

Metrics 101: Homework 1

Zhaopeng Qu

9/28/2017

目录

1 Due Date	2
2 Learning Objectives	2
3 Rural-Urban Migrant Earnings(30 points)	2
3.1 Background	2
3.2 Question:	3
4 Medical Insurance (30 points)	3
4.1 Background	3
4.2 Question	4
5 RAND Health Insurance Experiment(40 points)	5
5.1 Background	5
5.2 Question	5
6 祝各位同学中秋国庆假期快乐!	6

1 Due Date

- *Due to Oct.15 11:00 pm.*
- *Please send your homework(log file for Stata/ PDF for R) to: jl2017f@126.com (计量 2017fall) on time.*
- *IMPORTANT: Rename your file with ID_Name_Major (eg. 151090001_ 户纳东 _ 经济.scml). The homework with wrong filename will not be archived or scored properly.*

2 Learning Objectives

- understand the framework of causal inference
- use *Stata* or *R* to generate and review some basic statistics
- learn the syntax of Markdown by yourself
- practice the operation of R on Datacamp

3 Rural-Urban Migrant Earnings(30 points)

3.1 Background

- Rural to urban migrants in China is one of the most contentious policy debates in the US. Understanding the economic performance of migrants helps to shape our stance on government's migration policies.
- A first simple fact about migration is that on average migrants earn less than urban residents. Based on [Chinese Household Income Project\(CHIPs\)](#) data in 2002, the table below presents the means and standard deviations of migrants and urban residents' earnings and working hours in 2002.

Variables	urban	migrant
Monthly earnings	1148 [772]	812 [576]
Working hours	43 [11.5]	72 [20]

Variables	urban	migrant
Obs.	1811	1253

3.2 Question:

1. provide at least two interpretations to the mean comparisons displayed in Table 1. That is, what are the possible underlying causal mechanisms.(5 points)
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. (5 points)
3. Following the framework of Rubin Causal Model, please formalize the question above to make you piont.(10 points)
4. Are there any policy implications of your interpretations of the data? If so, briefly discuss these implications. If not, briefly discuss why your interpretations are not enough to give policy suggestions.(10 points)

4 Medical Insurance (30 points)

4.1 Background

- the National Health Interview Survey (NHIS) is an annual survey of the U.S. population with detailed information on health and health insurance.
- Among many other things, the NHIS asks: “Would you say your health in general is excellent, very good, good, fair, or poor?” We used this question to code an index that assigns 5 to excellent health and 1 to poor health in a sample of married 2009 NHIS respondents who may or may not be insured.
- Based on the survey in 2009, the table below presents simple comparisons for a outcome variable,the health index and other characteristic variables of insured and uninsured **husbands** in American households.
- The causal relation of interest here is determined by a variable that

indicates coverage by private health insurance.

Variables	Some HI	No HI	Difference
Health Index	4.01 [.93]	3.70 [1.01]	0.31 (.03)
Age	43.98	41.26	2.71 (.29)
Education	14.31	11.56	2.74 (.10)
Employed	.92	.85	.07 (.01)
Family income	106,467	45,656	60,810 (1,355)
Obs.	8,114	1,281	

4.2 Question

1. Briefly describe some basic facts one can learn from the table. What are the statistics in the brackets and parentheses respectively? What is the difference between these two statistics? (10 points)
2. Is the mean difference in health index statistically significant? How do you reach your decision? (10 points)
3. Can we interpret the mean difference in health index as the causal impact of having health insurance? What is the “gap” between the simple mean difference and causal effect? Use potential outcome notations to make your answer concrete.(10 points)

5 RAND Health Insurance Experiment(40 points)

5.1 Background

- The RAND Health Insurance Experiment (HIE), which ran from 1974 to 1982, was one of the most influential social experiments in research history. The HIE enrolled 3,958 people aged 14 to 61 from six areas of USA. The HIE sample excluded Medicare participants and most Medicaid and military health insurance subscribers. HIE participants were randomly assigned to one of 14 insurance plans. Participants did not have to pay insurance premiums, but the plans had a variety of provisions related to cost sharing, leading to large differences in the amount of insurance they offered.
- use Stata or R to answer the question. Download the dataset *rand.dta* from frankqu.weebly.com. The dataset contains the following variables:

Variable Names	Discriptions
any_ins	= 1 if has any health insurance assigned; = 0 otherwise
female	= 1 if female; = 0 otherwise
blackhisp	= 1 if nonwhite; = 0 otherwise
educper	years of education
hosp	hospitalized last year
ghindx	pre-treatment outcome : general health index
cholest	pre-treatment outcome: cholesterol level (mg/dl)
ghindx	post-treatment outcome: general health index
cholestx	post-treatment outcome: cholesterol level (mg/dl)

5.2 Question

1. Generate basic summary statistics such as mean, standard deviation, and total number of observations for variables female, blackhisp,

educper, ghindx, and cholest, separated by the any_ins indicator. Your table for each variables should look similar as the table in Question 1. You can also combine all the information in just one table. (10 points)

2. Two variables in question 1 can be considered as baseline outcome variables. For these two variables, test the null hypothesis that there is no mean difference ($H_0 : \mu = 0$). (10 points) (You should write down the equation of the statistic which you are going to use.)
3. For the two post-treatment variables, repeat the exercises you just did (compute group mean, standard deviation, and run a t-test). (10 points)
4. Interpret the results you obtain (the mean comparison results for characteristic variables, pre-treatment outcome variables, and post-treatment outcome variables). (10 points)

6 祝各位同学中秋国庆假期快乐！



图 1: moon