

## Семинар 4 (с решением)

## 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   | 2                |
| Мечталкин                | 0      | 2   | 3   | 1                |
| Душа компании            | 0      | 0   | 1   | 1                |
| Перфекционист            | 1      | 1   | 1   | 0                |
| Творческая натура        | 1      | 1   | 2   | 1                |
| Переживалкин             | 0      | 4   | 2   | -2               |
| Торопыжкин               | 1      | 4   | 7   | 3                |
| С улицы Бассейной        | 0      | 4   | 2   | -2               |
| Родом из сонного царства | 1      | 2   | 5   | 3                |

Найдите

1. treatment effect (TE) дедлайна для каждого работника (заполните последний столбик в таблице)

*Мы наблюдаем каждого работника в двух состояниях – в период до воздействия (без дедлайна) и при воздействии (при наличии дедлайна), поэтому можем в явном виде посчитать эффект воздействия как разность значения количества задач в условии дедлайна и количества задач при условии отсутствия дедлайна. Результаты приведены в последнем столбце.*

2. средний эффект воздействия (ATE) дедлайна

*Посчитаем среднее арифметическое по всем TE:  $\frac{7}{9}$*

3. условный эффект воздействия дедлайна для стажеров и постоянных работников, не являющихся стажерами. Проинтерпретируйте содержательно полученные результаты. Можно ли говорить о наличии aggregation bias?

*Сначала посчитаем среднее арифметическое по стажерам (1 – в столбце «Стажер»).*

$$CATE_1 = \frac{0 + 1 + 3 + 3}{4} = \frac{7}{4}$$

$$\text{А теперь для не-стажеров: } CATE_0 = \frac{2 + 1 + 1 - 2 - 2}{5} = 0$$

Так мы видим, что если для стажеров эффект наличия дедлайна на количество выполненных задач положительный, то для постоянных работников эффект отсутствует. Итоговый положительный эффект по всей выборке получен только благодаря стажерам, но при этом несправедливо распространять этот же результат на всех работников.

## Задание 2.

Рассмотрим значения outcome variable  $Y$  – количества ответов ученика на занятии 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»

В этом задании у нас другой формат данных: если ранее мы наблюдали один и тот же объект в двух состояниях, то теперь у нас две группы – экспериментальная и контрольная – которые составляют разные объекты. Если мы предполагаем, что участники эксперимента распределены на группы случайным образом, то можно полагать, что распределения контрольных характеристик сбалансированы, и можно говорить, что различия в среднем значении количества ответов – исключительно результат

треттмента, то есть, осведомленности о видеосъемке. Давайте посчитаем «наивную оценку»  $ATE$  как разницу в средних между экспериментальной группой и контрольной группой:

$$naiveATE = \frac{6 + 7 + 2 + 4 + 4}{5} - \frac{4 + 8 + 3 + 5 + 0}{5} = 4.6 - 4 = 0.6$$

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.

*В первой группе – отличников – четыре школьника попали в группу треттмент и только один – в контрольную группу. Во второй группе – не-отличников – напротив, только один человек – в группе воздействия, остальные – в контрольной группе. Да, выборка у нас, конечно, маленькая, чтобы делать радикальные выводы, но тем не менее, такое распределение может сигнализировать о том, что распределение на группы экспериментальную и контрольную не было рандомизированным, а следовательно, и условие «при прочих равных» вряд ли соблюдается. Поэтому  $ATE$  можно отдельно пересчитать по подгруппам отличников и не-отличников.*

3. Соблюдается ли в данном случае SUTVA? Свой ответ обоснуйте. Приведите проявления нарушения SUTVA, если таковые имеются.

*Первое допущение заключается в том, что воздействие для всех объектов одинаковое. В данном случае, если всем школьникам экспериментальной группы было сказано о проведении видеосъемки на занятии одновременно в одной и той же манере, то допущение можно считать верным.*

*Второе допущение – отсутствие «эффекта заражения». В данном случае такой эффект вполне возможен, так как школьники из экспериментальной и контрольной групп взаимодействуют. К примеру, если школьники из экспериментальной группы в силу того, что знают, что во время занятия идет видеосъемка, будут более активны и чаще отвечать, то вполне вероятно, что школьники из контрольной группы подхватят эту «волну» и также будут более активно включаться в дискуссию несмотря на то, что они не были проинформированы о видеосъемке. Это пример нарушения второго допущения.*

**Задание 3.**

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

*Среднее значение успеваемости по математике =  $7.4 - 0.5 = 6.9$*

*Среднее значение успеваемости по иностранному языку =  $6.8 + 1.2 = 8$*

2. Сделайте вывод: значим ли naive ATE для математики? Для иностранного языка?

*Можно построить доверительный интервал для naive ATE.*

*95%-ый CI для математики:*

$$-0.5 - 1.96 * 0.408 \leq \text{naiveATE}(\text{Maths}) \leq -0.5 + 1.96 * 0.408$$

$$-1.299 \leq \text{naiveATE}(\text{Maths}) \leq 0.299$$

*Доверительный интервал содержит 0, поэтому можно считать эффект инновационной программы на успеваемость по математике незначимым.*

*95%-ый CI для иностранного языка:*

$$1.2 - 1.96 * 0.125 \leq \text{naiveATE}(\text{ForeignLanguage}) \leq 1.2 + 1.96 * 0.125$$

$$0.955 \leq \text{naiveATE}(\text{ForeignLanguage}) \leq 1.445$$

*А вот для иностранного языка – эффект значимый положительный.*

**Задание 4.** 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?

*Мы не просто видим достаточно большой процент неответивших в двух группах, но и также то, что non-response rate отличается на 15%, если сравнить контрольную группу и группу воздействия. Такие различия могут свидетельствовать о том, что у участников группы воздействия были основания не отвечать на поставленные вопросы (к примеру, это могут те участники, которые не получили ожидаемого эффекта), и как следствие, сравнивать ответы по оставшей совокупности участников было бы нецелесообразно, есть риск получить смещенные результаты.*

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) |

*В этом случае мы наблюдаем обратную ситуацию. Значимых различий в response-rates нет. Однако это еще не повод утверждать, что группы являются сравнимыми. Для того, чтобы иметь возможность интерпретировать результат в терминах причинно-следственной связи, необходимо обеспечить прочие равные условия, то есть, сбалансированные распределения по контрольным характеристикам. В условии уже сказано, что пол и возраст - значимые*

контроли. Нам остается проверить, есть ли значимые различия в доле мужчин и среднем возрасте между экспериментальной группой и контрольной группой. Можно в очередной раз построить доверительные интервалы:

95%-ый CI для доли мужчин:

$$0.4 - 1.96 * \sqrt{\frac{0.4 * 0.6}{150}} \leq p(treatment) \leq 0.4 + 1.96 * \sqrt{\frac{0.4 * 0.6}{150}}$$

$$0.3216 \leq p(treatment) \leq 0.4784$$

$$0.6 - 1.96 * \sqrt{\frac{0.4 * 0.6}{150}} \leq p(control) \leq 0.6 + 1.96 * \sqrt{\frac{0.4 * 0.6}{150}}$$

$$0.5216 \leq p(control) \leq 0.6784$$

Доверительные интервалы для доли мужчин, построенные для экспериментальной и контрольной групп, не пересекаются, значит есть значимые различия. Это уже говорит о несбалансированности распределений и невозможности непосредственного сравнения.

95%-ый CI для среднего возраста:

$$45 - 1.96 * \frac{20}{\sqrt{150}} \leq E(treatment) \leq 45 + 1.96 * \frac{20}{\sqrt{150}}$$

$$41.79 \leq E(treatment) \leq 48.2$$

$$52 - 1.96 * \frac{15}{\sqrt{150}} \leq E(control) \leq 52 + 1.96 * \frac{15}{\sqrt{150}}$$

$$49.59 \leq E(control) \leq 54.4$$

Соответствующие доверительные интервалы для среднего возраста близки, однако все равно не пересекаются, что также говорит о несбалансированности распределений.

**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