**Difference-in-Differences Workshop**

This workshop uses simulated data based on a real study testing the effectiveness of a Tool on surgical safety, measured by NOTECHS score. Code for this workshop is available for SAS, Stata, and R software. These can be downloaded from **https://github.com/eleanormurray/DiD\_Tutorial**.

Note: Remember that since the data in this workshop is synthetic, any conclusions drawn from the following analyses should not be taken as evidence for or against real world causal effects.

1. **Getting to know the data**

Run Code Section 0: Data set up & formatting. If using SAS Studio, it might make sense to omit the optional lines of code in this section pertaining to style. Running the format code from lines 9-27 of the ‘tutorial\_style\_template.sas’ is important, though, before running the main difference-in-differences code. Answer the following questions based on your output from Section 0.

**Question 1.1**

How many surgical observations (n) are in the did\_sim data set?

20,000

**Question 1.2**

How many variables are in the did\_sim data set?

48

**Question 1.3**

How many surgeries occurred before introduction of the Tool (DBT)? How many surgeries occurred in the post-intervention time period?

9,951 surgeries occurred in the pre-intervention time period

10,049 surgeries occurred in the post-intervention time period

**Question 1.4**

In this study, departments were assigned to either a control group or a treatment group where surgeons were offered the Tool to use. How many surgical observations are in each treatment group?

4,776 surgical observations are in the control group

15,224 surgical observations are in the intervention group

1. **Causal Assumptions**

There are three major causal assumptions that must be met before drawing causal interpretations from the difference-in-differences analyses. Run Code Section 1: Checking Assumptions. Answer the following questions based on the output and assumptions laid out in the source paper.

**Question 2.1**

In the context of this study design, what is the causal question?

What is the average treatment effect among the treated (ATT), or rather the impact on surgical performance of the Tool for the surgeries in which Tool use was offered?

**Question 2.2**

The means procedure is performed to assess the average NOTECHS score in the control and intervention departments at both the pre- and post-COVID baseline time periods. Compare these outcome values.

At the pre-COVID baseline, the mean NOTECHS scores are similar across the intervention groups, with less than half a point difference (0.1108) between groups. At the post-COVID baseline, the mean NOTECHS scores are also similar, with less than a point difference (0.5110) between intervention groups.

**Question 2.3**

What is the pre-intervention trend difference (control and intervention departments) in NOTECHS score? Provide 95% confidence limits and the standard error. The pre-intervention trend difference provides support for which causal assumption?

Pre-Intervention trend difference = 0.40

95% CI: (0.06, 0.74)

Standard error: 0.17

The pre-intervention trend difference provides support for the parallel trends assumption holding.

**Exercise 2.4**

Briefly describe the positivity assumption in the context of the study.

All departments in the study had a non-zero probability of being selected for use of the Tool.

1. **Crude Analyses**

Run Code Section 2: Unadjusted models. Two different linear regression models are created, one excluding the pre-COVID data and one using the full data set, and a Poisson regression is also modeled. Answer the following questions based on the source paper and output from only the regression models which include the full data set.

**Exercise 3.1**

Draw a directed acyclic graph (DAG) of the assumed data structure including the following variables: Pre-COVID Baseline (Yt=-12), Baseline (Yt=0), Z, Change in NOTECHS Score, Device Type, Case Complexity, and Department.

**A diagram of a covid-19

Description automatically generated with low confidence**

**Exercise 3.2**

Provide the regression equation for the unadjusted linear model using parameter estimates β and the notation from Box 1. Which parameter can be used to calculate the difference-in differences estimator?

E(Yt | Z, A) = β0,t + β1A + β2Z + β3AZ

The parameter β3 will equal the difference-in-differences estimate

**Exercise 3.3**

Fill in the table below with the predicted values and difference-in-differences estimator from the unadjusted linear regression model. Here, it will be helpful to run Optional Code 2.2.2.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Post-COVID Baseline | Post-Intervention | *Row differences* |
| Intervention | 36.33 | 37.19 | 0.86 |
| Control | 36.02 | 38.37 | 2.35 |
| *Column Differences* | 0.31 | -1.18 | DiD = -1.49 |

**Question 3.4**

What effect does the Tool have, if any, on average NOTECHS score in the intervention group? Report on the significance of the difference-in-differences estimate.

The difference-in-differences estimate of -1.49 indicates that the average effect the Tool had on NOTECHS score in the intervention group was a decrease by 1.49 points. With a 95% confidence interval of (-1.89, -1.09) and p-value less than 0.0001, the mean unconditional difference-in-differences estimate is considered statistically significant at the 0.05 level.

**Exercise 3.5**

Provide the regression equation for the unadjusted Poisson model using parameter estimates α and the notation from Box 1.

logE(Y | Z, A) = α0t + α1A + α2Z + α3AZ

**Exercise 3.6**

Fill in the table below with the predicted values and difference-in-differences estimator from the unadjusted Poisson regression model.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Post-COVID Baseline | Post-Intervention | *Row differences* |
| Intervention | 36.34 | 37.19 | 0.85 |
| Control | 36.02 | 38.37 | 2.35 |
| *Column Differences* | 0.32 | -1.18 | DiD = -1.5 |

**Question 3.7**

How do the results from the unadjusted linear and Poisson models compare?

As expected, the difference-in-differences estimate, point estimates, and row and column differences in the crude Poisson model are extremely similar to those in the crude linear model. Point estimates only begin to differ at the hundredth place. The purpose of this comparison is to demonstrate the interchangeability of using either model in the difference-in-differences analysis of NOTECHS score.

1. **Adjusted Analyses**

Run Code Section 3: Adjusted Linear Models. Answer the following questions based on the source paper and your output.

**Question 4.1**

Which of the variables in the DAG should be adjusted for as confounders? Why?

Device type and case complexity should be adjusted for as confounders since they both vary between departments. It is also possible that device type and case complexity may vary before and after introduction of the Tool within the intervention departments.

**Exercise 4.2**

Provide the regression equation for the adjusted linear model using parameter estimates γ and the notation from Box 1.

E(Y | Z, A, L) = γ0t + γ1A + γ2Z + γ3AZ + γ4’L

**Question 4.3**

What is the mean conditional difference-in-differences estimate? Provide 95% confidence limits and the standard error.

Mean conditional difference-in-differences estimate: -1.24

95% CI: (-1.54, -0.94)

Standard error: 0.15

**Question 4.4**

The first step in adjusting for confounders required using a multivariable regression model conditional on confounder levels. However, the unconditional effects may be of greater interest in this analysis. What is the name of the process used to estimate these unconditional (marginal) effects after adjustment? Why is this process considered a necessary step in the adjusted analysis?

Standardization.

In this study, there is interest in knowing what the differences-in-differences estimate would have been (in this same population) if the confounders had been distributed randomly between the intervention groups. So, standardizing the estimate to the distribution of the confounders in the study sample is required.

**Exercise 4.5**

Fill in the table below with the predicted values and unconditional (marginal) difference-in-differences estimator from the adjusted linear regression model. Here, it will be helpful to run Optional Code 3.2.2.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Post-COVID Baseline | Post-Intervention | *Row differences* |
| Intervention | 36.22 | 37.24 | 1.02 |
| Control | 36.08 | 38.34 | 2.26 |
| *Column Differences* | 0.14 | -1.1 | DiD|L = -1.24 |

**Question 4.6**

What effect does the Tool have, if any, on average NOTECHS score in the intervention group, after adjusting for confounders?

The difference-in-differences estimate of -1.24 (for both conditional and unconditional) indicates that the average effect the Tool had on NOTECHS score in the intervention group was a decrease by 1.24 points. With a 95% confidence interval of (-1.54, -0.94) and p-value less than 0.0001, the mean conditional difference-in-differences estimate is considered statistically significant at the 0.05 level.

**Question 4.7**

How do the crude and adjusted difference-in-differences analyses compare? What conclusions can be drawn about the Tool regarding surgical safety?

The average effect of the Tool decreasing NOTECHS score in the intervention departments is consistent across the crude and adjusted difference-in-differences analyses. However, in the adjusted model, the difference-in-differences estimate is less extreme, or rather closer to the null of 0. One could conclude that the Tool slightly decreased surgical performance following its introduction and recommend additional training. However, despite significant result, a change in 1.24 points on a 44-point scale is quite small, so one could also conclude that the Tool did not meaningfully affect surgical safety in the month immediately following introduction of the Tool in the intervention group.