

Семинар 4

Задание 1.

Рассмотрим значения outcome variable Y – количества заданий, выполняемых работником за день – при разных значениях treatment variable T . Задания являются одинаковыми по сложности и объему. Treatment variable принимает значение 1, если работник выполняет задания в условиях наличия дедлайна, и значение 0 – если определенного дедлайна нет. В таблице единицей во втором столбце обозначены стажеры.

Работник	Стажер	T=0	T=1	Treatment effect
Неваляшкин	0	3	5	
Мечталкин	0	2	3	
Душа компании	0	0	1	
Перфекционист	1	1	1	
Творческая натура	1	1	2	
Переживалкин	0	4	2	
Торопыжкин	1	4	7	
С улицы Бассейной	0	4	2	
Родом из сонного царства	1	2	5	

Найдите

1. treatment effect (TE) дедлайна для каждого работника (заполните последний столбик в таблице)
2. средний эффект воздействия (ATE) дедлайна
3. условный эффект воздействия дедлайна для стажеров и постоянных работников, не являющихся стажерами. Проинтерпретируйте содержательно полученные результаты. Можно ли говорить о наличии aggregation bias?

Задание 2.

Рассмотрим значения outcome variable Y – количества ответов ученика на занятия в школе – при разных значениях treatment variable T . Treatment variable принимает значение 1, если школьник осведомлен, что в классе проходит видеосъемка. Treatment variable принимает значение 0 для тех, кто попал в контрольную группу (индивиду ничего не говорится о видеосъемке).

Участник	Treatment	Количество ответов
1	1	6
2	0	4
3	1	7
4	1	2
5	1	4
6	0	8
7	1	4
8	0	3
9	0	5
10	0	0

Найдите

1. average treatment effect (ATE) - «naive estimate»
2. Можно ли полагаться на полученную в первом пункте оценку ATE, если известно, что первые пять участников – отличники в школе, а последние пять участников показывают более низкие результаты по успеваемости? Если нет, скорректируйте свой ответ.
3. Соблюдается ли в данном случае SUTVA? Свой ответ обоснуйте. Приведите проявления нарушения SUTVA, если таковые имеются.

Задание 3.

Представьте, что мы хотим выявить, как утренняя гимнастика влияет на сонливость днем. Участники исследования были разделены на 2 группы: группу воздействия (им нужно было выполнять гимнастику по утрам) и контрольную (они не выполняли утреннюю гимнастику). Среди участников были и взрослые, и дети. Ниже дана таблица совместного распределения по итогам исследования, где $Y = 1$ – отсутствие сонливости днем, $Y = 0$ – испытывал сонливость днем.

Группа / Подвыборка	Взрослые		Дети	
	$Y = 1$	$Y = 0$	$Y = 1$	$Y = 0$
Группа воздействия	0.15	0.225	0.1	0.025
Контрольная группа	0.0375	0.0875	0.2625	0.1125

Найдите

1. средний эффект воздействия утренней гимнастики на дневную сонливость по всей выборке в целом

2. средний эффект воздействия утренней гимнастики на дневную сонливость для взрослых
3. средний эффект воздействия утренней гимнастики на дневную сонливость для детей
4. сопоставьте полученные результаты, заметили ли Вы какие-нибудь расхождения? Если да, то с чем они связаны? Критически порассуждайте о процедуре проведенного эксперимента

Задание 4.

Ниже представлены результаты следующего эксперимента. Измерялся эффект обучения в дошкольном учреждении по инновационной программе на последующие успехи в первом классе. Outcome variable Y – успеваемость по тому или иному предмету (по 10-балльной шкале).

Предмет	Y (контрольная группа)	Naive ATE	Стандартная ошибка для Naive ATE
Математика	7.4	-0.5	0.408
Иностранный язык	6.8	1.2	0.125

1. Найдите значения outcome variable для treatment group
2. Сделайте вывод: значим ли naive ATE для математики? Для иностранного языка?

Задание 5. Самостоятельно придумайте примеры – контекст исследований – в которых встречаются сразу несколько ограничений для каузального вывода. *Примечание: ограничение еще не означает невозможность сделать каузальный вывод, однако в анализе требуются определенные поправки.*

- множественные причины и отложенный эффект
- spillover effect и слабая различимость концептов
- неоднородность каузального эффекта и отложенный эффект
- spillover effect и нарушение допущения постоянства воздействия

Задание 6. После проведения эксперимента участникам была отправлена по почте ссылка на опрос с целью выявления изменений, произошедших после участия в эксперименте.

1. В результате 30% среди участников контрольной группы и 45% среди участников treatment group не стали отвечать или не ответили на часть вопросов из анкеты. Поразмышляйте, о чем могут говорить эти значения non-response и как они могут повлиять на интерпретацию результатов опроса?
2. Рассмотрим другую ситуацию. Между Response-rates в группе воздействия и контрольной группе нет статистически значимых различий. Можно ли полагаться в оценивании среднего эффекта воздействия только на те ответы, которые доступны нам (то есть, тех респондентов, кто ответил на все вопросы)? Сделайте вывод, опираясь на информацию по доступной выборке (то есть, в таблице – характеристики тех, кто ответил) из следующей таблицы (в круглых скобках даны стандартные отклонения), если принимать, что пол и возраст – значимые характеристики для объяснения различий в скорости реакции. При необходимости приведите в подтверждение своего ответа расчеты. Всего участников в каждой группе 150.

Группа ответивших	Доля мужчин	Возраст
Treatment group	0.4	45 (20)
Control group	0.6	52 (15)