Tutorial 2 Solutions, Econometric Theory I

Exercise 1: measuring the effect of tracking on students' test scores

This exercise uses data from "Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. The American Economic Review, 101(5), 1739-1774." In this paper, the authors randomly assigned 108 primary schools with 2 1st grade classes in Kenya to a tracking group and to a non-tracking group. In the tracking group, the 2 1st grade classes in the school were formed based on students' ability: students faring below the school median in a test in the beginning of the year go to class A, while students faring above the school median go to class B. In the non-tracking group, classes A and B are randomly formed, and therefore both classes have students below / above the median.

The data set tracking.dta contains data on 5170 students in these 108 primary schools. schoolid is the unique identifier of their school. tracking is an indicator equal to 1 for students whose school was randomly assigned to the tracking group. bottomhalf is an indicator equal to 1 for students who fared below the median of their school in the beginning of first grade test. Finally, scoreendfirstgrade is students' standardized score in the end of first grade test.

Throughout the exercise, you need to bear in mind that the randomization took place at the school level, not at the student level. Therefore, your unit of observation in all of what follows should be schools, not students. Hint: the Stata command "collapse" might prove useful.

1. Estimate the effect of tracking on students' end of first grade test scores.

Solution

See program. Tracking increases students' average test score by 14% of a standard deviation. This effect is significant at the 10% level, not at the 5% level.

2. Use a 10% level randomization inference test to assess whether the finding of question 1 is robust or whether it rests on an unappropriate asymptotic approximation.

Solution

See program. The result seems robust. A 10% level randomization test of $y(0) = y(1)$ is also rejected.

3. One might fear that tracking benefits strong students while harming weaker ones. Assess whether this is a legitimate concern.

Solution

See program. To assess whether this concern is legitimate, we can estimate the effect of tracking among students below the median in the beginning of the year, and among students above the median in the beginning of the year. To estimate, say, the effect of tracking among students below the median, we can compute for each school the mean of end of 1st grade test score among students initially below the median, and then we can regress this variable on the tracking indicator. We can proceed similarly to estimate the effect of tracking among students above the median. The results show that students below and above the median benefit similarly from tracking.

Exercise 2: some properties of the ATE estimator in randomized experiments

1. Prove the following theorem:

Theorem 0.0.1 *Using draws from a uniform distribution to generate draws from other continuous distributions*

Let F denote a strictly increasing cdf. If U follows the uniform $[0,1]$ distribution, then the cdf of $F^{-1}(U)$ is F.

Hint: the cdf of $F^{-1}(U)$ if the function $x \mapsto G(x) = P(F^{-1}(U) \leq x)$. You need to show that $G(x) = F(x)$.

Solution

$$\begin{aligned} P(F^{-1}(U) \leq x) &= P(U \leq F(x)) \\ &= F(x). \end{aligned}$$

The following questions require using Stata.

2. Set the number of observations to 1000, create a variable $y(0) = U$, where U follows a $N(0,1)$ distribution, and create a variable $y(1) = 0.5y(0) + 0.5V + 0.2$, where V is a $N(0,1)$ random variable independent of U. $y(0)$ and $y(1)$ represent the potential outcomes of 1000 units that participate in a randomized experiment. Hint: it follows from the previous question that if ε follows a uniform distribution, $\Phi^{-1}(\varepsilon)$ follows a $N(0,1)$ distribution. The Stata command for Φ^{-1} is invnormal().

Solution

See program.

3. Given the way we created $y(0)$ and $y(1)$, is the effect of the treatment homogeneous or heterogeneous across units?

Solution

The effect of the treatment is heterogeneous: $y(1) - y(0) = 0.5(V - y(0)) + 0.2$, and $V - y(0)$ varies across units.

4. Compute ATE_{1000}, the average effect of the treatment in this fixed population of 1000 units. Store this number in a scalar.

Solution

See program. You should find a result close to 0.2.

5. Compute the correlation between $y(1)$ and $y(0)$ and the variance of $y(1)-y(0)$ in this finite population of 1000 units.

Solution

See program. You should find a correlation close to 0.707 and a variance close to 0.50.

6. Write a 800 iterations loop, where in each iteration you:

1. Create a variable containing a random number.

2. Sort the 1000 observations according to that random number.

3. Create a dummy variable D equal to 1 for the first 500 observations in the sorted data set.

4. Create a variable $Y = (1 - D)y(0) + Dy(1)$.

5. Regress Y on D, using the robust option. Check that the coefficient of D is equal to $\widehat{ATE}_{1000}$. Check that the variance of that coefficient is equal to $\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}$.

6. Compute $IC(0.05)_{+} = \left[\widehat{ATE}_{1000} - 1.96\sqrt{\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}}, \widehat{ATE}_{1000} + 1.96\sqrt{\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}}\right]$, and the indicator $1\{ATE_{1000} \in IC(0.05)_{+}\}$.

7. Store $\widehat{ATE}_{1000}$, $\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}$, and $1\{ATE_{1000} \in IC(0.05)_{+}\}$ in a matrix.

Compute the mean of $\widehat{ATE}_{1000}$ over those 800 replications, and compare it to ATE_{1000}. Compute the variance of $\widehat{ATE}_{1000}$ over those 800 replications, and compare it to the mean of $\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}$ over those replications. Compute the mean of $1\{ATE_{1000} \in IC(0.05)_{+}\}$. Explain your results in light of the lecture notes.

Solution

See program.

You should find that the mean of $\widehat{ATE}_{1000}$ is close to ATE_{1000}, as predicted by Theorem **??**. You should also find that the variance of $\widehat{ATE}_{1000}$ is lower than the mean of $\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}$. This reflects the fact that $\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}$ is an unbiased estimator of $V(\widehat{ATE}_{1000})$ only if the effect of the treatment is constant. Otherwise, $\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}$ is an unbiased estimator of an upper bound of $V(\widehat{ATE}_{1000})$. Here, the effect of the treatment is not constant, hence the result.

Accordingly, you should find that across those 800 possible assignments of the treatment, $ATE_{1000} \in IC(0.05)_{+}$ more than 95% of the time. $IC(0.05)_{+}$ is a conservative confidence interval.

7. Write a 800 iterations loop, where in each iteration you:

    1. Create a variable $y(0) = U$, where U follows a $N(0, 1)$ distribution, and create a variable $y(1) = 0.5y(0) + 0.5V + 0.2$, where V is a $N(0, 1)$ random variable independent of U, with 1000 observations each.

    2. Create a variable containing a random number.

    3. Sort the 1000 observations according to that random number.

4. Create a dummy variable D equal to 1 for the first 500 observations in the sorted data set.
5. Create a variable $Y = (1 - D)y(0) + Dy(1)$.
6. Regress Y on D, using the robust option. The coefficient of D is equal to $\widehat{ATE}_{1000}$. The variance of that coefficient is equal to $\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}$.
7. Compute $IC(0.05)_{+} = \left[\widehat{ATE}_{1000} - 1.96\sqrt{\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}}, \widehat{ATE}_{1000} + 1.96\sqrt{\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}}\right]$, and the indicator $1\{0.2 \in IC(0.05)_{+}\}$.
8. Store $\widehat{ATE}_{1000}$, $\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}$, and $1\{0.2 \in IC(0.05)_{+}\}$ in a matrix.

Compute the mean of $\widehat{ATE}_{1000}$ over those 800 replications, and compare it to 0.2. Compute the variance of $\widehat{ATE}_{1000}$ over those 800 replications, and compare it to the mean of $\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}$ over those replications. Compute the mean of $1\{0.2 \in IC(0.05)_{+}\}$ (i.e. $1\{ATE_{1000} \in IC(0.05)_{+}\}$). Use the lecture notes to explain why the results change wrt to those you obtained in question 6.

Solution

See program.

In this design, we assume that the population is infinite, and the 1000 observations we get to observe are a random sample from that population. Therefore, randomness arises both from the sampling of these 1000 observations, and from the random allocation of the treatment to 500 units out of a 1000. In this context, our parameter of interest is $ATE = E(Y(1) - Y(0)) = 0.2$.

You should find that the mean of $\widehat{ATE}_{1000}$ is close to 0.2, as predicted by Theorem **??**. You should also find that the variance of $\widehat{ATE}_{1000}$ is close to the mean of $\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}$, as predicted by Theorem **??**. Accordingly, you should find that across those 800 possible assignments of the treatment, $0.2 \in IC(0.05)_{+}$ around 95% of the time.

8. Write a 800 iterations loop, where in each iteration you:

1. Create a variable $y(0) = U$, where U follows a $N(0, 1)$ distribution, and create a variable $y(1) = y(0) + 0.1771$, with 1000 observations each.
2. Create a variable containing a random number.
3. Sort the 1000 observations according to that random number.
4. Create a dummy variable D equal to 1 for the first 500 observations in the sorted data set.
5. Create a variable $Y = (1 - D)y(0) + Dy(1)$.

6. Regress Y on D, using the robust option.

7. Compute the indicator $1\left\{\left|\frac{\widehat{ATE}_{1000}}{\sqrt{\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}}}\right|>1.96\right\}$.

8. Store that indicator in a matrix.

Compute the mean of $1\left\{\left|\frac{\widehat{ATE}_{1000}}{\sqrt{\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}}}\right|>1.96\right\}$ over those 800 replications. Use the lecture notes to explain your result. Hint: notice that in this simulation design, $V(Y(0))=V(Y(1))=1$. Then, where does 0.1771 come from?

Solution

See program.

You should find that the mean of $1\left\{\left|\frac{\widehat{ATE}_{1000}}{\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}}\right|>1.96\right\}$ is close to 80%. This is because $0.1771=2.8\sqrt{\frac{1000}{500^2}}$, the MDD in a test of $ATE=0$ against $|ATE|\geq MDD$ with level 0.05 and power 0.80.

9. Assume you want to use a randomized experiment to measure the effect of a treatment. Your experiment will have 1000 subjects. 500 will be treated, while 500 will remain untreated. Moreover, given the nature of the treatment, you think it makes sense to assume that $V(Y(0))=V(Y(1))$. No one has ever measured the effect of the specific treatment you are interested in, but a literature review of the effects of treatments with a similar cost shows that they typically produce effects in the range of 10% of a standard deviation of the outcome. Should you embark in this experiment?

Solution

10% of a standard deviation of the outcome is a lower effect size than the minimum detectable difference of this experiment (17.71% of a standard deviation of the outcome, see previous question). If the true effect is 10% of a standard deviation of the outcome, there are more than 20% chances you will end up concluding the effect of the treatment is not significantly different from 0. It would be quite risky to embark in this experiment.

10. Write a 800 iterations loop, where in each iteration you:

1. Create variables $y(0)=U$, where U follows a $N(0,1)$ distribution, and $y(1)=y(0)$, with 10 observations each.
2. Create a variable containing a random number.
3. Sort the 10 observations according to that random number.
4. Create a dummy variable D equal to 1 for the first 5 observations in the sorted data set.
5. Create a variable $Y=(1-D)y(0)+Dy(1)$.

6. Regress Y on D. Store $\widehat{ATE}_{10}$, the coefficient of D in that regression.

7. Then, write a 100 iterations loop, where in each iteration you:

   (a) Create a variable containing a random number.

   (b) Sort the 10 observations according to that random number.

   (c) Create a dummy variable $\tilde{D}$ equal to 1 for the first 5 observations in the sorted data set.

   (d) Regress Y on $\tilde{D}$. Store the coefficient of $\tilde{D}$ in that regression.

8. Compute the 5th percentile and the 95th percentile of the coefficients of $\tilde{D}$ in the previous loop. Let $q^{\tilde{D}}_{0.05}$ and $q^{\tilde{D}}_{0.95}$ denote those percentiles.

9. Compute the indicator $1\{\widehat{ATE}_{10} \in [q^{\tilde{D}}_{0.05}, q^{\tilde{D}}_{0.95}]\}$.

10. Store $1\{\widehat{ATE}_{10} \in [q^{\tilde{D}}_{0.05}, q^{\tilde{D}}_{0.95}]\}$ in a matrix.

Compute the mean of $1\{\widehat{ATE}_{10} \in [q^{\tilde{D}}_{0.05}, q^{\tilde{D}}_{0.95}]\}$ over those 800 replications. Use the lecture notes to explain your result.

Solution

See program.

You should find that across those 800 replications, the mean of $1\{\widehat{ATE}_{10} \in [q^{\tilde{D}}_{0.05}, q^{\tilde{D}}_{0.95}]\}$ is close to 0.10. $1\{\widehat{ATE}_{10} \in [q^{\tilde{D}}_{0.05}, q^{\tilde{D}}_{0.95}]\}$ is the randomization inference test with level 10% of the null hypothesis that $Y(0) = Y(1)$ we discussed in the notes. Indeed, in the intermediary loop we compute 100 possible values of $\widehat{ATE}_{10}$, under the null hypothesis that $Y(0) = Y(1)$. Doing so, we estimate the distribution of $\widehat{ATE}_{10}$ under that null, and comparing $\widehat{ATE}_{10}$ to the 5% and 95% percentiles of that estimated distribution is a test of that null. In this design, the null $Y(0) = Y(1)$ is satisfied, which is why we reject the null 10% of the time. Notice that this test has the expected 10% level despite the fact we consider a situation with only 10 observations. The validity of this test does not rest on an asymptotic approximation.