ØAMET4100 · Spring 2019
Worksheet 2A

Instructor: Fenella Carpena

January 17, 2019

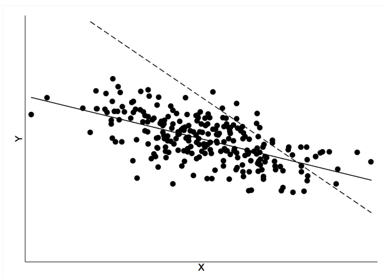
1 Population Regression vs. Sample Regression, Ordinary Least Squares (OLS) Estimation

Exercise 1.1 Consider the population regression line $Y_i = \beta_0 + \beta_1 X_i + u_i$ and the sample regression line $\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i$. Note that for the sample regression line, $\widehat{u}_i = Y_i - \widehat{Y}_i$. Explain the difference between $\widehat{\beta}_1$ and β_1; between the residual $\widehat{u}_i$ and the regression error u_i.

Exercise 1.2 Consider the population regression line $Y_i = \beta_0 + \beta_1 X_i + u_i$.

(a) What minimization problem does OLS solve in order to estimate β_0 and β_1? Express the problem mathematically (there is no need to solve the minimization problem).

(b) The figure below shows data from a sample of 250 observations of X and Y. One of the lines is the sample regression line, $\widehat{\beta}_0 + \widehat{\beta}_1 X_i$; the other is the population regression line $\beta_0 + \beta_1 X_i$. Is the sample regression line solid or dashed? Explain.



2 Interpreting OLS Regression Coefficients

Exercise 2.1 A regression of *wage* (hourly wage, measured in 1979 dollars per hour) and *educ* (years of schooling) using data from a random sample of 526 American workers yields the following:

$$\widehat{wage} = -0.90 + 0.54 \cdot educ$$

(a) Interpret the intercept of this regression.

(b) It turns out that all workers in the data have at least 8 years of education. Does this help reconcile your answer in part (a)?

(c) Interpret the slope of this regression.

(d) A worker has 16 years of education. What is the regression's prediction for the worker's hourly wage?

(e) Suppose a worker obtains 4 more years of schooling. What is the regression's prediction for the change in the worker's hourly wage?

3 Measures of Fit

Exercise 3.1 SER and R^2 are "measures of fit" for a regression. Explain how the SER measures the fit of a regression. What are the units of SER? Explain how R^2 measures the fit of a regression. What are the units of R^2?

Exercise 3.2 Sketch a hypothetical scatterplot of data for an estimated regression with $R^2 = 0.9$. Sketch a hypothetical scatterplot of data for a regression with $R^2 = 0.5$.

Exercise 3.3 Suppose that the R^2 from the regression in Exercise 2.1 is equal to 0.242. How would you interpret this R^2?

4 Least Squares Assumptions

Exercise 4.1. For each least squares assumption, provide an example in which the assumption is valid and then provide an example in which the assumption fails.

Exercise 4.2. A professor decides to run an experiment to measure the effect of time pressure on final exam scores. He gives each of the 400 students in his course the same final exam, but some students have 90 minutes to complete the exam while others have 120 minutes. Each student is randomly assigned one of the exam times based on the flip of a coin. Let Y_i denote the number of points scored on the exam by the i^{th} student (where $0 \leq Y_i \leq 100$). Let X_i denote the amount of time the student has to complete the exam (where $X_i = \{90, 120\}$), and consider the regression model $Y_i = \beta_0 + \beta_1 X_i + u_i$.

(a) Explain what the term u_i represents. Why will different students have different values of u_i?

(b) Explain why $\mathbb{E}[u_i|X_i] = 0$ for this regression model.

(c) Are the other least squares assumptions satisfied?

5 Simple Linear Regression in Stata

Exercise 5.1 The table below shows regression output from Stata.

```
Linear regression                                      Number of obs =        74
                                                       F(  1,     72) =     17.28
                                                       Prob > F      =    0.0001
                                                       R-squared     =    0.2196
                                                       Root MSE      =    2623.7

------------------------------------------------------------------------------
             |               Robust
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         mpg |  -238.8943   57.47701    -4.16   0.000    -353.4727    -124.316
       _cons |   11253.06   1376.393     8.18   0.000     8509.272    13996.85
------------------------------------------------------------------------------
```

Identify the following from the above regression output.

(a) Dependent and independent variables

(b) Sample size

(c) R^2

(d) SER

(e) $\widehat{\beta}_0$ and $\widehat{\beta}_1$