

## LABOR ECONOMICS PROJECT

### INSTRUCTIONS

Economics 5850 w/ Dr. Doetsch  
Due December 3, 2018 at 2:20 pm

Late submissions are NOT accepted.

This assignment is intended to give you experience in researching, analyzing, organizing, interpreting, evaluating, and presenting real-world data and applying economic theory and analytical techniques to real-world labor markets. It is highly suggested that you work on it continually throughout the semester as we cover relevant topics.

The work you produce must be your own. **Substantial collaboration with other students constitutes academic misconduct. Sharing statistical results with other students constitutes academic misconduct.**

You are to answer the following questions using an unweighted sample from the 2016 American Community Survey (ACS)<sup>1</sup>. These microdata are available for download in the "Files" section of this course's Carmen page.

The sample is a 1/1,000 sample ( $n = 28,285$ ) of persons in the noninstitutionalized working-age population (age 18-64) residing in Ohio, Michigan, Pennsylvania, Indiana, Kentucky, or West Virginia. It is contained in files by the name of *ProjectSample2019* on Carmen. Each observation is a real, live human being living in Ohio or a neighboring state in 2016 and each column is a different variable. Files are available in .csv (comma separated value), .dta (Stata), and .gdt (Gretl) formats on Carmen. You can import .csv files into R.

The codebook for this data is in the appendix to this assignment. Another appendix describes using binary & categorical variables in your analyses. You should review and understand each appendix before you begin.

---

<sup>1</sup> For simplicity, the sample you are receiving is unweighted. This means that it is not representative. For the sake of this assignment, assume that it is. Your estimates will be biased, so take your results with a (small) grain of salt. Because your sample is unweighted, you will produce slightly different numbers than any official BLS reports.

You should be familiar with running regressions and using statistical analysis software from the pre-requisites for this course. If you are rusty with regressions or displaying data, you may avail yourself of multiple online tutorials.

Tables are graded on presentation and clarity. Every table should have an appropriate title. There should be labels on rows and columns as appropriate. A good table attempts to convey the most information with the least number of cells, columns, and rows. Tables should be intuitive and easily read. A reader should be able to understand what a table presents from reading the table itself. Carefully think about and plan out what your table should look like and how it should be effectively organized before you create it.

Do not submit raw regression outputs; instead, make a nice table. For regression report tables, please include the relevant t-statistic in parentheses under the coefficient. Report every coefficient, every t-statistic,  $R^2$ , adjusted- $R^2$ , sample size, and any sample restrictions. If you create new variables, please define them. Regressions are graded on presentation, clarity, and application of theory.

Discussions should run the indicated word count. They should reflect your grasp of labor economic theory and basic empirical results. You should be able to judge which information in each table or results from each analysis are interesting and worth commenting on. Try to include as much important detail as you can, given the small amount of space. Discussions will be graded on relevance, insight, and application of theory.

The text of your report must be double-spaced in 12-point professional font with one-inch margins. Use a single staple in the upper left-hand corner to attach multiple sheets of paper together. Present your labeled answers to each section in order, please.

Do not turn in a copy of these instructions. Do not repeat the questions in the text of your answers.

If you create any new variables (e.g. natural log transformations, interactions, binaries) from the ones you are given, please specify the details on how you created them and include them in any summary statistics.

Please keep in mind that part of the learning process is the frustration of figuring out how to do things yourself. Also note that there are often no clear answers on how to best approach the various little questions that come up during an analysis. Do your best and explain what you do when such situations arise.

## 1) EXECUTIVE SUMMARY

### 1A) Table: Summary Statistics

Produce a table of means, standard deviations, minimums, and maximums for every variable that you use in the foregoing analysis. You may wish to create this table last.

### 1B) Table: Labor Market Statistics

Produce a table of estimates of salient aggregate labor market statistics (mean earnings, labor force participation rates, unemployment rates, employment-population ratios) by state.

### 1C) Discussion: Executive Overview & Summary

Consider your results from 1B. Compare and contrast the aggregate labor market of Ohio with that of its neighboring states. How do conditions in the Ohio labor market compare? How do these states compare to the country as a whole (recall this is 2016 data)? Do you believe these numbers are satisfactory to make cross-state comparisons of labor market conditions? Explain. (300 words)

## 2) LABOR FORCE PARTICIPATION

### 2A) Table: Labor Force Participation

Using the data provided, produce one table of statistics presenting the following:

- a) Labor force participation rate by sex
- b) Labor force participation rate by sex and marital status
- c) Labor force participation rate by sex and educational attainment (dropout, high school, some college, college, postgrad)

### 2B) Regression: Labor Force Participation

Use an OLS linear probability model to model the effects of correlates on labor force participation and how they differ by sex. Run three regressions: one including both sexes, one for men, and one for women.

Dependent variable: *lfp*

Independent variables: Use your judgment, guided by theory  
Sample restrictions: None

### 2C) Discussion: Labor Force Participation

Consider your results from 2A and 2B. Discuss the differences in the labor force participation rate of working-age people of different demographic groups and different individual characteristics for this sample. Does economic theory adequately explaining these differences in labor force participation? Did you expect to find what you did here? Is anything noteworthy? Interpret any interesting coefficients. Are there any important omitted variables that may introduce bias into your estimates? Discuss any weaknesses in your analysis. Explain. (300 words)

### 3) HUMAN CAPITAL & EARNINGS

#### 3A) Table: Wage and Salary Income by Educational Attainment

Using the provided data, produce one table presenting the following statistics for the sample for each of the five educational attainment categories (dropout, high school, some college, college, postgrad):

- a) Mean yearly wage and salary income of employed workers
- b) Mean weekly hours of employed workers
- c) Employment-population ratio
- d) Unemployment rate
- e) Mean age
- f) Mean marriage rates
- g) Mean children
- h) Nativity status composition
- i) Sex composition
- j) Racial composition

#### 3B) Regression: The Determinants of Earnings

Use a Mincerian Wage equation (OLS model) to model the correlates of yearly labor market earnings.

Dependent variable: *earnings* or *lnearnings*

Independent variables: Use your judgment, guided by theory

Sample restrictions: Employed persons only

#### 3C) Calculation: Earnings Predictions

Use your coefficient estimates and values from your personal characteristics<sup>2</sup> to predict your personal earnings at age 40 in 2106 \$ with the 2016 coefficients. Provide a brief table explaining the values used and showing your work. Briefly discuss if you think your prediction is accurate. (50 words)

#### 3D) Discussion: Earnings Determination

Consider your analyses from 3A, 3B, and 3C regarding the value of human capital. Did you find what you expected to find? Is anything noteworthy? Discuss your findings and any weaknesses of your analysis. (300 words)

---

<sup>2</sup> You will have to use your best guess for what your personal characteristics by age 40 will be. Be realistic, though (the data can guide you here). This, of course, assumes you'll live in Ohio or a neighboring state.

#### 4) WORK HOURS

##### 4A) Table: Usual work hours per week

Produce one table of summary statistics (means, standard deviations of variables) for employed persons, comparing usual work hours per week by sex, race, and education level.

##### 4B) Regression: Determinants of Work Hours

Use a simple OLS model to model the correlates of usual weekly hours.

Dependent variable: *hours* or *lnhours*

Independent variables: Use your judgment, guided by theory

Sample restrictions: Employed persons only

##### 4C) Discussion: Work Hours

Consider your results from 4A and 4B. Do substitution or income effects dominate? Do your results confirm what one expects from economic theory? Did you find anything noteworthy? Explain your reasoning and explain any weaknesses in your analysis. (300 words)

## 5) SEX DISCRIMINATION

### 5A) Table: Sex Comparisons

Produce one table of summary statistics (means, standard deviations of variables) for the sample comparing employed males and employed females on the following variables:

- a) Wage and salary income by sex
- b) Educational attainment by sex
- c) Wage and salary income by sex and educational attainment

### 5B) Oaxaca Decomposition by Hand: Men and Women

Estimate two Mincerian OLS wage equations restricted to males and females, respectively.

Dependent variable: *lnearnings*

Independent variables: use your judgment, guided by theory

Sample restrictions: Employed persons only

Using the Oaxaca Decomposition technique and your coefficient estimates, calculate by hand (aided by Excel/a calculator/whatever) how much of the log male-female earnings gap is "explained" and "unexplained"<sup>3</sup>. Report your findings.

**Show your work.**

### 5C) Discussion: Sex Discrimination

Consider your analyses from 5A and 5B regarding male-female differences in labor market outcomes and earnings. Interpret your results. Discuss your findings and any weaknesses of these analyses. What variables have you left out that are important? How do you think their inclusion would affect your conclusions? (300 words)

---

<sup>3</sup> Use as many decimal places as you can in all your calculations for better numbers. It makes a very large difference.



## 6) WAGE INEQUALITY

### 6A) Table: Wage Structure Summary Statistics

Break up the sample of employed persons into four earnings quartiles. The first quartile is the lowest 25% of earners. The second quartile is the next 25% (25<sup>th</sup> to 50<sup>th</sup> percentiles). The third quartile is the next 25% (50<sup>th</sup> to 75<sup>th</sup> percentiles). The fourth quartile is the top 25% of earners (above 75<sup>th</sup> percentile).

Build a table of summary statistics by quartile, showing inter-quartile differences in characteristics like earnings, age, sex, race, education, hours, and any other variables you think pertinent or interesting.

### 6B) Discussion

Consider your table in 6A. Discuss what you find interesting. Discuss your findings and any weaknesses in your analysis. Explain. (300 words).

## 7) CHOOSE YOUR OWN ADVENTURE

### 7A) Table and/or Regression

Produce an interesting question that may be answered from this data. Creative, challenging, interesting, or rigorous questions will earn comparatively more points. Use the data to answer your question with a table and/or regression. Explain what you do and why.

### 7B) Discussion

Discuss your answer to your question from 7A. Discuss your findings and any weaknesses in your analysis. Explain. (300-600 words).

## APPENDIX

## CODEBOOK

Variable Name	Type	Description
<i>age</i>	Continuous	Person's age in years
<i>age2</i>	Continuous	Age squared
<i>asian</i>	Binary	=1 if person identifies as Asian racial category
<i>black</i>	Binary	=1 if person identifies as African-American/black racial category
<i>children</i>	Continuous	Number of own children in household
<i>college</i>	Binary	=1 if educational attainment is a Bachelor's degree
<i>dropout</i>	Binary	=1 if educational attainment is less than high school or equivalent
<i>earnings</i>	Continuous	Yearly labor market wage and salary earnings (2016 \$)
<i>employed</i>	Binary	=1 if person is currently employed
<i>female</i>	Binary	=1 if person is female
<i>hispanic</i>	Binary	=1 if person identifies as Hispanic ethnic category. Note: this is an ethnic category; one can be Hispanic of any race
<i>hours</i>	Continuous	The number of hours usually worked per week
<i>hsonly</i>	Binary	=1 if educational attainment is high school graduation, GED, or equivalent
<i>immigrant</i>	Binary	=1 if the individual was foreign-born
<i>lfp</i>	Binary	=1 if person participates in labor force (employed or unemployed)
<i>lnearnings</i>	Continuous	The natural log of earnings
<i>lnhours</i>	Continuous	The natural log of hours
<i>lnotherincome</i>	Continuous	The natural log of other income
<i>male</i>	Binary	=1 if person is male
<i>married</i>	Binary	=1 if person is currently married
<i>native</i>	Binary	=1 if person was born in USA or territories
<i>notinlfp</i>	Binary	=1 if person is not in the labor force
<i>otherincome</i>	Continuous	Includes all non-labor income: income (earned and unearned) from other family members, interest, business income, dividends, capital income, rents, government transfers, etc. (2016 \$)

<i>otherrace</i>	Binary	=1 if the individual identifies as American Indian, biracial, multiracial, or other racial category
<i>postgrad</i>	Binary	=1 if educational attainment is a Master's degree, PhD, or professional degree (e.g. MD, JD)
<i>somecollege</i>	Binary	=1 if educational attainment is Associate's degree, trade school, or some college (no completed degree)
<i>statefip</i>	Categorical	A categorical variable indicating state of residence. Codes are available at this <a href="#">[link]</a> .
<i>unemployed</i>	Binary	=1 if the individual is unemployed (seeking work)
<i>vet</i>	Binary	=1 if the individual is a veteran of US military service (during wartime or peacetime)
<i>white</i>	Binary	=1 if person identifies as white racial category
<i>yearsinus</i>	Continuous	If a person is foreign-born, how many years have they been in the US. Caution: value is 0 if native born!

## APPENDIX

### BINARY & CATEGORICAL VARIABLES HELP

#### Mutually exclusive binary variables

Some of the binary variables above are mutually exclusive because they fall into categories. You cannot include all variables within each group in your regressions because of perfect collinearity. Hence, you must choose a base group, leave it out of your regressions, and interpret binary in reference to that base group. Below is a list of all such cases. Be careful: you generally want a large base group. Stars indicate conventional base groups in economics.

Sex: *female, male\**

Educational Attainment: *dropout, hsonly\*, somecollege, college, postgrad*

Race: *asian, black, otherrace, white\**

Employment Status: *employed\*, unemployed, notinlf*

Labor Force Status: *lfpart, notinlf*

Birthplace: *immigrant, native\**

#### Categorical variables

*Statefip* is categorical in nature. You cannot include it in a regression without transforming it first. There are numerous ways to do this. You could make a single binary variable for each category, e.g. an *Ohio* binary variable, an *Indiana* binary variable, etc. Even better, you can use your program of choice to create a dummy variable for each category (aka "factor variables"). Be sure to choose your base category carefully as it affects your interpretation on each binary variable.

Here is how to do it in Stata: [\[link\]](#)

Here is how to do it in R: [\[link\]](#)

Here is how to do it in Gretl: [\[link\]](#) (Note: First you must declare a variable as discrete in Gretl [\[link\]](#)).