

# Регрессионный анализ: панельные данные и каузальность

Каузальный вывод: введение

22 ноября 2022

# Планы на сегодня:

- статистическая инференция VS каузальный вывод
- начнем знакомиться с базовыми понятиями causal inference
- сформулируем основную проблему каузального вывода
- обозначим допущения для возможности каузального вывода
- обозначим преимущества эксперимента для выявления каузального эффекта

# «Классика» статистической инференции (1)

Какова задача статистической инференции?

# «Классика» статистической инференции (1)

Какова задача статистической инференции?

Перенести выводы с выборки на более широкую совокупность.

Как мы получаем информацию?

Мы делаем вероятностное суждение на основе

# «Классика» статистической инференции (1)

## Какова задача статистической инференции?

Перенести выводы с выборки на более широкую совокупность.

## Как мы получаем информацию?

Мы делаем вероятностное суждение на основе

- Оценивания параметров (точечного и интервального)

# «Классика» статистической инференции (1)

## Какова задача статистической инференции?

Перенести выводы с выборки на более широкую совокупность.

## Как мы получаем информацию?

Мы делаем вероятностное суждение на основе

- Оценивания параметров (точечного и интервального)
- Проверки статистических гипотез

# «Классика» статистической инференции (2)

## Примеры вопросов

# «Классика» статистической inferенции (2)

## Примеры вопросов

- Какой средний уровень математической тревожности среди студентов-политологов?



# «Классика» статистической инференции (2)

## Примеры вопросов

- Какой средний уровень математической тревожности среди студентов-политологов?
- В терминах сравнения. Одинаково ли в среднем тревожны относительно математических предметов студенты-магистранты и студенты бакалавриата?
- Совместная изменчивость. Связаны ли математическая тревожность и боязнь социальных взаимодействий? Если связаны, то какой характер взаимосвязи?

Примечание: Чем больше выборка, тем лучше для свойств оценок.

# Каузальный вывод

## Логика исследования

- ❶ идентификация причинно-следственного эффекта. Есть ли такая возможность?
- ❷ возвращаемся к статистической инференции как вспомогательному инструментарию
- ❸ каковы свойства полученных оценок?

Примечание: размер выборки на первом этапе далеко не решающий фактор!

## Что мы хотим узнать?

Выявить, что на что влияет. Какие факторы **ВЛИЯЮТ** на математическую тревожность? Претендуем на определение направления причинно-следственной связи.

# Основные понятия causal inference (1)

## Объект изучения

- Treatment (intervention) variable – переменная воздействия. Та переменная, которая оказывает эффект.
- Outcome variable – зависимая переменная (предполагаемый отклик в модели). На эту переменную оказывается воздействие.
- Treatment (intervention) effect

# Основные понятия causal inference (2)

## Treatment effect и его разновидности

- $TE = Y(T = 1, Z) - Y(T = 0, Z)$  В общем виде, каузальный эффект – разница в значении зависимой переменной при разных значениях treatment variable и при прочих равных условиях.

- Средний эффект воздействия.

$$ATE = E(Y|T = 1, Z) - E(Y|T = 0, Z)$$

- Условный средний эффект воздействия (ATE для подвыборки наблюдений, к примеру, эффект только для выборки мужчин):

$$ATE = E(Y|T = 1, G = g1, Z) - E(Y|T = 0, G = g1, Z)$$

# Желаемые условия

Что нам хотелось бы иметь для идентификации  
каузального эффекта?

Минимальный набор:

# Желаемые условия

Что нам хотелось бы иметь для идентификации каузального эффекта?

Минимальный набор:

- Изменчивость treatment variable + outcome variable
- Нам нужно иметь значения зависимой переменной сразу в нескольких состояниях: при условии того, что есть эффект воздействия, и нет эффекта воздействия
- Дело осложняется тем, что мы еще должны сохранять условие «при прочих равных»

Хотим оказаться в параллельном мире:  
counterfactual outcomes

What if ...

# Основная проблема каузального вывода



# Основная проблема каузального вывода

## Проблема пропущенных данных

- Можем ли мы, к примеру, одновременно наблюдать одну футбольную команду как выигравшую и проигравшую конкретный матч? Или одного и то же человека, как одновременно имеющего работу и безработного?
- Значения outcome variable при условии гипотетических состояний treatment variable называют counterfactual outcomes
- Проблема каузального вывода как проблема пропущенных данных

# Что нам дает эксперимент для выявления ТЕ?

## Эксперимент в лабораторных условиях

Мы можем исследовать влияние добавления соли на время закипания воды. Для этого

- 1 нужно обеспечить разные значения treatment variable (отсутствие соли, наличие соли в воде)
- 2 нужно обеспечить все остальные факторы, влияющие на время закипания воды, неизменными (объем воды, материал емкости и т.д. будут одинаковыми)

# Что нам дает эксперимент для выявления ТЕ?

## Эксперимент в лабораторных условиях

Мы можем исследовать влияние добавления соли на время закипания воды. Для этого

- 1 нужно обеспечить разные значения treatment variable (отсутствие соли, наличие соли в воде)
- 2 нужно обеспечить все остальные факторы, влияющие на время закипания воды, неизменными (объем воды, материал емкости и т.д. будут одинаковыми)

Примечание: Нам совершенно необязательно нужна большая выборка. Возможно работать и с двумя идентичными объектами, один из которых подвергается воздействию, а второй – нет.

# Ограничения в социальных науках

## Проблемы использования логики лабораторного эксперимента

# Ограничения в социальных науках

## Проблемы использования логики лабораторного эксперимента

- ❶ не можем подобрать два идентичных объекта (условие «прочие равные» не соблюдается)
- ❷ сталкиваемся со множеством посторонних факторов – confounding factors, эффект которых не можем разграничить
- ❸ трудности обеспечения идеальных лабораторных условий, как в естественных науках

# Как в социальных науках проводится эксперимент?

## Условия разделения на treatment и control groups

- Treatment group – группа, в которой объекты подвергаются воздействию ( $T = 1$ )
- Control group – контрольная группа, в которой объекты НЕ подвергаются воздействию ( $T = 0$ )
- Объекты делятся на группы случайным образом (рандомизация)
- Маленькие выборки в данном случае не спасут!

# Как в социальных науках проводится эксперимент?

## Условия разделения на treatment и control groups

- Treatment group – группа, в которой объекты подвергаются воздействию ( $T = 1$ )
- Control group – контрольная группа, в которой объекты НЕ подвергаются воздействию ( $T = 0$ )
- Объекты делятся на группы случайным образом (рандомизация)
- Маленькие выборки в данном случае не спасут!

Важно: Рандомизация и достаточный объем выборки обеспечивают возможность допущения об идентичности объектов в двух группах

# Допущения, упрощающие каузальный вывод

## SUTVA: Stable Unit Treatment Value Assumption

- 1 Единственная версия treatment: «воздействие на меня такое же, как на тебя»
- 2 Отсутствует «эффект заражения» (spillover effect): «мой результат – только мой результат»



# Одинаковый пинок, независимый полет



# Как оценить АТЕ на основе результатов эксперимента

## Naive average treatment effect

- $NaiveATE = E(Y|treatment) - E(Y|control)$
- То есть, Naive АТЕ – разница между средними значениями зависимой переменной в группе воздействия и контрольной группе

