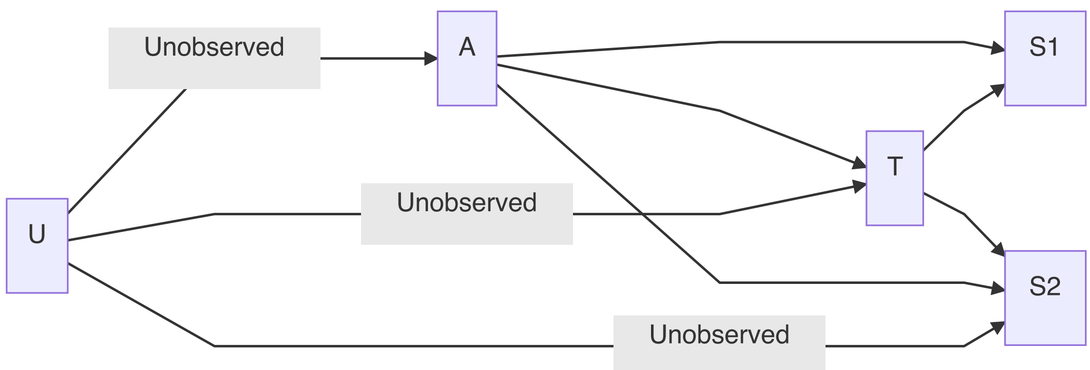
Exam 2 Prep - Panel Data Models and Difference in Difference Designs

> You are interested in the impact of a tutoring program on calculus exam scores. Following the first exam tutoring is offered to anyone who wants it. You know the score on the first exam, who received tutoring, and the score on a second exam after tutoring occurred for those that wanted it.

1. You suspect those that have more innate ability (A), which does not change over time, and those that scored higher on exam 1 (S1) are less likely to sign up for tutoring (T). Draw the DAG for the causal inference problem.



2. Describe how a difference in difference approach would calculate the effect of tutoring. Why does this remove unobserved individual factors that do not change over time?
    - A difference in difference (DID) approach calculates the effect of tutoring by taking the average of the treatment effect, then subtracting the originial from the average. DID removes unobservable individual factors that remain constant over time by eliminating individual variation and controlling for unknown unknowns. DID works by comparing each unit to itself over time then comparing their differences. This cancels out things that did not change over time and, as a result, cancels out any unobservable variables that remained constant.

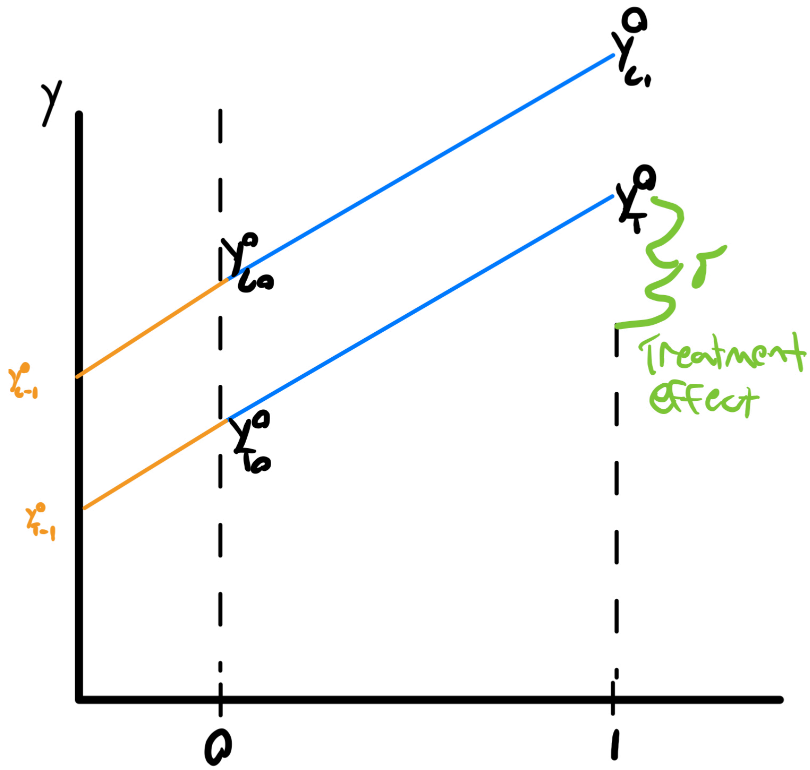
	T_1	T_0
1	y_1^1	y_1^0
2	y_2^1	y_2^0

$$\begin{aligned}\Delta y^1 &= (y_2^1 - y_1^1)\\ \Delta y^0 &= (y_2^0 - y_1^0)\end{aligned} \tag{1}$$

$$\Delta y^1 - \Delta y^0$$

That last line is the difference in difference.

3. What is the parallel trends assumption and why is it crucial to difference in difference designs? Draw a sketch to illustrate.
   - DID assumes parallel trends which is that the difference between the treatment and control groups are constant over time. This can be observed by looking at the data over a large period of time. It requires that the difference btween the treatment and control groups are constant over time. The parallel trends assumption is important to ensure that the data does not indicate a biased estimation of causal effect.



4. How would you employ randomization inference to test the null hypothesis of no association between tutoring and exam scores in the difference in difference model?
   - Randomization inference can be used to test the null hypothesis that there is no association between tutoring and exam scores in the DID model by assuming that there is no treatment effect, also known as Fisher's Sharp Null. In theory, this allows us to specify all counterfactual outcomes. That is, if there is no treatment effect for tutoring’s impact on the calculus exam scores, then the control units for all of the outcomes would be the same even if the students were placed in treatment, which, in this case, is tutoring. Similarly, the treatment units’ outcomes would have been unchanged had they been placed in the control group. Thus,

under Fisher's Sharp Null hypothesis, we can map all of the possible outcomes from our data. Then, all we have to do is construct all possible random assignments and calculate the test statistic for each one. This helps us account for any uncertainty about unobserved potential outcomes.

5. Write the panel data regression model expressing the score on exam t, Sit as a linear function of tutoring (Tit) and individual specific fixed effects (ui). Show that if you subtract the individual average, the individual specific effect drops out, leaving a model in "demeaned" variables. Discuss why including individual dummy variables and demeaning achieve the same thing.

$$T_{it} = \alpha + \beta Treat_{it} + u_i + \sum_{it} + \beta x_{it}$$

$$u_i \text{ represents the specific fixed effects}$$

$$\text{then take the average...}$$

$$\frac{\sum_i T_{it}}{y} = \frac{\sum_i \alpha}{y} + \delta \frac{\sum_i Treat_{it}}{y} + \frac{\sum_i u_i}{y} + \beta \frac{\sum_i x_{it}}{y} \tag{2}$$

$$\bar{T} = \alpha + \delta \bar{Treat}_i + u_i + \beta \bar{x}_i$$

$$\text{then subtract the individual average and drop individual specific effects...}$$

$$T_{it} - T_i = \alpha\delta(Treat_{it} - \bar{Treat}_i) + \beta(x_{it} - \bar{x}_i) + \sum_{it}$$

   - Using dummy variables instead of demeaning the variables is okay because they are both used to remove individual heterogeneity that is time invariant. This removes individual specific fixed effects (u_i) and also eliminates individual variation and controls for unknown confounders about the individual that stay constant throughout the model. Thus, the model is the same whether you use dummy variables or de-mean it.

6. Why is the panel data regression model from #5 a difference in difference model? Why does this implementation allow for a more general and flexible approach? Hint: think about when the treatment timing varies across units and when there are many observable factors that account for much variation in the dependent variable.

   - The model from question five is a difference in difference model because all of the individual variation is in dummy variables or was de-meaned. This allows us to omit any variables that are constant across time. DID allows us to use a more general and flexible approach because it can help us exploit situations where the treatment effect is staggered over time by taking the difference of the average before and after the difference between treatment and non-treatment.

7. Consider the four models show in the table below. The top number in each cell is the regression coefficient and the bottom number in parentheses is its standard error. Which model is "best" in this context? Why? Interpret the results.

Model	1	2	3
Individual Fixed Effects?	No	No	Yes
Tutoring Indicator	-0.405 (1.009)	-1.388 (1.409)	11.361 (1.291)
Exam 2 Indicator		1.322 (1.322)	-7.029 (1.045)

- The third model is the best. At first glance, model one looks pretty good because it has the lowest standard error in any of the models. Of course, that doesn't quite sit right because only tutoring is used as an indicator and not the exam two scores as well. So then, looking for the lowest standard error, we see model three. This makes sense because it is essentially the same as model two, but uses individual fixed effects. So then, I, as the over-achiever I am, recreated the models in R and lo and behold, the third model is indeed, the best. It has the highest R^2 and the lowest p-value. Of course, those aren't always the best metrics, but here they are much much better than in models one and two.

> Extra: If you have time to do this it would be a good idea. The data behind the models in #7 is in the file tutoring2.csv. Implement the models and conduct randomization inference to test the null hypothesis of no effect for the simplest difference in difference calculation.