do you call this difference?