THIRD ASSIGNMENT IN MICROECONOMETRICS

onoma , A.M.:
onoma, A.M.:
onoma, A.M.:
onoma, A.M.:

1   Introduction

In this report, we will provide an overview of the basic statistical measures of some variables. We subsequently illustrate suitable graphical displays (histograms). Then we develop a theoretical linear regression model that examine the linear relationship between a dependent variable (hourly wage) and two (Multiple Linear Regression) independent variables (years of schooling and age) and this will be followed by two regression analysis using the Ordinary Least Squares (OLS) method.

2   Related Literature and Theoretical Frame

The objective in paper of Thomas Lemieux from 2003 was to determine the effect of education on the wage of worker, for which a quantitative research approach was applied. By studying the relation between the variables, they used the Mincer linear model, and they observed the positive relation between professional training and wage. Moreover, they show that effect correlates with the remaining years of working until retirement, but the older employees haven't sufficiently motivation.

3   Data

Our study used data collected by the Statistical Service of the country of Ano Magoula. The data refer to a specific time period, in 2021 and the data set contains 21 variables and 753 observations. Specifically, there are some variables which refer to women, such as hours (wife's hours of work in 2021), age (wife's age), educ (wife's educational attainment, in years), wage (wife's average hourly earnings, in 2021 dollars), repwage (wife's wage reported at the time of the 2022 interview), motheduc (wife's mother's educational attainment, in years), fatheduc (wife's father's educational attainment, in years), exper (actual years of wife's previous labor market experience) and variable exper which is the squared value of the previous one. Then, concerning men, there are a variable hushrs which measure husband's hours worked in 2021, a variable husage(husband's age), huseduc (husband's educational attainment, in years), huswage (husband's wage, in 2021 dollars). Also, childt6 captures the number of children who are less than 6 years old in household and childge6 the number of children between ages 6 and 18 in household. There is a variable named city and indicates the size of the city and variables faminc and nwifeinc which are linked to family income, in 2021 dollars and non-wage family income (= [faminc – (wage*hours)/100]). Moreover, a variable inlf is an indicator oh the labor force

participation. Finally, there are variables called mtr and unem which present the marginal tax rate and the unemployment rate in county of residence (in percentage points), respectively.

4 Questions

4.1 Question 1

Table 1: Statistic Table of Ano Magoula

Variable	Obs	Mean	Std.dev.	Min	Max
inlf	753	0.5683931	0.4956295	0	1
hours	753	740.5764	871.3142	0	4950
kidslt6	753	0.2377158	0.523959	0	3
kidsge6	753	1.353254	1.319874	0	8
age	753	42.53785	8.072574	30	60
educ	753	12.28685	2.280246	5	17
wage	428	4.177682	3.310282	0.1282	25
repwage	753	1.849734	2.419887	0	9.98
hushrs	753	2267.271	595.5666	175	5010
husage	753	45.12085	8.058793	30	60
huseduc	753	12.49137	3.020804	3	17
huswage	753	7.482179	4.230559	0.4121	40.509
faminc	753	23080.59	12190.2	1500	96000
mtr	753	0.6788632	0.0834955	0.4415	0.9415
motheduc	753	9.250996	3.367468	0	17
fatheduc	753	8.808765	3.57229	0	17
unem	753	8.623506	3.114934	3	14
city	753	0.6427623	0.4795042	0	1
exper	753	10.63081	8.06913	0	45
nwifeinc	753	20.12896	11.6348	-0.0290575	96
expersq	753	178.0385	249.6308	0	2025

Provided by microdataset1 from Stata

In this table we show the summary statistics of the variables of the data set. First of all, we observed significant differences between the genders. Specifically, the wife's wage is lower than husband's per 3,3$/h. In addition, the mean education of women is not significantly lower than men but the average of men works three times (hushrs) as much as the average women (hours).

4.2 Question 2

In this question, the tables display the 10th, 25th, 50th, 75th, and 90th percentiles of the following variables: wage, education, age and experience. To rephrase this, it's the percentage of data that falls at or below a certain observation.

4.3 Question 3

Now, we continue the analysis by constructing histograms for each of the five variables wage, educ, age, hours and faminc.

```
Type: Numeric (double)

          Range: [.1282,25]              Units: .0001
  Unique values: 373                 Missing .: 325/753

           Mean: 4.17768
      Std. dev.: 3.31028

    Percentiles:    10%     25%     50%     75%     90%
                  1.4815  2.2584  3.4819  4.9726  7.6218
```

Figure 1: Provided by microdataset1 from Stata

Firstly, as far as the hourly wife's wage, we can see that the 10th percentile is 1.4815 dollars which means that 10% of women got less than 1.48dollars/hour and remaining 90% got more than 1.48 dollars/hour. In addition, the 25th percentile earns up to 2.25 dollars, the 75th up to 4.97 dollars , and the 95th up to 7.62 dollars per hour. Also, the median that we are requested to find and explain is the 50% percentile. We can see that the median of wage is 3.4819 wife's average hourly earnings, in 2021 dollars.

```
Type: Numeric (byte)

          Range: [5,17]                 Units: 1
  Unique values: 13                  Missing .: 0/753

           Mean: 12.2869
      Std. dev.: 2.28025

    Percentiles:    10%     25%     50%     75%     90%
                     10      12      12      13      16
```

Figure 2: Provided by microdataset1 from Stata

The distribution of women's educational attainment is shown in the table above. As we can see the sample's median level of education is 12 years. Also, up to 10 years of education are seen in the 10th percentile, 12 years in the 25th percentile, 13 years in the 75th percentile, and 16 years in the 90th percentile. The sample's median level of education is 12 years.

Figure 3: Provided by microdataset1 from Stata

This figure presents a histogram of wife's average hourly earnings, in 2021 dollars. The biggest concentration of the graph distinguishes on the left part of figure, indicating a right skewed distribution. We can note that the wife's average hourly earnings, which were more frequent, had the range of 2 to 4 dollars per hour.
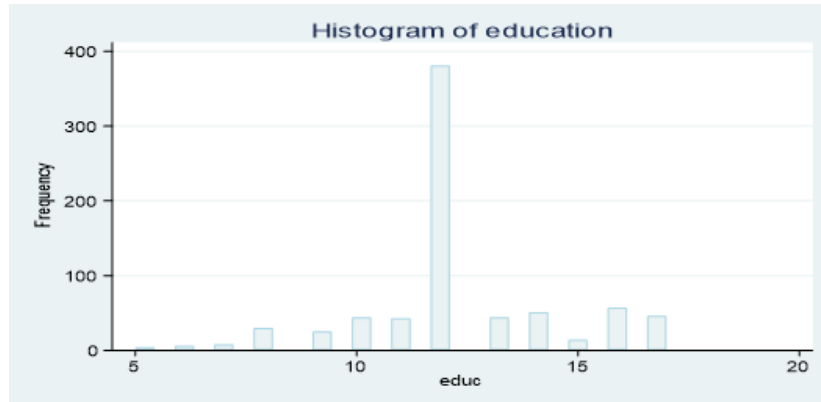
Figure 4: Provided by microdataset1 from Stata

As displayed on the above histogram the 50th percentage of wife's educational attainment fluctuates between 11 and 13 years whereas the rest of them have lees than 11 years and more than 13 years of education.
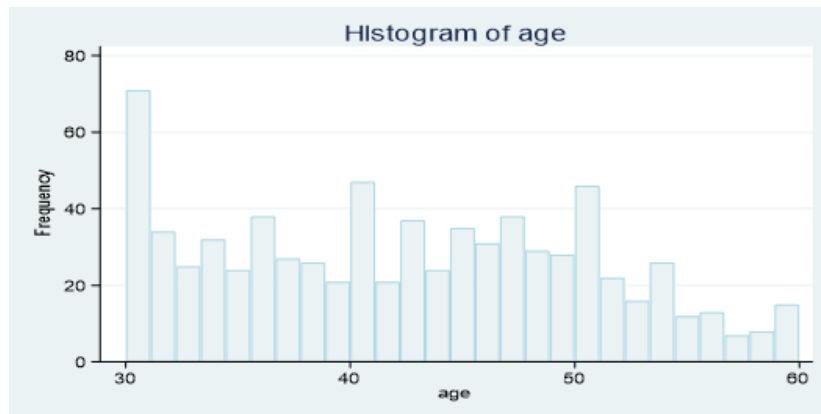

Figure 5: Provided by microdataset1 from Stata

Here we notice that the histogram is divided in 3 parts, wifes who are between ages of 30 and 40, those from 40 up to 50 and those from 50 up to 60. The most frequent group of wifes is those who are 30 years old and the less frequent is those who are close to 60 years old. The ages in between seem to be almost equally distributed.
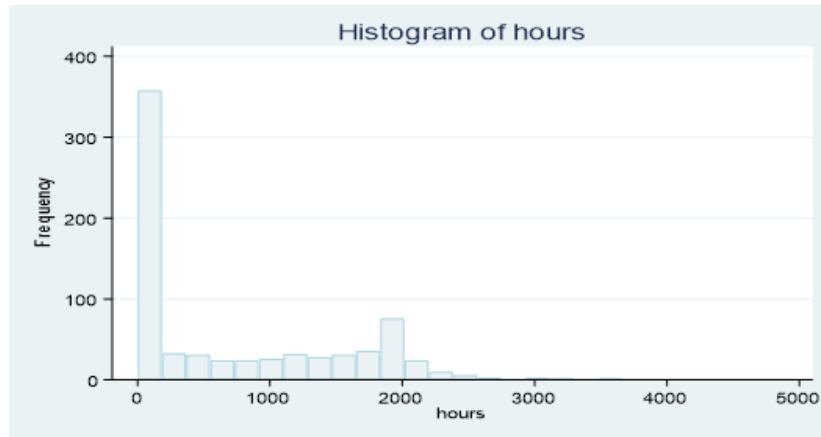
Figure 6: Provided by microdataset1 from Stata

We observe a peak at point 0. This happened because there is a great percent of women who are unemployed during 2021 according to our data.
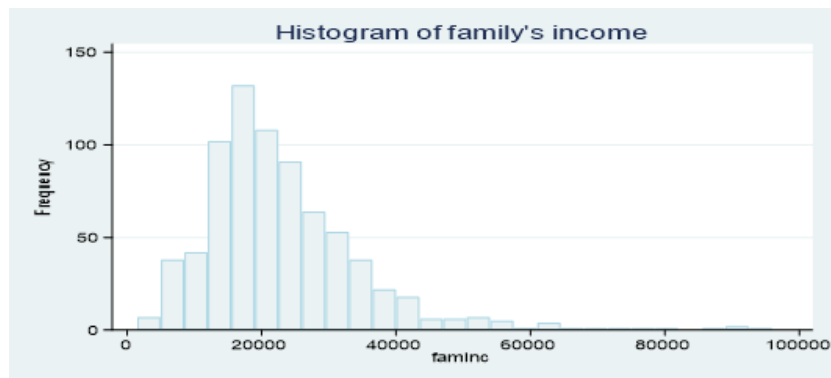

Figure 7: Provided by microdataset1 from Stata

The family income is depicted in the above chart by a frequency histogram. It has a peak, and it is almost symmetrical. We can observe that the highest frequency is close to 20.000 dollars, where you also find the highest class.

4.4 Question 4

By studying the article, we observed that education has a positive relationship in education and wage, which it due to the increasing marginal productivity of labor associated with higher education. So, they define and calculate the relationship between dependent variable wage, the education and age of workers (independent variables). So, we quote a simple linear model of the effects of age and education:

$$y \quad = \quad \beta_0 \quad + \quad \beta_1 X_1 \quad + \quad \beta_2 X_2 \quad + \quad u_t \ (1)$$

y = dependent variable

β_0 = fixed term

β_1 = regression coefficient

β_2 = second regression coefficient

u_t = disruptive term

In the data set, Ano Magoula 2021, the sample separates to women and men. So, we recreate the linear model,and we replace the dependent variable with wage of workers (men and women) and the independent variables with education and age for each gender. More specifically:

$$Wage \quad = \quad \beta_0 \quad + \quad \beta_1 educ \quad + \quad \beta_2 age \quad + \quad u_t \ (2)$$
$$Huswage \quad = \quad \beta_0 \quad + \quad \beta_1 huseduc \quad + \quad \beta_2 husage \quad + \quad u_t \ (3)$$

Mincer Econometric Model Approach

The Mincer model presents the dependent variable to logarithm and the independent variables to the square of it. In this data set, Ano Magoula 2021, we create two relations between women and men of sample.

$$LnWage \quad = \quad \beta_0 \quad + \quad \beta_1 educ \quad + \quad \beta_2 age^2 \quad + \quad u_t \ (4)$$
$$LnHuswage \quad = \quad \beta_0 \quad + \quad \beta_1 huseduc \quad + \quad \beta_2 husage^2 \quad + \quad u_t \ (5)$$

4.5 Question 5

Let's perform a regression analysis using the variables wage and educ. We expect that better educational attainment would be associated with higher earnings. Below, we show the Stata command for testing this regression model followed by the Stata output.

4.6 Question 6

Now, we regress the dependent variable, wage, on two predictor variables in the data set, educ and age.

```
Linear regression                          Number of obs    =      428
                                           F(1, 426)        =    53.78
                                           Prob > F         =   0.0000
                                           R-squared        =   0.1169
                                           Root MSE         =   3.1144
```

wage	Coefficient	Robust std. err.	t	P>|t|	[95% conf. interval]	
educ	.4953084	.0675396	7.33	0.000	.3625561	.6280607
_cons	-2.092367	.8243551	-2.54	0.011	-3.712677	-.4720574

Figure 8: Provided by microdataset1 from Stata

The value of coefficient is equal to 0.4953 and this means that if education be increased by 1 year (marginal change) then, the wage is going to be increased by 0.4953 dollars per hour. Also, the positive sign at the coefficient indicates a positive relation between the variables of wage and education — which is what we would expect. Now, the most common hypothesis test in econometrics is the t-test of the null hypothesis that a coefficient equals zero. We observe that P > |t| (p-value) = 0.000. So, choosing a significance level of 0.05 for our test of slope means that we reject the hypothesis of a zero slope, and the variable of education is statistically significant at the specific level. In addition, the R-squared of .1169 means that approximately 11% of the variance of wage is accounted for by the model, in this case, educ.

```
Linear regression                          Number of obs    =      428
                                           F(2, 425)        =    28.69
                                           Prob > F         =   0.0000
                                           R-squared        =   0.1193
                                           Root MSE         =   3.1139
```

wage	Coefficient	Robust std. err.	t	P>|t|	[95% conf. interval]	
educ	.4989633	.0675659	7.38	0.000	.3661583	.6317683
age	.0207362	.020897	0.99	0.322	-.0203382	.0618106
_cons	-3.008975	1.134387	-2.65	0.008	-5.238524	-.7794255

Figure 9: Provided by microdataset1 from Stata

Let's examine the output from this multiple regression analysis, focus on whether they are statistically significant and, if so, the direction of the relationship. As with the previous regression, we look to the p-value of the F-test to see if the variable of age is also significant. We can see that if we choose a 0.05 level of significance, the p-value (=0.322) is larger than this level, meaning that the regression coefficient for age is not significantly different from zero. We can confirm the above by comparing in absolute value the results of t with the value of the critical value index we derive from the Pearson. So, as we observe that the t value is lower than the critical value, the null hypothesis is not rejected while the coefficient of the variable age is statistically insignificant. In addition, the R-squared is 0.1193 and it shows that the independent variable explain 11.93% of the total wage variability.

Now, we will study the relationship between the variable of wage, the education of workers, the age of them and the square of age, which represents the polynomial of second degree for the effect of age on wages. In order to determine if there is a direct, or positive, relationship between them, we can calculate a statistical significance test, same as before. So, we conclude that none of the variables are statistically significant except for the variable wage.

```
Linear regression                              Number of obs  =      428
                                               F(3, 424)      =    19.23
                                               Prob > F       =   0.0000
                                               R-squared      =   0.1193
                                               Root MSE       =   3.1174

             │              Robust
        wage │ Coefficient  std. err.     t    P>|t|     [95% conf. interval]
─────────────┼────────────────────────────────────────────────────────────────
        educ │  .4991781    .0677566     7.37   0.000     .3659976    .6323587
         age │  .062309     .2402102     0.26   0.795    -.4098421    .5344601
        age2 │ -.0004851    .0029162    -0.17   0.868    -.0062171    .005247
       _cons │ -3.873235    4.946489    -0.78   0.434    -13.59593    5.849458
```

Figure 10: Provided by microdataset1 from Stata