



**Цикл
просветительских
лекций
от лучших
экономистов
России и мира**

Лекции проводятся
при поддержке



Ольга Кузьмина
профессор РЭШ

**Простой непростой анализ
ЭКОНОМИЧЕСКИХ ДАННЫХ:**
как делать корректные выводы и не попасть впросак

в рамках Цикла лекций по финансовой грамотности
Экономического лектория РЭШ-2019

Лекции проводятся
при поддержке

Анализ экономических данных: зачем?

- Экономисты анализируют данные, чтобы:

... пытаться предсказывать будущее

- Вероятность дефолта заёмщика в зависимости от его пола, возраста, кредитной истории и т.д.
- Уровень безработицы в стране в зависимости от исторических уровней безработицы и инфляции

... делать рекомендации на основе «эффекта воздействия»

- Поступить ли в престижный вуз в расчете на ощутимую прибавку в зарплате после его окончания?
- Нанять ли женщину или мужчину на руководящую позицию в компании, чтобы увеличить стоимость фирмы?

Анализ экономических данных: зачем?

- Экономисты анализируют данные, чтобы:

... пытаться предсказывать будущее

- Вероятность дефолта заёмщика в зависимости от его пола, возраста, кредитной истории и т.д.
- Уровень безработицы в стране в зависимости от исторических уровней безработицы и инфляции

... делать рекомендации на основе «эффекта воздействия»

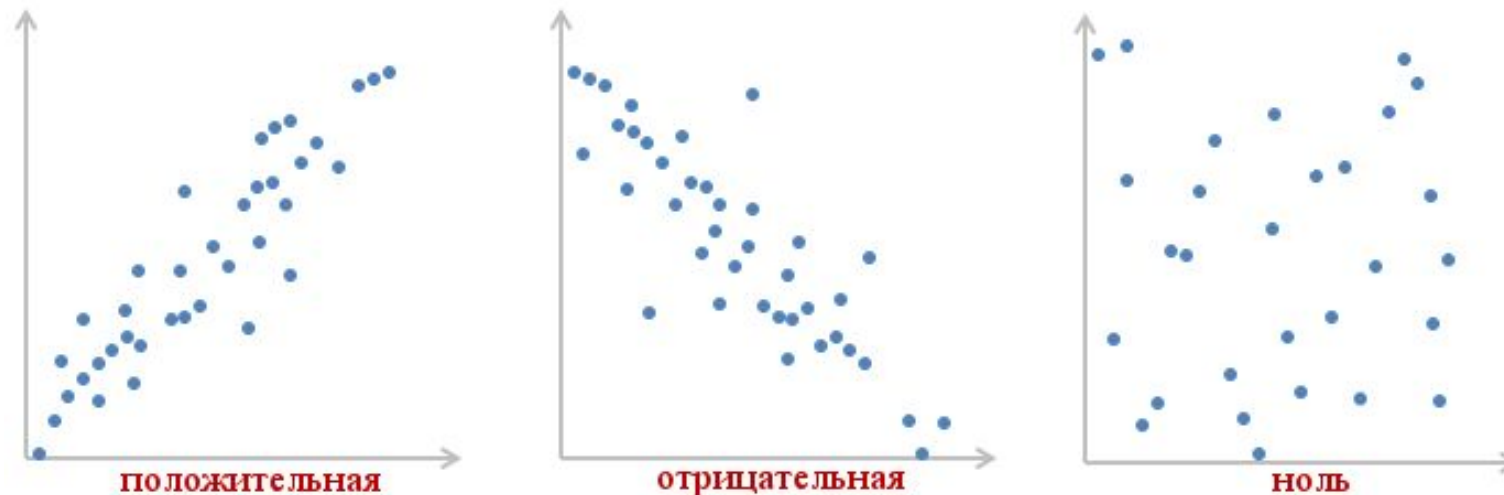
- Поступать ли в престижный вуз в расчете на ощутимую прибавку в зарплате после его окончания?

- Нанять ли женщину или мужчину на руководящую позицию в компании, чтобы увеличить стоимость фирмы?

сегодня

Анализ экономических данных: как?

- **Ингредиент 1:**
статистический анализ
- **Корреляция:**
Базовая мера статистической связи / закономерности между двумя переменными



Слепой перебор и ложные корреляции

**Число людей, утонувших в бассейнах,
коррелирует с
числом фильмов с Николасом Кейджем**



tylervigen.com



Слепой перебор и ложные корреляции

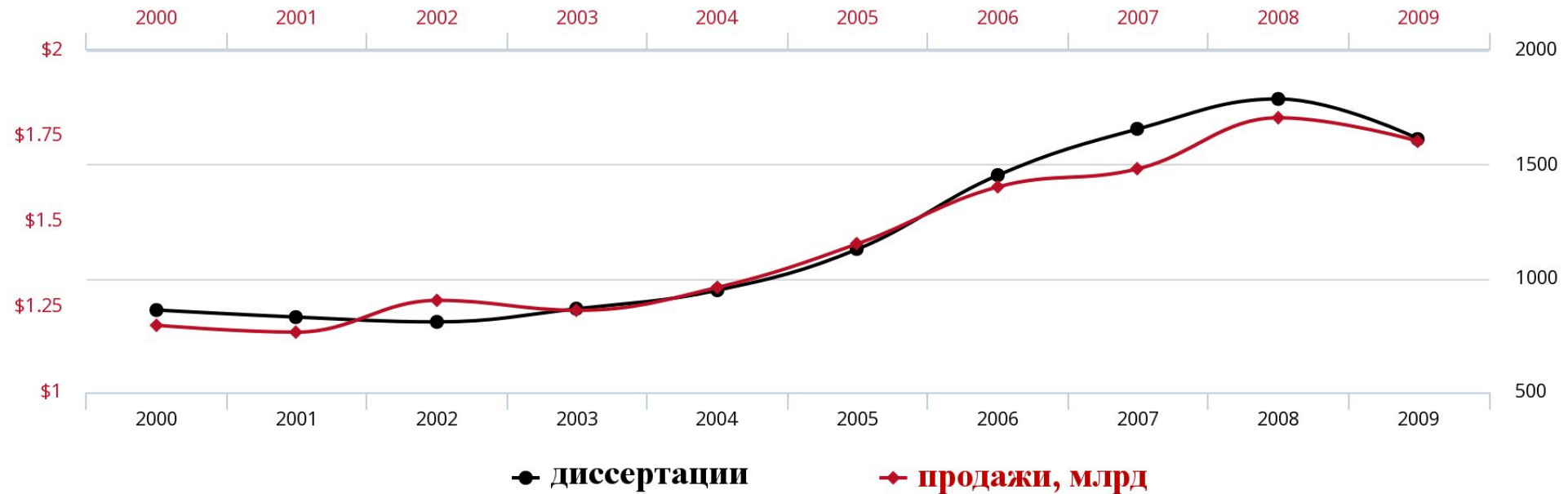
**Подушевое потребление сыра
коррелирует с
числом людей, умерших, запутавшись в простынях**



tylervigen.com

Слепой перебор и ложные корреляции

**Продажи компьютерных игр-аркад
коррелируют с
числом диссертаций по компьютерным наукам**



tylervigen.com



Слепой перебор и ложные корреляции



**Продажи японских автомобилей
коррелируют с
числом самоубийств за рулём**



tylervigen.com

От закономерностей к рекомендациям

- **Вывод 1: Не каждая корреляция – истинная**

Особенно в данных финансовых рынков

Усугубляется психологическими ловушками

- **Что делать:**

Задаваться адекватными экономическими вопросами

не перебирать «всё подряд»

избегать «дата майнинга»

Использовать адекватный статистический анализ

большая выборка

аккуратное обращение с временными рядами

Корреляция или зависимость?

- Допустим, анализ данных выявил **корреляцию**:
 - Люди с высоким уровнем образования зарабатывают в среднем больше, чем с низким
 - Компании с большей долей женщин в совете директоров стоят больше.
 - Люди, посещающие госпитали, чувствуют себя хуже, чем не посещающие

- Означает ли это, что:
 - Уровень образования действительно **влияет** на зарплаты?
 - Женщины лучше **управляют** компаниями?
 - Посещение госпиталя **ухудшает** здоровье людей?

Корреляция \neq причинно-следственная связь

- Для экономических рекомендаций критично выявить именно ПСС
- Корреляции симметрична,
ПСС – направлена
- Корреляция не учитывает других факторов,
ПСС – всегда «при прочих равных»
- **Ингредиент 2:**
регрессионный анализ / эконометрика
учёт дополнительных объясняющих факторов

Корреляция или зависимость?

- Люди с высоким уровнем образования зарабатывают в среднем больше, чем с низким

- Фактор способностей:

Может, изначально более способные дети получают больший уровень образования, а более способным и так платят больше?

- Фактор дохода родителей:

Может, они родились в более богатых семьях, а дети богатых родителей зарабатывают больше по другим причинам?

- Фактор ...

- Фактор ...

От закономерностей к рекомендациям

- **Вывод 2: Не каждая истинная корреляция выживет при учёте других факторов**
- **Что делать:**
*Использовать адекватный эконометрический анализ
учёт важных объясняющих факторов*

Можно ли учесть всё?

- Люди с более высоким уровнем образования зарабатывают в среднем больше, чем с более низким
 - Фактор мотивации:
Может, при том же уровне способностей и дохода родителей, более мотивированные дети получают больший уровень образования, а более мотивированным и так платят больше?
 - Фактор ...

Наблюдение во времени?

- Может быть, сравнить результаты тех, кто получил высокий уровень образования, ДО и ПОСЛЕ обучения?
 - *Такое сравнение (более-менее) учтёт факторы способностей, дохода родителей, мотивации, и т.д.*
 - *Но зарплата людей может меняться во времени и независимо от получения дополнительного образования:
Меняется рынок труда, возраст человека, и т.д.*

Наблюдение во времени?

- Может быть, сравнить результаты тех, кто получил высокий уровень образования, ДО и ПОСЛЕ обучения?
 - *Такое сравнение (более-менее) учтёт факторы способностей, дохода родителей, мотивации, и т.д.*
 - *Но зарплата людей может меняться во времени и независимо от получения дополнительного образования:
Меняется рынок труда, возраст человека, и т.д.*
- **Вывод 3: После – не значит вследствие**

Можно ли учесть всё?

■ НЕВОЗМОЖНО УЧЕСТЬ ВСЁ!

факторы могут быть:

- *неизвестны,*
- *не измеряемы,*
- *не наблюдаемы,*
- *непредсказуемо меняться во времени,*
- *может не быть доступных данных,*
- *...*

добровольный отбор участников, исходя из их выгоды:

*за дополнительным образованием могут пойти те,
кто ожидает от него большую прибавку к зарплате*



«Революция доверия» в экономике и финансах

- **Ингредиент 3:**
дизайн исследования / экспериментальный метод
 - На смену «наивному» эконометрическому анализу:
→ методы из биостатистики и медицины
- ~20-30 лет в экономике
~10-15 лет в финансах



«Революция доверия» в экономике и финансах

- Идеальный дизайн исследования:
рандомизированный эксперимент
(РКИ в медицине)
(А/В тестирование в маркетинге)
- Участники попадают в экспериментальную и контрольную группы случайным образом
- И это **гарантирует** одинаковость всех факторов между группами при достаточном объеме выборки

Бросание монетки: СТОИМОСТЬ И ЭТИКА

- Можем ли мы случайно выбрать / назначить:
 - Кого из желающих поступить в лучший ВУЗ России принять, а кому – отказать?
 - Кому из больных дать настоящее лекарство, а кому – плацебо?
 - Кому из пользователей сайта показать рекламный баннер на синем фоне, а кому – на красном?

Бросание монетки: СТОИМОСТЬ И ЭТИКА

- Рандомизированные эксперименты проводятся все чаще и чаще
- Но часто это слишком дорого или этически невозможно
 - В этих случаях «случайность» можно позаимствовать из природы и административных законов с помощью т.н. около-экспериментальных методов

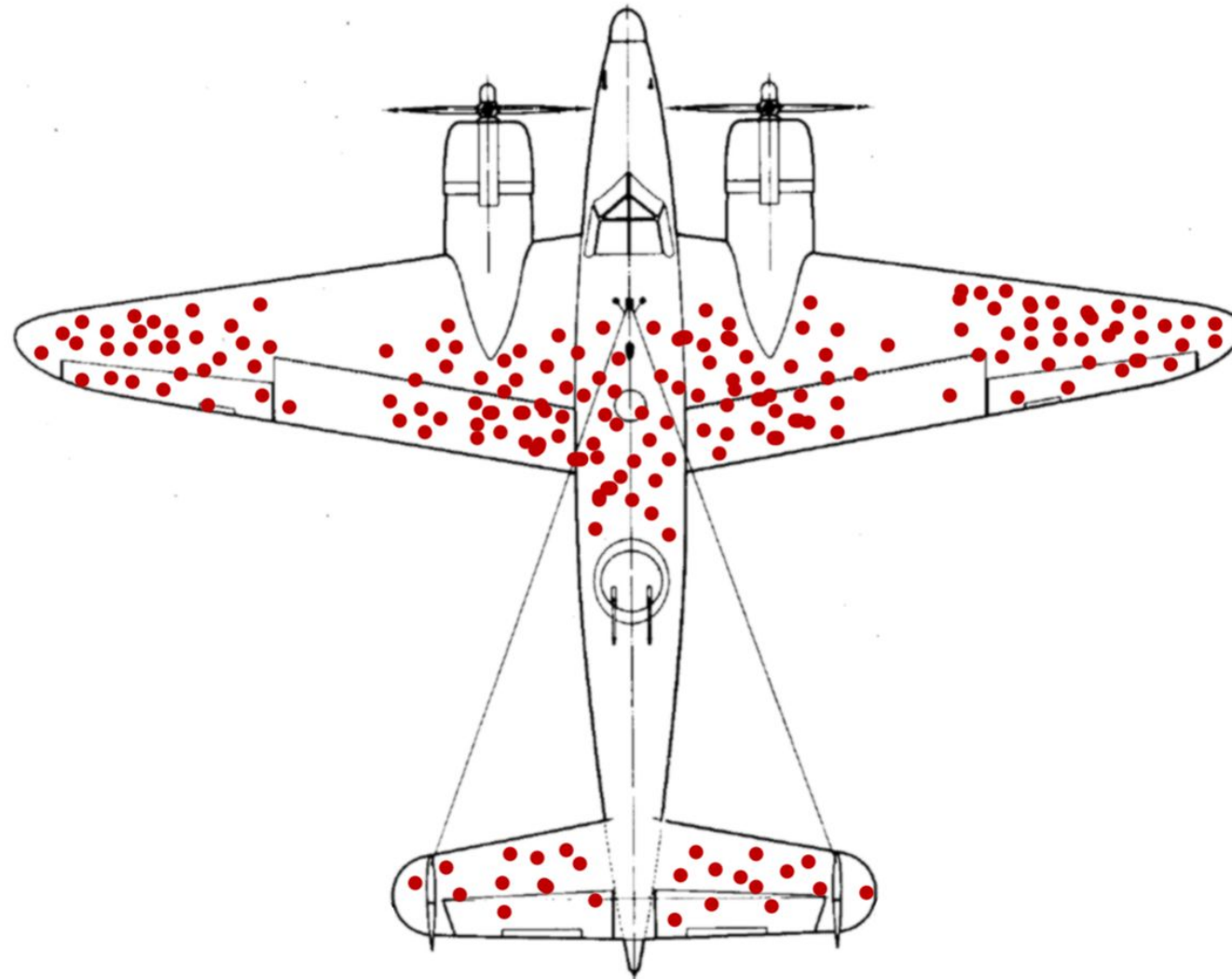
Отсев в экспериментах

- Даже хороший эксперимент может пострадать
 - (а плохое исследование – и подавно)

- Например, от **отсева**:
 - Пусть мы случайно выбрали, кого принять в лучший ВУЗ России, а кому – отказать.
 - Но в момент измерения зарплат выпускников, оказалось, что некоторые участники изначального эксперимента выпали из поля зрения:
 - эмигрировали, ушли в декрет, ушли в теневой сектор, и т. д.

- Если исчезновение из выборки связано с

Отсев: ошибка выжившего



Отсев в экспериментах

- Например:
 - Если эмигрируют чаще те, для кого эффект от образования большой, то:
*эффект от образования, измеренный на оставшихся, будет **занижен**.*
 - Если в теневой сектор уходят чаще те, для кого эффект от образования маленький, то:
*эффект от образования, измеренный на оставшихся, будет **завышен**.*

От закономерностей к рекомендациям

- Вывод 4а: **Ошибка выжившего сместит оценку**
- Вывод 4б: ... не всегда в предсказуемую сторону

- **Что делать:**

Использовать адекватный эконометрический анализ

Необходимо измерять результат для изначальной выборки

Если это невозможно, то использовать статистическую коррекцию на отсев

Что говорят исследования: образование и зарплаты

- Каждый дополнительный год образования в среднем увеличивает зарплату на:

Данные других стран:

- 5-8% (корреляция с учетом основных факторов)
- 7-10% (корреляция на «близнецах»)
- 8-15% (около-экспериментальные оценки ПСС)

Данные РФ:

- 8-10% (корреляция с учетом основных факторов)





Что говорят исследования: доля женщин и эффективность

- С увеличением доли женщин, стоимость/производительность компаний/команд:

Данные других стран:

- ↑ (корреляция)
- 0 или ↓ (корреляция с учётом основных факторов)
- 0 или ↓ (около-экспериментальные оценки)
- ↑ (экспериментальные оценки, с учетом отсева)

Данные РФ:

- ↑ (корреляция с учетом основных факторов)

Основные выводы

1. Не каждая корреляция – истинная
2. Не каждая истинная корреляция выживет при учёте других факторов
3. После – не значит вследствие
4. Остерегайтесь ошибки выжившего

От закономерностей к рекомендациям

- **Адекватный экономический вопрос**
Если сразу кажется странным, это повод задуматься, а не результат ли это поиска несуществующих закономерностей
- **Адекватный дизайн исследования**
В идеале: рандомизированный эксперимент или близкий дизайн исследования (инструментальные переменные, разрывный дизайн, и т.д.), на большой выборке
- **Адекватный эконометрический анализ**
*учёт дополнительных факторов (если требуется)
учёт ошибки выжившего и отбора в выборку (если требуется)*

Спасибо за внимание!

**Российская
экономическая
школа**

www.nes.ru

lectorium@nes.ru

Яндекс Дзен



**Лекции проводятся
при поддержке**

