

The International College of Economics and Finance
Econometrics-2021-2022

Home assignment 3. Dummy Variables
To be submitted by November 07, 23:55

1. [30 marks] A researcher investigating regional variations in wages uses data on 1,000 individuals in a U.S. survey to fit the following regressions (standard errors in parentheses; *RSS* is residual sum of squares):

$$\hat{EARN} = -5.49 + 1.40S \quad \text{RSS} = 65.50$$

(2.00) (0.15)

$$\hat{EARN} = -5.43 + 1.34S - 0.26NC + 3.28NE + 1.29W \quad \text{RSS} = 64.14$$

(2.03) (0.15) (0.83) (0.97) (1.00)

where *EARN* is hourly earnings in U.S. \$, *S* is the number of years of schooling of the individual, and *NC* (north central), *NE* (north east) and *W* (west) are regional dummy variables for the residence of the individual (south is the reference category).

1.1. [10 marks] ☐ Explain how the dummy variables are defined here.

☐ Explain the meaning and the role of the coefficients.

☐ Represent the second equation graphically.

1.2. [10 marks] ☐ Do the significance tests for the coefficients of the dummy variables.

☐ Do a test of the explanatory power of the dummy variables as a group.

1.3. [10 marks] ☐ Suppose the researcher had taken the western region as the reference category. What differences would this have made to the second regression?

☐ Suppose that the researcher wished to investigate whether the impact of schooling on earnings was different in the four regions. How could this be done?

2. [20 marks] A researcher has observations on *EARN*, earnings measured in \$ per hour, *S*, years of schooling, and *ASVABC*, the score on a test of cognitive ability (here in %), for 2,868 individuals in the United States in 1994. She also defines dummy variables *MALE* and *UNION*. *MALE* is equal to one if the individual is male and 0 if she is female. *UNION* is equal to 1 if the individual belonged to a union in 1994, 0 otherwise. She runs the following regressions, whose results are summarized in the table below:

(1) a regression of the logarithm of earnings on *S*, *ASVABC*, and *MALE*, using the entire sample

(2) the same specification, for those individuals who belonged to a union

(3) the same specification, for those individuals who did not belong to a union

(4) the same specification with the dummy variable *UNION* added, for the entire sample.

The table gives the regression coefficients. Standard errors are given in parentheses, *RSS* is the residual sum of squares, and *n* is the number of observations in the regression. Dependent Var.: Logarithm of Hourly Earnings

	1: whole sample (2868)	2: union only (286)	3: non-union only (2582)	4: whole sample (2868)
<i>S</i>	0.066 (0.004)	0.028 (0.012)	0.070 (0.005)	0.066 (0.004)
<i>ASVABC</i>	0.013 (0.001)	0.011 (0.003)	0.013 (0.001)	0.013 (0.001)
<i>MALE</i>	0.214 (0.017)	0.286 (0.049)	0.199 (0.018)	0.209 (0.017)
<i>UNION</i>	-	-	-	0.189 (0.028)
<i>Constant</i>	0.819 (0.055)	1.545 (0.164)	0.750 (0.058)	0.803 (0.055)
<i>R</i> ² / <i>RSS</i>	0.249 / 588.3	0.195 / 43.7	0.260 / 522.5	0.261 / 579.7

2.1. [10 marks] Give an economic interpretation of the first equation.

2.2. [10 marks] ☐ Determine whether the earnings functions are different for union members and non-members

(i) using a Chow test

(ii) using a *t* test on the coefficient of the union dummy variable.

☐ Explain the relationship between these two approaches to testing for differences in the earnings functions for union and non-union workers. Is it possible that they might lead to different conclusions?

3. [50 marks] For your data set **ha03_data**__ explore whether earnings of respondents are different for different categories defined by several dummy variables - **MALE**, **FEMALE**, **MARRIED** (refer to **ha03_data description.pdf**).

3.1. [10 marks] Generate variable **H=S-12** and run simple regression model **EARNINGS** on **H** and compare it with regression **EARNINGS** on **H** and **MALE**.

- ☐ Explain why coefficients of **H**, intercepts and other characteristics of these regressions are different. Give a graphical illustration of the equations being investigated.
- ☐ Is there a discrimination of the workers in earnings on the base of gender?
- ☐ What would change if instead of regression **EARNINGS** on **H** and **MALE** the regression **EARNINGS** on **H** and **FEMALE** is estimated?

3.2. [10 marks] ☐ Is discrimination in earnings based on gender confirmed by the Chow test? Do Chow test both manually (on the base of subsamples) and automatically (using command).

- ☐ What is the relation of the test performed in 3.1 to the Chow test in 3.2?
- ☐ What kind of test is equivalent to the Chow test in the context of this question?

3.3. [10 marks] Repeat Chow test and equivalent test for the group of dummy variables, but now on the basis of multiple regression equation **EARNINGS** on **H**, **ASVABC** and **AGE**.

3.4. [10 marks] Taking as a basis a simple regression of **EARNINGS** on **H**, consider the simultaneous effect of gender (variable **FEMALE**), family status (variable **MARRIED**), and place of residence (variable **URBAN**).

- ☐ Having calculated the regression with inclusion of three dummy variables, estimate their significance and joint significance, and also offer an explanation of the obtained signs of coefficients.
- ☐ Having calculated the regression of **EARNINGS** on **H**, **FEMALE**, **MARRIED** and the interaction dummy **FEMALE*MARRIED**, give full interpretation to all its coefficients.

3.5. [10 marks] ☐ Taking the regression of **EARNINGS** on **H**, investigate the presence of discrimination on the basis of ethnicity (**ETHBLACK**, **ETHHISP**, **ETHWHITE** variables).

- ☐ Explore the impact of the change of reference category on the regression results. What is new in comparison with the similar question in 3.1?

If you have any questions please ask at Vladimir.tcherniak@gmail.com