

# Problem Set #1

## Evaluating the Impact of Health Insurance on Health Outcomes

Prof. Matt Reimer\*

Econ A390: Methods for Public Policy Evaluation

Please answer all the questions below. For all parts, you should provide a log file that contains STATA (or other program) commands and the results. *Please type your explanations and responses directly into this Word document.* You are encouraged to work as a group; however, there may be no more than three students per group. Everyone is responsible for submitting one typed file for grading on Blackboard under Assignments. Please indicate the members of your group at the top of the file. You might find it helpful to read through the assignment, then read your class notes and textbook, and only then try and do the questions. *Attempting this assignment without consulting your notes and the text may be hazardous to your health and your grade!!*

### Problem 1:

The Center for Disease Control and Prevention (CDC) has been tasked with figuring out whether health insurance has a positive effect on people's health. Specifically, the CDC is interested in knowing the impact on health from: 1) having no health insurance; 2) having health insurance with partial coverage; and 3) having health insurance with full coverage. In order to help determine which of these scenarios is likely to lead to better health, CDC has decided to conduct a random experiment. The CDC requested applications from people in the general population who were interested in participating in the experiment. The CDC then randomly selected 60 people from the pool of applicants to participate in the experiment. Before assigning people one of the three types of health insurance packages, the CDC recorded each person's age, income, education, gender, and health, where health was recorded as an index between 1 (not healthy) and 5 (extremely healthy). The CDC then randomly assigned people to the no insurance group with probability  $1/3$  and to the treatment group (gets some health insurance) with probability  $2/3$ . The people assigned to the treatment group are subsequently assigned to "partial coverage" group or the "full coverage" group with equal probability. This results in  $1/3$  of the participants without insurance,  $1/3$  with partial coverage, and  $1/3$  with full coverage. Five years later, the CDC recorded the health outcomes of all original participants. The CDC has hired you to help them analyze the data and determine which of the three scenarios is most effective at improving health. A description of the dataset "CDCexperiment.dta" can be seen in Table 1.

---

\*Telephone: 1-907-786-5430, e-mail: mreimer2@alaska.edu.

Variable	Description
id	unique person identifier
assignment	0=no insurance; 1=some insurance
insurance	0=no insurance; 1=partial coverage; 2=full coverage
pre_health	health index (1 to 5) before experiment
post_health	health index (1 to 5) after experiment
age	age of the person in years
income	income of person (in '000s of dollars)
education	years of schooling
sex	0=male; 1=female

Table 1: Variable descriptions for the dataset “CDCexperiment.dta”.

1. Were the treatment and control groups actually generated by random assignment? Explain, present and discuss your results.
2. Were the treated individuals actually randomly assigned to the partial and full coverage groups? Explain, present and discuss your results.
3. Estimate the effect of having some health insurance (partial or full) on health. Explain, present and discuss your results.
4. Estimate the effect of having health insurance with full coverage, relative to both no health insurance and health insurance with partial coverage. Which of the two forms of health insurance is more influential on health outcomes? Explain, present and discuss your results. *Hint: you want to do this in one regression.*
5. Discuss two concerns you may have about the statistical, internal, construct, and external validity of the randomized experiment.

These data were actually manufactured. Specifically, the variables in the dataset were randomly drawn from the distributions in Table 2. Define  $Y = \text{post\_health}$ ;  $D = 1$  if assignment=1 and  $D = 0$  if assignment=0;  $I = 0$  if no insurance,  $I = 1$  if partial insurance, and  $I = 2$  if full insurance.

6. Calculate the following (*hint: you can do this manually using the information in Table 2 and the definition of an expected value for a discrete random variable*).

a.  $E[Y|D = 0]$

- b.  $E[Y|D = 1]$
- c.  $E[Y|D = 1] - E[Y|D = 0]$
- d.  $E[Y|I = 0]$
- e.  $E[Y|I = 1]$
- f.  $E[Y|I = 2]$
- g.  $E[Y|I = 1] - E[Y|I = 0]$
- h.  $E[Y|I = 2] - E[Y|I = 1]$

7. How well did your estimates approximate the true values of the estimands? Discuss reasons why the estimates were close or far from your estimands.

Variable	Individuals	Random Draw From:
post_health	$I = 0$	$\Pr(Y = 1) = \Pr(Y = 2) = \dots = \Pr(Y = 5) = 0.2$
post_health	$I = 1$	$\Pr(Y = 1) = \Pr(Y = 2) = 0.1; \Pr(Y = 3) = 0.2; \Pr(Y = 4) = \Pr(Y = 5) = 0.3$
post_health	$I = 2$	$\Pr(Y = 1) = \Pr(Y = 2) = \Pr(Y = 3) = 0.1; \Pr(Y = 4) = \Pr(Y = 5) = 0.35$

Table 2: Data generating process for the dataset “CDCexperiment.dta”.