Problemset for Applied Econometrics due June 4, 2014

Maximilian Kasy

May 13, 2014

Warm-up Problem

1) Suppose that $X_1, ..., X_n$ are independent, normally distributed random variables with mean μ_X and variance σ_X^2, and that $Y_1, ..., Y_n$ are independent and normally distributed with mean μ_Y and variance σ_Y^2, and that the Y_i's are independent of the X_i's. Discuss how you would perform a t-test of the null hypothesis that $\mu_x - \mu_y = 1$.

Randomized experiments

2) Suppose the following data arose from an experiment where treatment D was randomly assigned. Calculate estimates for the treatment effect on the mean of Y and for the treatment effect on the median of Y.

i	D_i	Y_i
1	1	3
2	0	1
3	0	2
4	1	3
5	0	0
6	1	4

3) Calculate the Fisher p-value for the null hypothesis that treatment D has no effect on outcome Y using the following data.

i	D_i	Y_i
1	1	3
2	0	1
3	0	2
4	1	3

4) Consider the power function introduced on p.40 of the treatment effects lecture (first set of slides). Suppose the true difference in means is $\mu_1 - \mu_0 = 1$ and $\sigma_0 = 1$, $\sigma_2 = 2$ and $p = 0.25$. What sample size N do we need to achieve power $\beta = 0.9$ for a test of the null hypothesis that $\mu_1 = \mu_0$? What about $\beta = 0.99$?

Conditional independence

5) Consider the following data set. Suppose that treatment assignment was such that treatment is randomly assigned conditional on X_i. Calculate an estimate for the average treatment effect on the treated, α_{ATET}, using matching. Proceed by first finding matches and building a matched table (as in the third set of slides).

i	D_i	X_i	Y_i
1	1	1	3
2	1	3	1
3	1	2	2
4	1	10	3
5	1	4	0
6	1	2	4

i	D_i	X_i	Y_i
7	0	3	3
8	0	2	1
9	0	0	2
10	0	4	3
11	0	4	0
12	0	10	4

6) Show (give a proof) that the following two objects are identical, assuming that X can only take the values $1, \ldots, k$:

$$\sum_{x=1}^{k} \left(E[Y|D=1, X=x] - E[Y|D=0, X=x]\right) \cdot P(X=x)$$

and

$$E\left[Y \cdot \frac{P(X)}{P(X|D=1)} \Big| D=1\right] - E\left[Y \cdot \frac{P(X)}{P(X|D=0)} \Big| D=0\right].$$

Robustness, sensitivity, falsification

7) Consider the paper "Worms: identifying impacts on education and health in the presence of treatment externalities" by Miguel and Kremer (the paper can be found on the course website). Discuss the following questions:

a. What are the experimental units? What is the treatment and its assignment mechanism?

b. Does the "Stable Unit Treatment Value" assumption hold in this context?

c. What might be alternative outcomes to consider? What about alternative controls?

d. Can you think of falsification tests suitable in this context (for instance. testing for a treatment effect on pretreatment outcomes)?

Difference-in-Differences

8) Suppose you have estimation results from a regression of the form

$$\Delta Y_i = \delta + \alpha \cdot D_i + X_i'\beta + D_i \cdot X_i'\gamma + u_i,$$

and suppose that the assumption of common trends holds conditional on X. How could you use the estimated regression coefficients to estimate the average treatment effect? The average treatment effect on the treated? The average treatment effect on the untreated?

9) Consider the linear model

$$Y_i = \mu + \gamma \cdot D_i + \delta \cdot T_i + \alpha \cdot (D_i \cdot T_i) + \varepsilon_i,$$

where $E[\varepsilon|D,T] = 0$. Show that

$$\begin{aligned} \alpha &= \{E[Y|D=1,T=1] - E[D=0,T=1]\} \\ &- \{E[Y|D=1,T=0] - E[Y|D=0,T=0]\}. \end{aligned}$$

10) Difference-in-Differences estimation of causal effects relies on the assumption of common trends. We expect this assumption to also hold in the "pre-period," between times $T = -1$ and $T = 0$. We can test common trends in this "pre-period."

Construct a t-test for the hypothesis

$$E[Y(0) - Y(-1)|D=1] - E[Y(0) - Y(-1)|D=0] = 0.$$

Be explicit about how you would calculate the standard error entering the t-statistic, assuming that your data are repeated cross-sections.

Instrumental variables

11) Assume that Z is a binary random variable and that

$$E[D|Z=1] \neq E[D|Z=0].$$

Show that under these conditions alone

$$\frac{\text{Cov}(Y,Z)}{\text{Cov}(D,Z)} = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]}$$

12) For both of the papers on instrumental variables presented in class, discuss who you think might be the compliers, the never-takers, the always-takers?

Also, come up with a story why the exclusion restriction might be violated in either of these papers.

13) Consider the proof that the two-stage-least-squares estimand is equal to the local average treatment effect on slide 8 of the slides on Instrumental Variables:

$$\begin{aligned} &\frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]} \\ &= \frac{E[Y_1 D_1 + Y_0(1-D_1)|Z=1] - E[Y_1 D_0 + Y_0(1-D_0)|Z=0]}{E[D_1|Z=1] - E[D_0|Z=0]} \\ &= \frac{E[Y_1 D_1 + Y_0(1-D_1)] - E[Y_1 D_0 + Y_0(1-D_0)]}{E[D_1] - E[D_0]} \\ &= \frac{E[(Y_1 - Y_0)(D_1 - D_0)]}{E[D_1 - D_0]} \\ &= \frac{E[Y_1 - Y_0|D_1 - D_0 = 1]}{\Pr(D_1 - D_0 = 1)} \Pr(D_1 - D_0 = 1) \\ &= E[Y_1 - Y_0|D_1 > D_0]. \end{aligned}$$

Where in this proof do we use the assumptions of

- independence,
- exclusion restriction,
- relevance of the instrument,
- and monotonicity of the first stage?