Assignment 1

Basic Econometrics Fall 2022

Andreas Dzemski

Due **Sep 8, 18:00h**.
Only hand-ins following the submission guidelines will be accepted.
Starred (*) problems will not be graded.[1]

Problem 2

For this problem you will be using the dataset "cps.dta". This file contains data on average hourly earnings (in US Dollars) for 4733 working-age individuals.

1. Run the regression

$$\texttt{wage} = \beta_0 + \beta_1 \texttt{educ} + \beta_2 \texttt{exper} + U. \tag{a}$$

   Interpret the coefficient on `educ`.

2. Run the following code and explain what it does.

```
gen lwage = log(wage)
```

   Run the regression

$$\log(\texttt{wage}) = \beta_0 + \beta_1 \texttt{educ} + \beta_2 \texttt{exper} + U. \tag{b}$$

   Interpret the coefficient on `educ`.

3. Define a new variable `exper2` which is equal to the square of `exper`. Do you think that the marginal effect of `exper` on log(`wage`) is constant? Why might it make economic sense to include the quadratic term exper2 in the regression? Run the regression

$$\log(\texttt{wage}) = \beta_0 + \beta_1 \texttt{educ} + \beta_2 \texttt{exper} + \beta_3 \texttt{exper2} + U. \tag{c}$$

[1] Even though, we recommend you attempt to solve them. All problems will be discussed in the labs. To maximize your benefit from the discussion it is best if you have attempted to solve all problems yourself first.

What is the estimated (approximate) marginal effect of experience on log wages (= log(`wage`))? Plot the "estimated marginal effect of experience" against "experience" using the command `twoway function`. For which levels of experience is this marginal effect negative? Are any of the individuals in the sample in this experience range?

Problem 3

In this problem we continue to investigate regression (c).

1. Re-run regression (c) and find the standard error of the coefficient on `exper2` in the regression output.

2. The command

```
reg lwage educ exper exper2
predict U_hat, residuals
```

creates a new variable `U_hat`. What is stored in this variable? (*Hint*: You can use google or the Stata docs to find out). Run the following code and interpret the output.

```
sum U_hat
corr U_hat educ exper exper2
```

Explain your observations using your knowledge about the "mechanical" properties of OLS (*Hint*: Chapter 4.4 of the lecture notes).

3. Plot the empirical residuals against education. Is homoscedasticity a good assumption for the regression we are considering? Explain! Give an intuitive economic interpretation of your findings.

4. By default, Stata's `reg` assumes homoscedastic errors (i.e., Stata assumes that assumption OLS-5 holds). You can tell Stata to assume a heteroscedastic error term when it computes standard errors. You just have to run `reg` with the option `robust`.

```
reg lwage educ exper exper2, robust
```

Does using robust standard errors change the reported standard errors?

Problem 4

In this problem we continue to study regression (c). Use heteroscedasticity robust standard errors (i.e. use the `robust` option when you run a regression). Use significance level $\alpha = .05$.

1. Compute the numerical value of the test statistic for a two-sided t-test of the hypothesis

$$H_0 : \beta_3 = 0$$

from the estimated coefficient and its standard error. Where is this test statistic reported in the regression output? What is the economic interpretation of the null hypothesis? What is the critical region of the two-sided t-test for $\alpha = .05$ (*Hint*: use the Stata function `invnormal` as explained in the lecture notes). Does the test reject? Interpret the result. Is the variable `exper2` statistically significant in this regression?

2. When you run a regression, Stata automatically reports a confidence interval at confidence level $(1 - \alpha) = .95$. What is the reported confidence interval for the coefficient on education? Dropping the `robust` option changes the reported confidence interval. Why?

3. (*) Compute the confidence interval for the coefficient on education with confidence level $1 - \alpha = .90$.

Problem 5

We consider the regression

$$\begin{aligned} \log(\texttt{wage}) =& \beta_0 + \beta_1 \texttt{educ} + \beta_2 \texttt{exper} + \beta_3 \texttt{exper}^2 \\ &+ \beta_4 \texttt{midwest} + \beta_5 \texttt{west} + \beta_6 \texttt{south} + U. \end{aligned} \tag{d}$$

For this problem use significance level $\alpha = .05$.

1. Run the regression without robust standard errors. Interpret the estimates for the coefficients β_4, β_5 and β_6.

2. What is the null hypothesis H_0 for testing the joint significance of the regional dummies `midwest`, `west` and `south`? What is the economic content of this hypothesis?

3. We can test H_0 using an F-test. If H_0 is true, then the appropriate F-statistic has an F-distribution. What are its numerator and denominator degrees of freedom? The critical value for $\alpha = .05$ is given by

```
di invF(df1, df2, 0.95)
```

(replace df1 by the numerator degrees of freedom and df2 by the denominator degrees of freedom).

4. Test the null hypothesis H_0 in Stata by running

```
test midwest west south
```

immediately after running regression (d). What is the realized test statistic $\hat{F}(\omega_0)$? Does the realized test statistic exceed the critical value computed above? Are the regional dummies `midwest`, `west` and `south` jointly significant?

5. Manually compute the p-value of the test (*Hint*: Use the Stata function `F` in display mode.) Suppose that we observed $\hat{F}(\omega_0) = 1.8$. What is the p-value corresponding to this value of the test statistic?

6. Use the command `test` to test the null hypothesis that

$$H_0 : \beta_1 = \cdots = \beta_6 = 0.$$

This test is called the "test of model significance". Interpret the the null hypothesis. What is the test result?

What have you learned?

Econometric skills: After working through this problem set, you can ✎ ☒

- ☐ model nonlinearities by transforming the outcome variable/including additional regressors.
- ☐ model differences between groups using dummy variables
- ☐ conduct and interpret a two-sided t-test
- ☐ heuristically assess the plausibility of the homoscedasticity assumption
- ☐ test the joint significance of multiple regressors and use such a test to test empirical claims about differences between groups
- ☐ test significance of the regression model.

Programming skills: After working through this problem set, you can ✎ ☒

- ☐ define a new variable by computing it from existing variables
- ☐ plot functions.