

Adverse selection and Difference-in-Difference example

Graded satisfactory (4% of course grade) or unsatisfactory (0% of grade).

Satisfactory means a serious attempt made to answer the questions. Your answers need not be lengthy. No credit for late assignments. Academic honesty is required.

Questions 4 to 6 use Stata and data in file ass3s25.dta.

1. For each of the following policies, **each taken in isolation**, state whether or not the policy will reduce adverse selection in private health insurance markets. **Provide a brief reason.**

- (a)** It is mandatory for individuals to purchase health insurance.
- (b)** It is mandatory for companies to provide health insurance to anyone who wants to buy it.
- (c)** Health insurance policies are offered to all consumers at the same price.

2.(a) Consider the market for used cars as presented in class and in the course notes.

Let X = value of the car.

Sellers know the value of the car they sell and their utility is $U(X) = X$.

Buyers only know that car value is uniformly distributed on (\$600, \$1000) and their utility is $1.5 \times X$.

Suppose the posted price for used cars is \$800. Will consumers buy a car at this price?

(b) Suppose all individuals face a loss distribution that is uniformly distributed on (\$0, \$20,000). Each individual knows his loss but the insurance company does not. If all individuals are risk neutral will the insurance company make a profit if it sells a complete-cover insurance policy for \$15,000 (and faces administration costs of \$2,000 per policy)?

(c) Give a definition of adverse selection in the context of health insurance.

3. For each of the following state with a brief explanation whether they are likely to keep total health costs down.

- (a)** Higher deductibles on health insurance policies.
- (b)** Price controls on medical services.
- (c)** Universal health insurance.

Questions 4 to 6 reproduce a study that looked at the impact on health outcome (child weight) of providing free health care to poor people in South Africa.

For the statistics methods used in this assignment see especially section 10 of tr132statistics.pdf, posted at the Canvas course site under Files / Statistics for 132.

The following data comes from the article Shinsuke Tanaka (2014), “Does Abolishing User Fees Lead to Improved Health Status? Evidence from Post-Apartheid South Africa”, *American Economics Journal: Economic Policy*, pages 282-312.

Background: In 1994 health care use fees (at community clinics) in South Africa were abolished for pregnant women and children under the age of six. This substantially increased access for (poor) black South Africans, particularly those in communities with a health clinic (compared to those in communities without a health care clinic).

The sample is of young black South African children with data from a longitudinal household survey in the KwaZulu-Natal province of South Africa.

We consider the change in child health (as measured by weight-for-age z-score) from 1993 (the pretreatment year) to 1998 (the post treatment year) for those in communities with a health care clinic (the high-treatment group) to those in communities without a health care clinic (the low-treatment group). Over time the weight-for-age z-score increased for both high- and low treatment groups. But it increased more for the high-treatment group (a differences in differences estimate). So the paper concludes that access to free health care increased child weight.

Dataset: ass3s25.dta

Key variables: waz = Weight for Age Z Score (standardized to mean 0 and variance 1)
high = 1 if clinic93=1 and =0 if clinic93=0
year = 93 if 1993 and = 98 if 1998
post = 1 if year==98 and =0 if year==93
postXhigh = post times high
idcommunity = identifier for community

Other variables: Give Stata command describe

Statistical methods: Discussed in class and Topic 10. Differences in differences
tr132statistics.pdf posted at Canvas / Files / Statistics for 132

4.(a) Read in Stata dataset **ass3s25.dta** and give commands **describe** and **summarize** to initially inspect the data.

(b) A z-score is a score standardized to have mean 0 and standard deviation 1. Here variable **waz** is presumably standardized for the population of all children. For this sample does variable **waz** appear to be approximately a standardized z-score.

(c) Does variable **waz** appear to be approximately normally distributed?
Use command **kdensity** with option **normal**.

(d) How many communities are in the dataset? One way to find this is to give command **tabulate idcommunity** and add up.

5. A differences in differences estimate is the difference across treatment groups in changes over time in the outcome of interest. Equivalently a differences in differences estimate is the change over time in the difference across treatment groups.

Consider the following table, which we will fill in to get the difference in differences estimate of the effect of free clinics on child weight.

	Treatment = High	Treatment = Low	Difference over treatment
Year==93	A	C	A - C
Year==98	B	D	B - D
Change over time	B - A	D - C	Answer = (B - D) - (A - C) = (B - A) - (D - C)

(a) Calculate each of A, B, C and D by computing the appropriate sample mean of **waz**.

For example, A is computed as **sum waz if year==93 & high==1**.

(Or use **mean waz if year==93 & high==1** or use **regress waz if year==93 & high==1**).

(b) In 1993, before free clinics were introduced, did the mean weight-for-age z-score differ across the high and low treatment groups? Conduct an appropriate test at level 0.05.

(c) Compute by hand B - A, the change over time for the high treatment group.

Compute by hand D - C, the change over time for the low treatment group.

Hence compute (B - A) - (D - C), the difference across treatment groups in these changes over time. This is the differences-in-differences estimate. What do you conclude?

(d) Compute by hand A - C, the difference across treatment groups in 1993.

Compute by hand B - D, the difference across treatment groups in 1998.

Hence compute (B - D) - (A - C), the change over time in the difference across treatment groups. This is also the differences-in-differences estimate.

6. The differences-in-differences estimate can be computed using regression methods. This additionally provides a way to test whether any difference is statistically significant.

Let T = 0 (before) and T = 1 (after) and S = 0 (low treatment) and S = 1 (high treatment).

The general regression for outcome y is

$$y = \beta_1 + \beta_2 T + \beta_3 S + \beta_4 T \times S + \text{error}$$

Here $\text{waz} = \beta_1 + \beta_2 \text{post} + \beta_3 \text{high} + \beta_4 \text{post} \times \text{high}$

Then the OLS estimate of β_4 is the differences-in-differences estimate.

(a) Give command **regress waz post high postXhigh, vce(robust)**

(b) Did you get the same differences-in-differences estimate as in the previous question?

(c) Test whether this estimate is statistically significant at level 0.05.

(d) The previous analysis assumed that each observation was independent. In fact it is likely that there may be some correlation for children in the same community. To control for this we use an alternative method to compute the standard errors of the regression coefficients.

Give command **regress waz post high postXhigh, vce(cluster idcommunity)**

(e) Again test whether the differences-in-differences estimate is statistically significant at level 0.05.