

Exercises 2: Introduction to Regression with R

Murshid Saqlain

Deadline February 12, 2019; msq@du.se

1. Simple regression with one regressor. Install the package *AER*. Load the data set *CPS1985* from this package. You can get information about the data by `?CPS1985`.
 - (a) To get a “picture” of the data obtain summary statistics of the variables (`summary()`). Use R to obtain the number of observations.
 - (b) Run a regression with *wage* as dependent variable and *education* as regressor (X-variable). Use the `lm()` function.
 - (c) What is the estimated average wage of an individual with 10 years of education according to the regression results?
 - (d) Use `summary()` on the regression object to obtain P-values and other information about the regression. What is the null and the alternative hypothesis that the P-value for education is based on?
 - (e) According to the regression results, is it a significant linear relationship between education and wage? Explain very shortly.
2. Simple regression with a dummy regressor. Continue to use the data set *CPS1985*.
 - (a) Run a regression with *wage* as dependent variable and *gender* as regressor (X-variable).
 - (b) According to the regression estimates, do women in the population earn less or more than men?
 - (c) Is the gender difference significant?
 - (d) Construct a (numerical) dummy variable coded 1 if female and 0 if male. Use this dummy in the regression instead of the factor variable *gender*. Do you get the same regression results?

- (e) It is possible to show that the sample average (\bar{Y}) is the least squares estimator of the population average. In the previous regression we have computed the least squares estimators of the averages of the populations for men and women. Compute the sample averages for both women and men. Compare the sample averages to the fitted values for men and women from the previous regression. Are the predicted averages from the regression equal to the sample averages?