

Introduction to Metrics : Homework 1

Zhaopeng Qu

3/26/2023

目录

1 Learning Objectives	1
2 Due Date and Formats	2
3 Theory and Application Exercises	2
3.1 Exercises in SW textbooks(50 points)	2
3.2 Practical Exercise: Gender Wage Gap(50 points)	3

1 Learning Objectives

- Be familiar with Stata and household survey datasets
- Learn and Practice cleaning data in reality
- Understand Rubin Causal Model and Regression
- Learn and Practice cleaning data in reality

2 Due Date and Formats

- **Due to April.17 2:00am.**
 - Late sending will lower your scores by a standard “10% per hour deduction”
 - 1. Upload your report(**inculding PDF or docx files**) and all raw files to generate it(such as `.stmd` and `.do` files for **Stata** users) to **教学立方** system. 关于该平台的使用可以参考 **使用指南：学生版**
 - “All files” may include Formal report in *Word* or *PDF*, *Rmarkdown*(`.Rmd` for **R** users) or *stmd*(`.stmd` for **Stata** users) to generate the final report, as well as *Rscript*(`.R` for **R** users) or *dofile*(`.do` for **Stata** users) to clean the survey data for the last question.
- **Language**
 - *English or Chinese*(Writing in English will receive a small bonus)
- **Important Rule: Plagiarism Will Be NOT Tolerant!**. If the TA has enough evidence to make me believe that your homework is **very very** “similar” with another finished by one of your classmates, then both of you will score **ZERO!**

3 Theory and Application Exercises

3.1 Exercises in SW textbooks(50 points)

- 2.20 (pp102)
- 2.23 (pp103)
- 3.17 (pp139)
- 3.18 (pp140)
- 4.6 (pp176)
- 4.12 (pp178)

- 5.8 (pp213)
- 6.6 (pp249)
- 6.10 (pp250)
- 7.11 (pp287)

3.2 Practical Exercise: Gender Wage Gap(50 points)

- Gender inequality is one of the most important economic and social inequality in our society. To understanding the inequality deeper, you will use some data sets in reality to explore some basic facts about the gender wage gaps.
1. Obtain the household survey data
 - Go to following websites and sign up an account to access the data
 - 中国居民收入调查,China Household Income Project(CHIP)
 - 中国家庭追踪调查,China Family Panel Studies(CFPS)
 - 中国健康与营养调查,China Health and Nutrition Survey(CHNS)
 - 中国健康养老追踪调查,China Health and Retirement Longitudinal Study(CHARLS)
 - 中国综合社会调查,Chinese General Social Survey(CGSS)
 - 中国劳动力动态调查,China Labor-force Dynamics Survey(CLDS)
 - 中国家庭金融调查,China Household Financial Survey(CHFS)
 2. Data Cleaning preparing for a simple summary table(30 points)
 - 选择某一年的截面数据, 比如 CHIP2002,CHARLS2011 或其他...
 - 将样本限定在目前居住在城市的城市调查户样本 (包含城乡移民和本地居民), 然后清理出如下变量, 并展示数据的 describe 列表。

Variable Names	Descriptions
annualwage	年工资 = 月工资 *12+ 年终奖金以及其他货币化补贴
workhour	年工作小时数 = 每天工作小时 * 周或月工作天数 *12
workstatus	工作状况: 工资/自我雇佣/失业/家务劳动等
female	= 1 if female; = 0 otherwise
educ	years of education (受教育年限)
hukou	户口状况: 城市 ==1 or 农村 ==0
exper	working experience
self-reported health	“excellent, very good, good ,fair and poor”and scale it from the number “1”to “5”
citycode(if available)	city code or name
provcode	province code or name

3. Descriptive Analysis(10 points)

- 1) 在居住在城市的调查样本中，男性和女性的平均小时工资各自是多少？是否存在统计上显著的差异？请用图来表示（提示：分组以及表示差异的 bar chart，并且在表示差异的图上，表示出置信区间）
- 2) 在居住在城市的调查样本中，男性和女性的在上述其他各特征变量两者之间是否存在显著差异？请用列表的方法表示你的结论、（提示：该表形式为三列：一列为男性各变量的统计信息，第二列为女性统计信息，第三列为他们的差异）

4. Based on the chart and the table above to answer following questions(30 points)

- 1) Is there a wage gap between urban residents and rural migrants? How much about it? Please provide at least two interpretations to the mean comparisons displayed in the chart. What are the

possible underlying causal mechanisms.

- 2) Suppose you want to test one of your interpretations, what kinds of mean comparisons would you make? Be specific about outcome variables and “treatment” variables.
- 3) Following the framework of Rubin Causal Model, please formalize the question above to make you point.
- 4) Are there any policy implication of your interpretations of the result? If so, briefly discuss these implications. If not, briefly discuss why your interpretations are not enough to have significant implications.

5. Wage determination

- A classical wage equation is following

$$\ln Y_i = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{exper}^2 + u_i \quad (4.1)$$

where

- $\ln Y_i$ is the Logarithmic values of **hourly wage** or (**annual wage** if you have no data for working hours)

- 1) At first, just run the regression as equation(4.1)
 - Assume that **education** and **exper** are the 1st interest and 2nd of variables in the regression respectively, please report the values of corresponding estimated coefficients.
 - How can you be sure that these values are true?
 - How to interpret the meaning of the β_1 in the regression?
 - How much is the returns to additional one year of working experience?
- 2) Then, now we consider to control some variables in the regression

above, thus run a regression as follows

$$\ln Y_i = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + X' \gamma + u_i \quad (4.2)$$

where X is a vector which represents demographic and other control variables such as gender, age, hukou, health status and living areas etc.

- put the corresponding variables in the following order **one by one** into the regression: city,gender,age(only if the **experience** variable in your data set is not directly computed by “age-educ-6”),hukou, health status etc.
 - Please suggest how does the value of estimated coefficient $\hat{\beta}_1$ change? Try your best to explain that why it the change in such a way.
- 3) Since the variable **hukou** can indicates whether the respondents are rural-urban migrants or not, please report the value and significance of estimated coefficient before **hukou**in the last specification(all variables in 4.2), try to explain it in economic implication.
 - 4) Assume that you still care about the effect of education on wage,now you would like to see if there is a significant difference of the effect between urban native and rural-urban migrants.

(Hint: you'd better put all regressions above into a table, then answer questions above in order)