

Семинар 4

Задача 1. Изучается эффект программы в области здравоохранения в США для пожилых людей на здоровье людей старше 65 лет – непосредственно участников данной программы. Рассматривается 5 лет до и после реализации программы. Те штаты, в которых данная программа не реализовалась, рассматривается как контрольная группа. Ниже представлены результаты оценивания средних значений самооценки здоровья (шкала для самооценки – 10-балльная, более высокое значение соответствует более высокой самооценке здоровья).

Группа/Период	До введения программы	После введения программы
Штаты: Группа воздействия	6.72	6.31
Штаты: Контрольная группа	6.25	6.36

1. Чему равна оценка коэффициента при дамми-переменной «Период» (1 – период после введения программы в области здравоохранения, 0 – период до введения соответствующей программы) в регрессионной модели DiD без включения контрольных переменных?
2. Проинтерпретируйте полученную оценку в предыдущем пункте. Что она содержательно показывает?
3. Вычислите оценку модели difference-in-differences без включения контрольных переменных, соответствующую оценке коэффициента при переменной взаимодействия между дамми для группы воздействия и периода после введения программы в области здравоохранения
4. Запишите спецификацию регрессионной модели DiD без контрольных переменных и проинтерпретируйте все оценки коэффициентов
5. В предположении о соблюдении допущения параллельности трендов рассчитайте значение counterfactual outcome в группе воздействия в период после введения указанной программы в области здравоохранения. Своими словами объясните, что показывает counterfactual outcome в контексте модели DiD
6. Представьте, что дизайн исследования был изменен: вместо сравнения здоровья людей возрастной группы от 65 лет и старше в разных

штатах, было предложено сравнить в одном штате, в котором была введена соответствующая программа в области здравоохранения, возрастные группы «от 65 лет и старше» (целевая группа программы) и «младше 65 лет», на которую не распространялось участие в программе. Оцените данную стратегию критически, как Вам кажется, какие ограничения у этой стратегии. Что бы Вы предприняли для того, чтобы эти ограничения преодолеть или хотя бы уменьшить.

Задача 2. Поразмышляйте, подойдет ли метод DiD для оценки эффективности антивирусных мер, вводимых во время пандемии российскими регионами, с точки зрения эффекта на заболеваемость и смертность от коронавирусной инфекции

1. Как бы Вы выбирали регионы на роль группы воздействия и контрольной группы, на что ориентировались?
2. Какие контрольные переменные Вы включили бы в модель? Свой выбор объясните
3. Какие «подводные камни» есть у применения DiD в таком контексте? Какие бы поправки / дополнения Вы внесли в эмпирическую часть в ответ на данные ограничения?

Задача 3. Ознакомьтесь с нижеприведенной выдачей (корреляционная матрица и оценки коэффициентов линейной регрессии y на x , z и w):

	y	x	z	w
y	1.0000	0.8035	-0.203	-0.216
x	0.8035	1.0000	-0.717	-0.692
z	-0.203	-0.717	1.0000	0.9514
w	-0.216	-0.692	0.9514	1.0000

	coef	std.error	t	Pr> t
(Intercept)	-10.7762	3.5310	-3.052	0.03796 *
x	1.5434	0.2138	7.218	0.00195 **
z	0.1507	0.3014	0.500	0.64971
w	0.8212	0.4382	1.874	0.09497 .
<hr/>				
Signif. codes:	0 ‘***’	0.001 ‘**’	0.01 ‘*’	0.05 ‘.’
Multiple R-squared:	0.9486	Adjusted R-squared:	0.8990	

1. Можно ли говорить о том, что модель страдает от мультиколлинеарности? Приведите не менее двух свидетельств из выдачи в качестве обоснования. Можно приводить как прямые, так и косвенные свидетельства мультиколлинеарности.
2. Проинтерпретируйте значения VIF. Что они показывают?

VIF:

x	z	w
2.154286	12.03485	7.954813