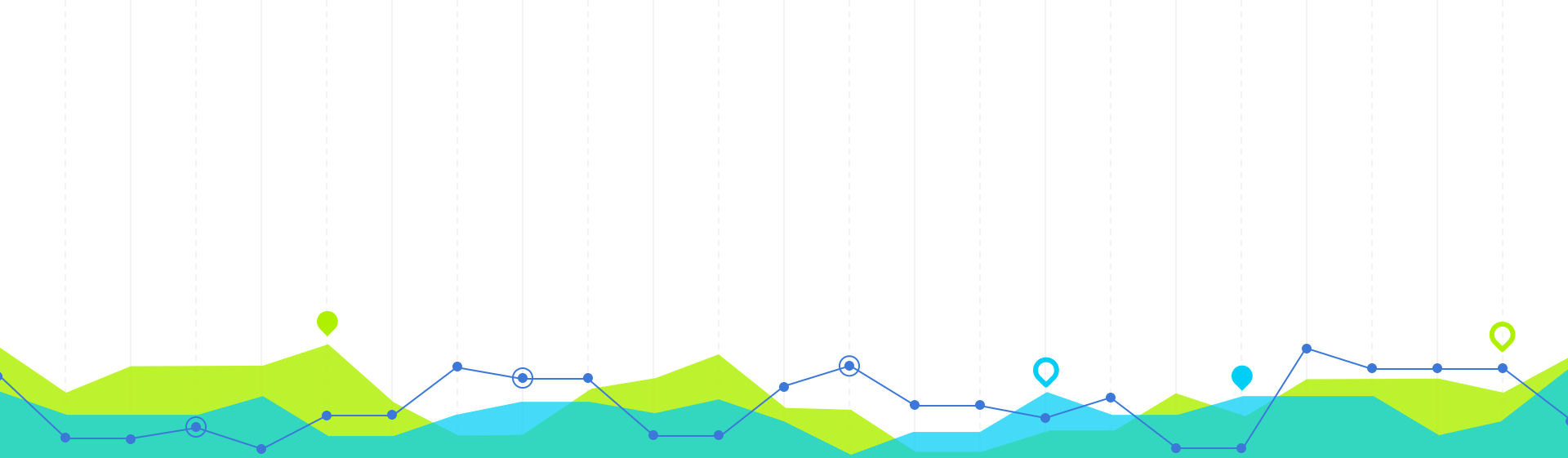




Основы биостатистики



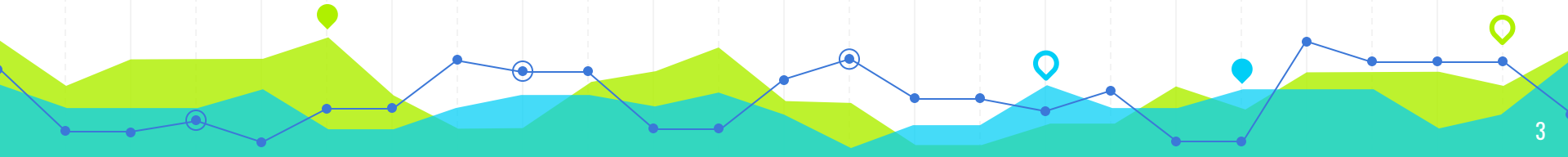
Определение биостатистики

1

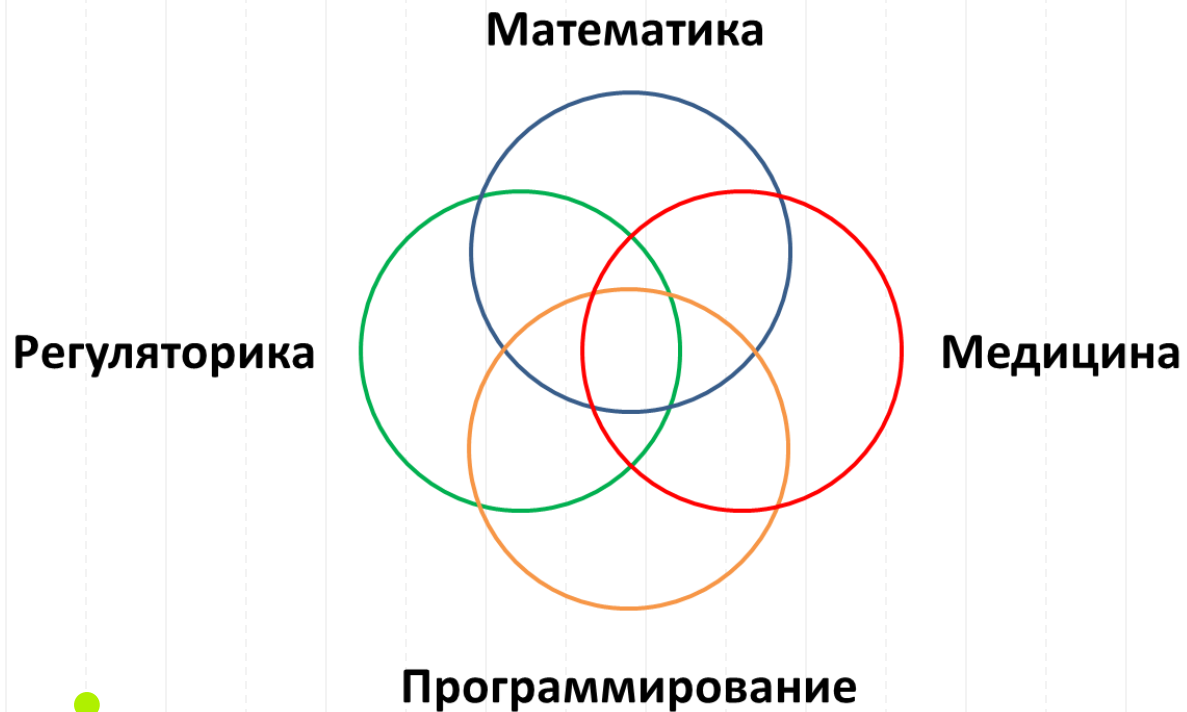


Биологическая статистика (биостатистика) — наука, связанная с созданием и применением статистических методов при проведении исследований в области биологии, медицины и общественного здравоохранения.

Синоним: биометрия, биометрика



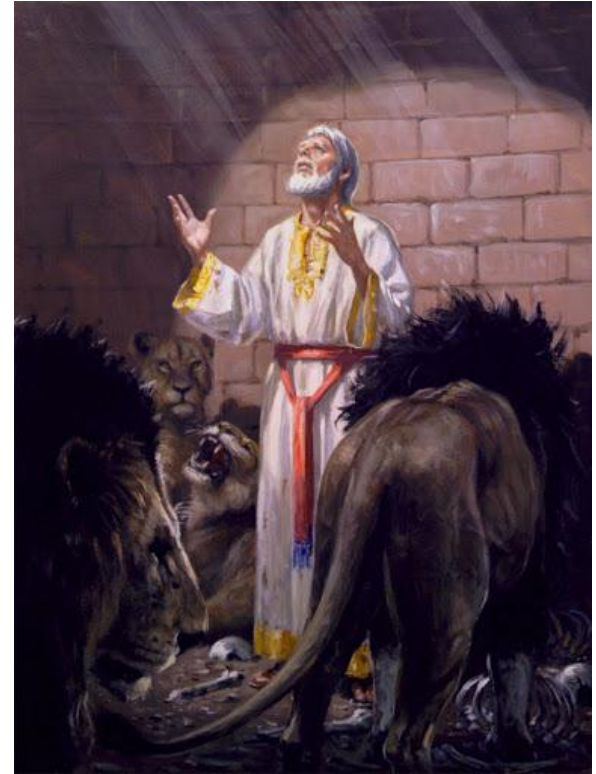
Современная биостатистика



История биостатистики

Книга Пророка Даниила, глава 1:

11. Тогда сказал Даниил Амелсару [...] :
12. сделай опыт над рабами твоими в течение десяти дней; пусть дают нам в пищу овощи и воду для питья;
13. и потом пусть явятся перед тобою лица наши и лица тех отроков, которые питаются царскою пищею, и затем поступай с рабами твоими, как увидишь.
14. Он послушался их в этом и испытывал их десять дней.
- 15 . По истечении же десяти дней лица их оказались красивее, и телом они были полнее всех тех отроков, которые питались царскими яствами.

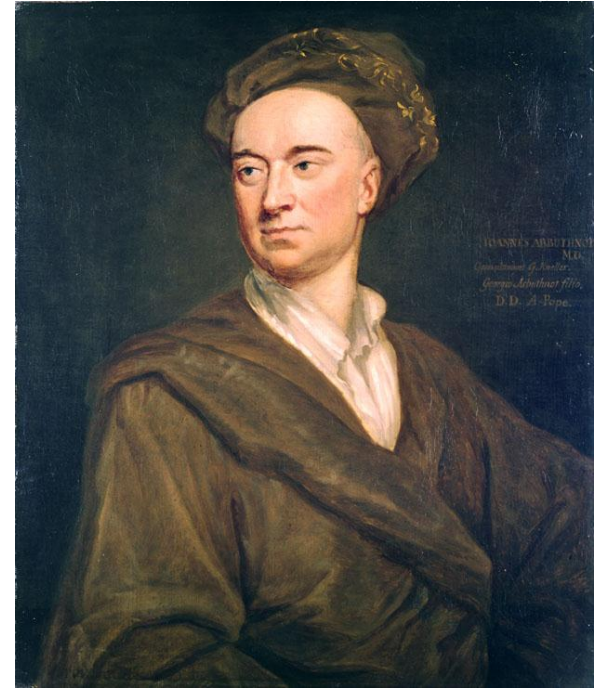


История биостатистики

До изобретения Дженнером вакцины от оспы, врачами практиковалась вариоляция. К сожалению, шанс смерти от вариоляции был очень высок (приблизительно 1:100).

Врач Джон Арбетнот 1722 году изучил статистику смертности в Лондоне выяснил, что шанс умереть от натуральной оспы во время эпидемии — приблизительно 1:10.

Таким образом, вариоляция снижала шансы смерти от оспы примерно в 10 раз.



Джон Арбетнот

История биостатистики

Голландский ученый **Даниил Бернулли** был одним из первых математиков, заинтересовавшихся исследованием эффективности медицинских процедур (18 век).

Ему удалось более строго доказать результаты **Джона Арбетнота** и учесть возраст жителей для повышения точности расчета эффективности.



Даниил Бернулли

История биостатистики

Французский врач и патолог Пьер-Шарль Александр Луи ввел в широкую медицинскую практику «численный метод».

Считал, что все методы лечения нужно предварительно проверять в ходе исследования на достаточно больших группах пациентов.

Применяя свой метод показал, что применение кровопускания при пневмонии было неэффективным и опасным (1835 г.).



Пьер-Шарль Александр Луи

История биостатистики

Гаварре - врач, получивший также инженерное образование, что позволило ему эффективно соединить идеи математики и медицины.

Автор первого учебника по медицинской статистике (1840 г):

Principes Generaux de Statistique Medicale



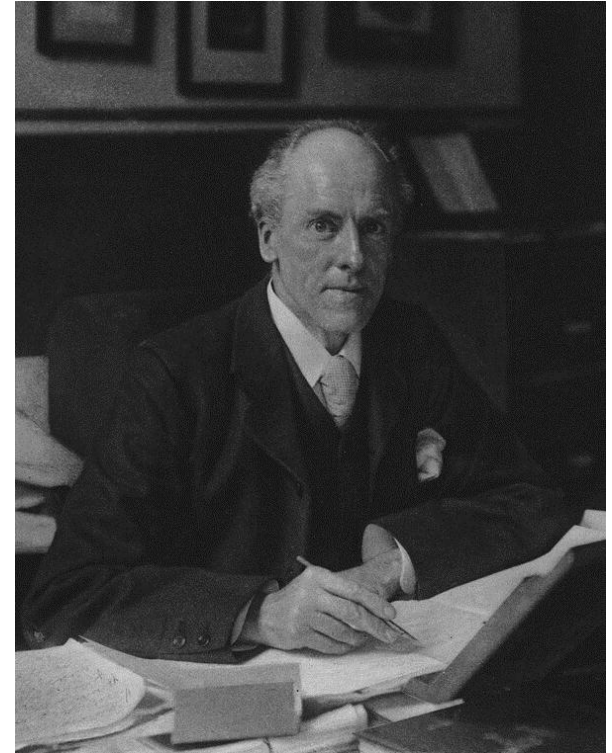
Луи Жюль Гаварре

История биостатистики

Карл Пирсон — профессиональный математик, разработавший множество методов статистики, использующихся в биологии и медицине.

Ввел в широкую практику такие понятия, как гистограмма, корреляция, стандартное отклонения, коэффициент асимметрии и пр. **Автор** критерия хи-квадрат.

Основал журнал «Биометрика», ставший площадкой для обмена идеями ученых со всего мира (1900 г).



Карл Пирсон

История биостатистики



Уильям Госсет



Рональд Фишер



Ежи Нейман

Биостатистика в современном мире

Биостатистика получила широкое распространение с развитием доказательной медицины (evidence-based medicine) и клинических испытаний (clinical trials).

В этой области методы статистики регулируются правилами ICH (The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use).



Клиническое исследование (КИ)

Любое исследование/испытание, проводимое с участием человека в качестве субъекта для выявления или подтверждения клинических и/или фармакологических эффектов исследуемых препаратов и/или выявления нежелательных реакций на исследуемые препараты, и/или изучения их всасывания, распределения, метаболизма и выведения с целью выполнения оценки безопасности и/или эффективности.

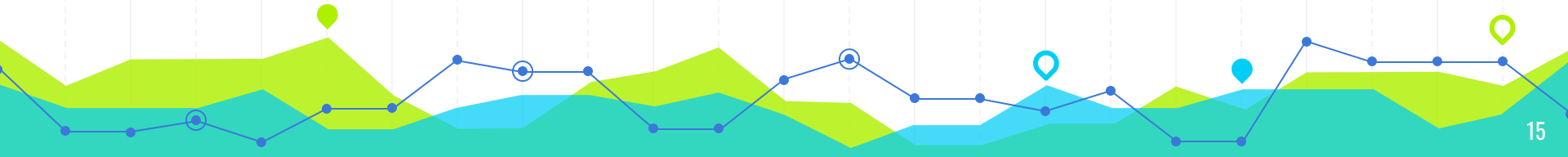
**Правила надлежащей клинической практики
Евразийского экономического союза**

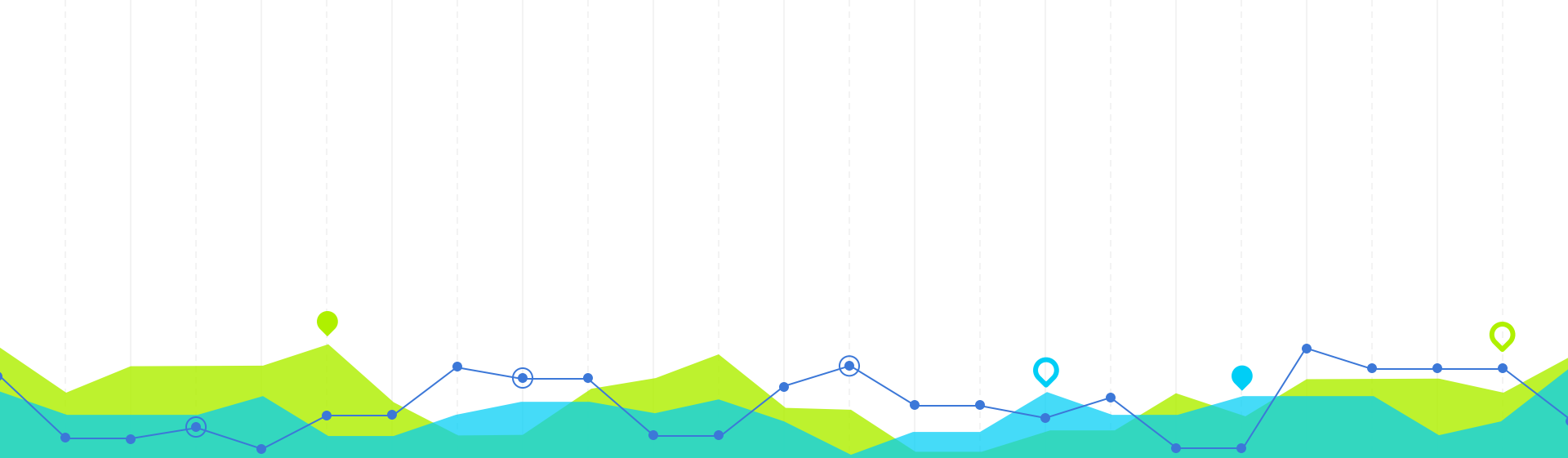
Помимо КИ биостатистика используется для исследований

- лабораторных животных,
- культур клеток;
- химических реакций
- и пр.

“

Признак исследуемого объекта (англ. - variables, features)
— свойство объекта, которое может быть объективно измерено в ходе исследования.





Последовательность этапов исследования

2

7 основных шагов в статистическом эксперименте

1. Определить множество исследуемых объектов (пациентов).
 - Определить критерии включения, не-включения, исключения (inclusion, non-inclusion, exclusion)



Лица, больные
заболеванием X

7 основных шагов в статистическом эксперименте

2. Сформулировать «Основной Вопрос Исследования»

● Выбрать воздействие и исход (exposure, outcome).

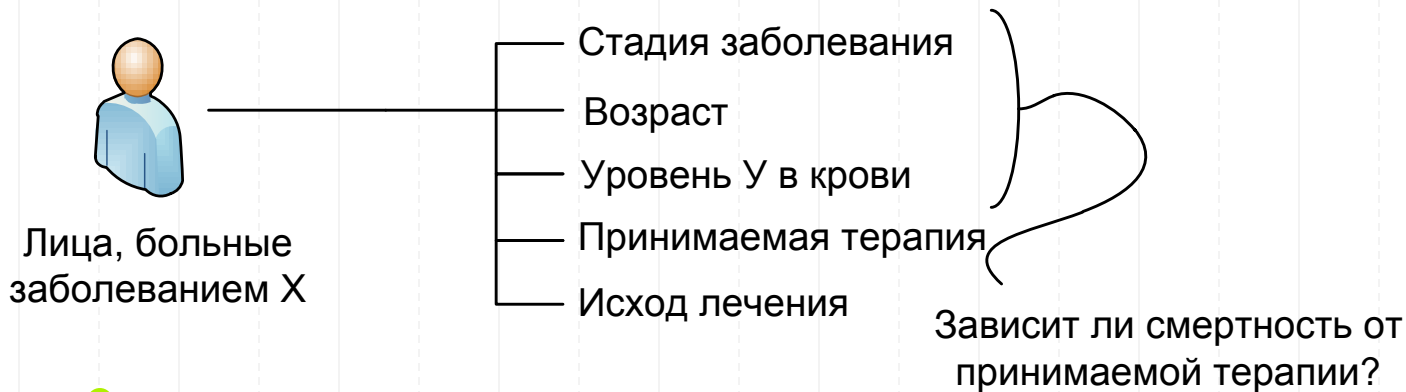


Лица, больные
заболеванием X

Зависит ли смертность от
принимаемой терапии?

7 основных шагов в статистическом эксперименте

3. Выделить признаки объектов, способные также повлиять на результат



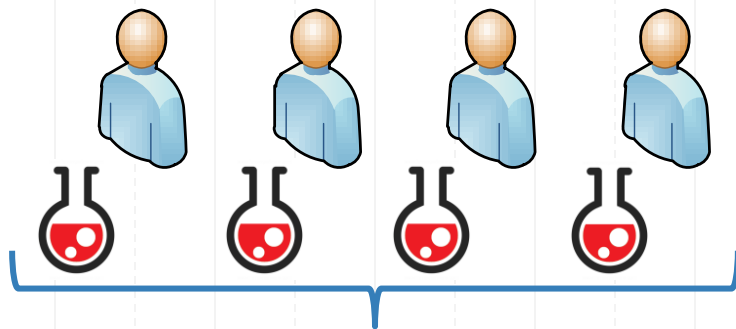
7 основных шагов в статистическом эксперименте

4. Выбрать дизайн исследования

Параллельное
Диагностическое
Рандомизированное Ретроспективное
Случай-контроль Скрининговое
Обсервационное Плацебо-контролируемое
Профилактическое Многоцентровое Пилотное
Перекрестное Одноцентровое Слепое
Одномоментное Интервенционное
Лонгитудинальное Наблюдательное
Двойное Открытое
Проспективное Неинтервенционное
Когортное

7 основных шагов в статистическом эксперименте

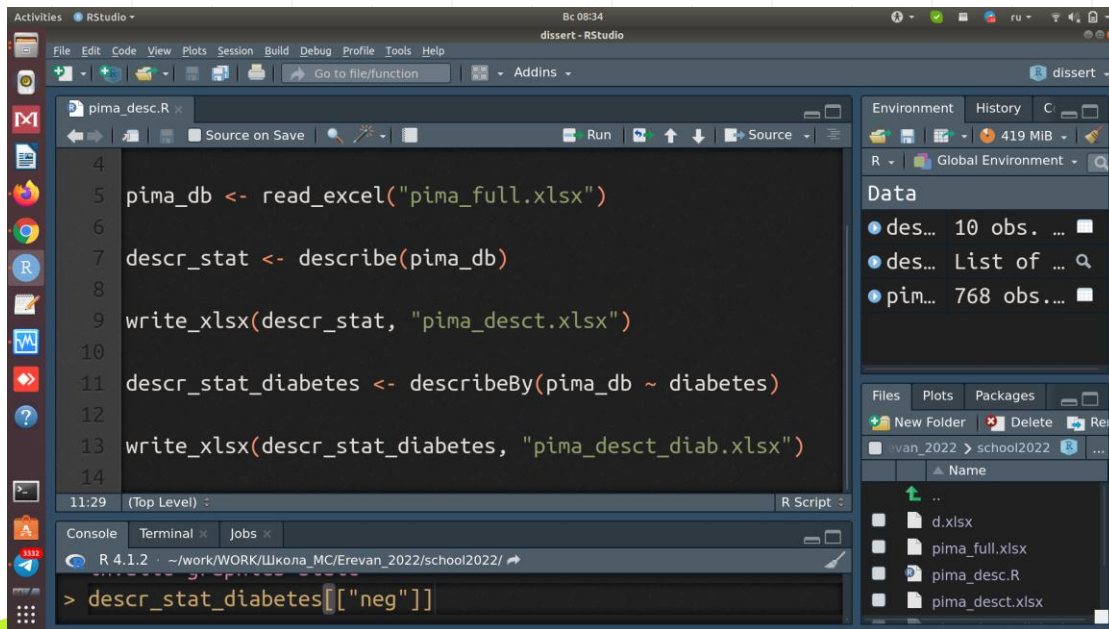
5. Сформировать выборку, содержащую значения признаков для каждого из пациентов.



Лица, больные заболеванием X

7 основных шагов в статистическом эксперименте

6. Произвести статистический анализ



The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for reading an Excel file, describing it, and saving the results.
- Environment:** Shows the current environment with 419 MiB of memory used.
- Data:** Lists the data objects: 'des...' (10 obs.), 'des...' (List of ...), and 'pim...' (768 obs.).
- Files:** Shows the file explorer with a folder named 'school2022' containing files like 'd.xlsx', 'pima_full.xlsx', 'pima_desc.R', and 'pima_desc.xlsx'.
- Console:** Shows the command prompt with the command `> descr_stat_diabetes[["neg"]]`.

```
4  
5 pima_db <- read_excel("pima_full.xlsx")  
6  
7 descr_stat <- describe(pima_db)  
8  
9 write_xlsx(descr_stat, "pima_desc.xlsx")  
10  
11 descr_stat_diabetes <- describeBy(pima_db ~ diabetes)  
12  
13 write_xlsx(descr_stat_diabetes, "pima_desc_diab.xlsx")  
14  
11:29 (Top Level) : R Script :  
Console Terminal Jobs :  
R 4.1.2 ~/work/WORK/Школа_MC/Erevan_2022/school2022/ >  
> descr_stat_diabetes[["neg"]]
```

7 основных шагов в статистическом эксперименте

7. Проинтерпретировать результаты



7 основных шагов в статистическом эксперименте

1. Определить множество исследуемых объектов (пациентов).
2. Сформулировать «основной вопрос исследования».
3. Выделить признаки объектов, способные повлиять на результат.
4. Выбрать дизайн исследования
5. Сформировать выборку, содержащую значения признаков для каждого из пациентов.
6. Произвести статистический анализ.
7. Проинтерпретировать результаты.

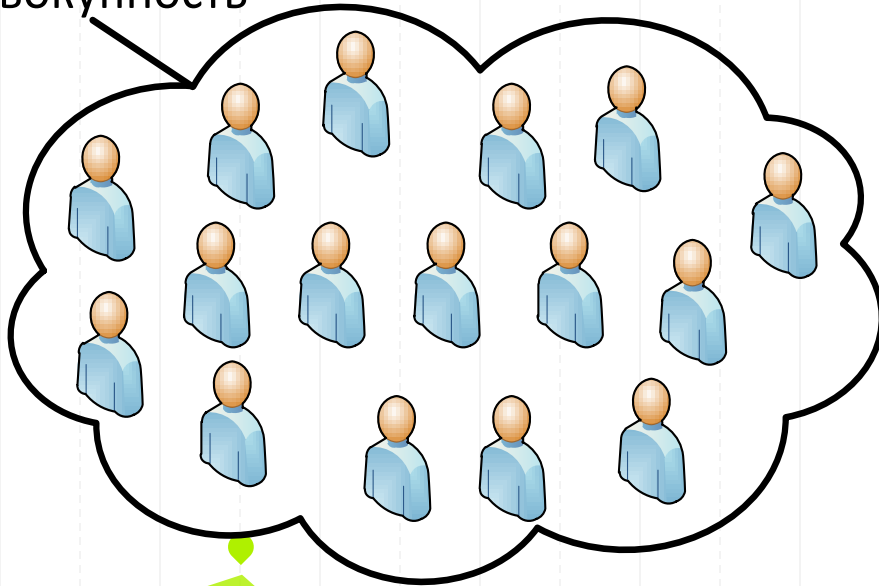


Выборки и их свойства

3

Генеральная совокупность

Генеральная
совокупность



- Множество **всех** пациентов, имеющих исследуемое заболевание (фенотип)
- Зачастую это множество **не ограничивается** только ныне живущими
- Анализ всех возможных пациентов из генеральной совокупности невозможен как **практически**, так и **теоретически**

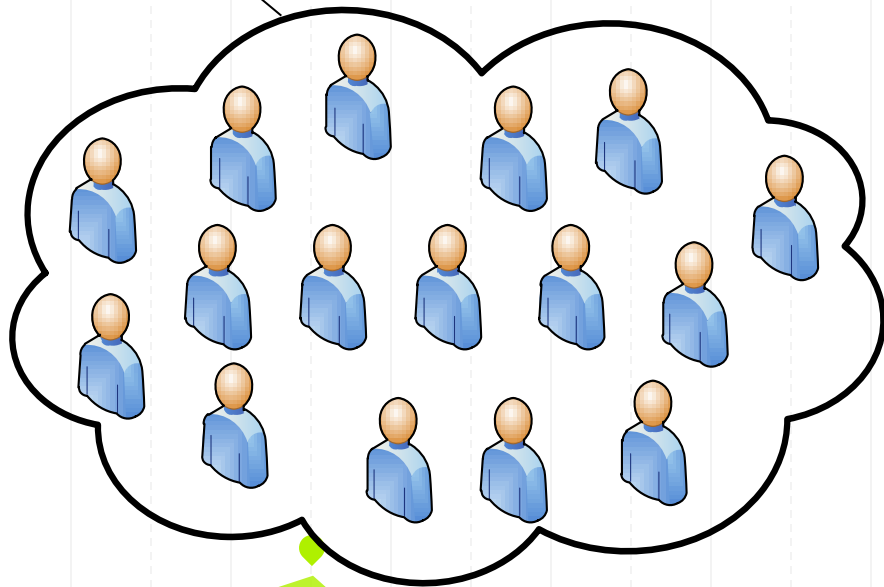
Но что, если бы обладали бесконечными сверхъестественными возможностями?

Мы хотим понять, лучше ли новая терапия чем плацебо.

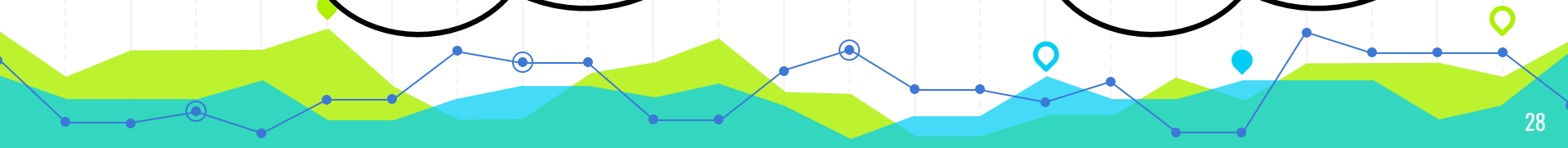
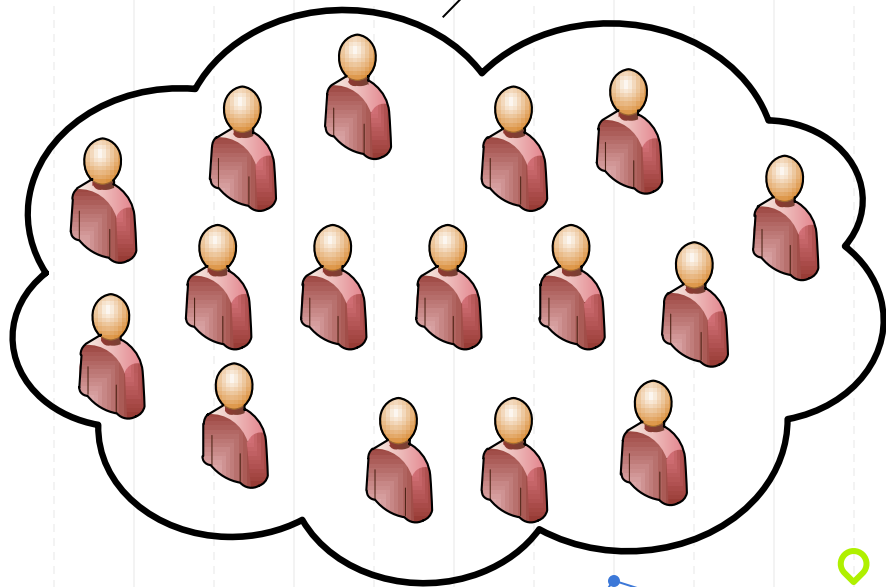


Но что, если бы обладали бесконечными сверхъестественными возможностями?

Вселенная 1:
всем даем плацебо



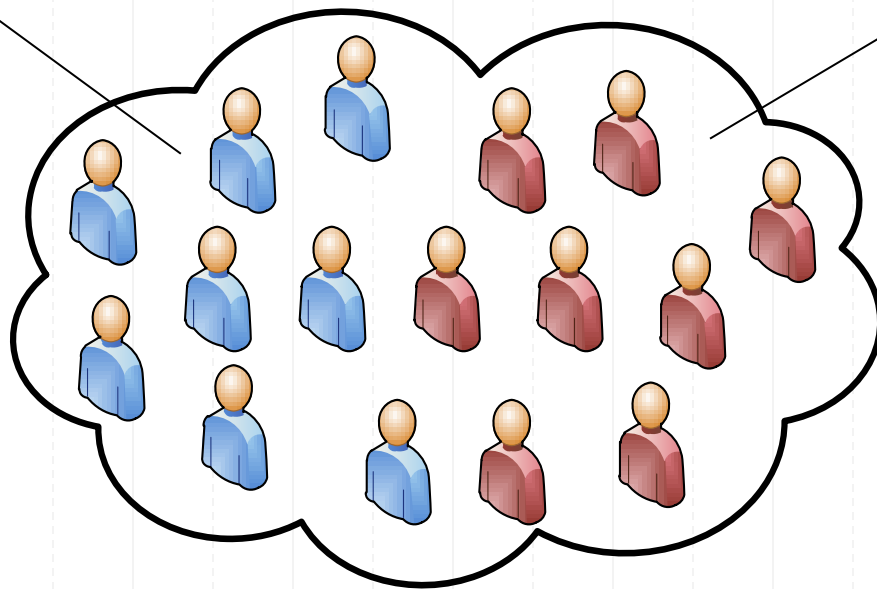
Вселенная 2: всем даем
новый препарат



Но что, если бы обладали бесконечными, **но не** сверхъестественными возможностями?

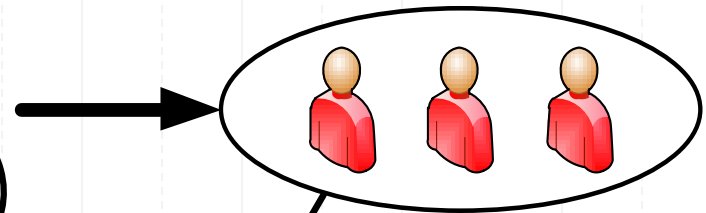
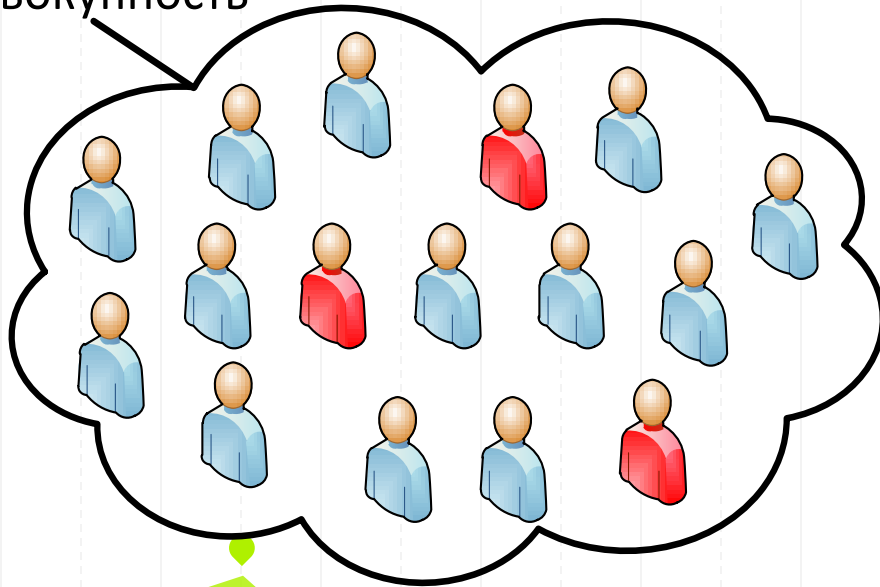
Случайная половина ГС:
всем даем плацебо

Случайная половина ГС:
всем даем новый препарат



Генеральная совокупность и выборка

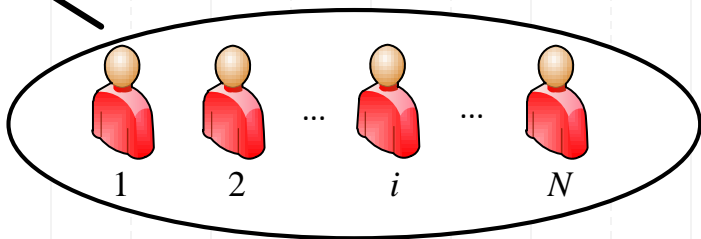
Генеральная
совокупность



Выборка

Элементы выборки

Выборка



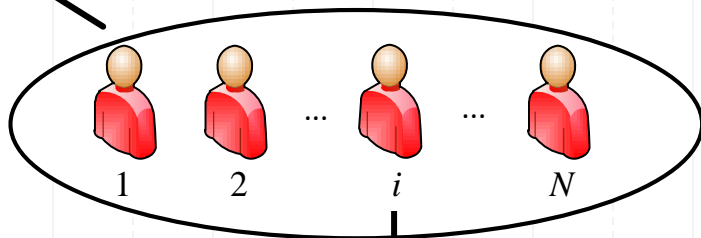
N – объем выборки

i – номер произвольного элемента выборки
($i = 1, 2, \dots, N$)

Как правило, порядок следования элементов не важен.

Элементы выборки

Выборка



Каждый элемент выборки обладает набором относящихся к задаче признаков (свойств)

1. ГЛАВНЫЕ ПРИЗНАКИ

2. Пол
3. Возраст
4. Фаза заболевания
5. Принимаемая терапия

...

Список признаков элемента выборки

Представление выборки в табличном виде

ID	Пол	Возраст	Фаза	Терапия	Наступление ремиссии
1	М	30	I	ЛТ	Да
2	Ж	58	II	ХТ	Да
3	Ж	42	IV	ИТ	Нет
...
N	М	70	I	ЛТ	Да

Выборка

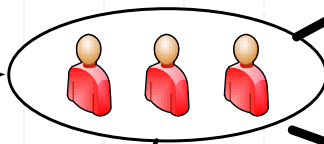
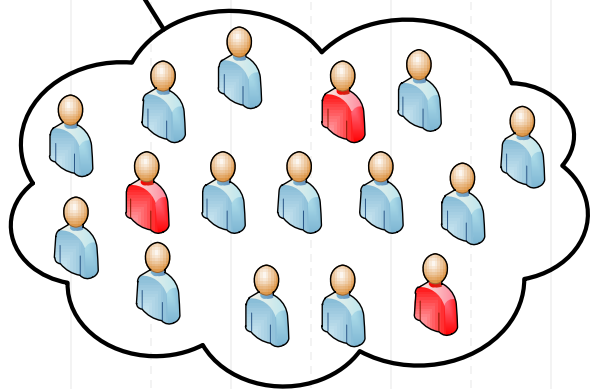
Элементы
выборки

Главные признаки
(воздействие и исход)

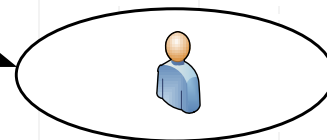
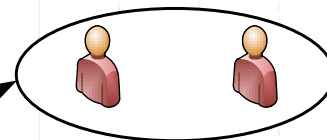
Дополнительные
признаки

Как все происходит в реальной жизни?

Генеральная совокупность
(пациенты с диагнозом X)



Выборка



Одномерные и многомерные данные

- Каждый элемент выборки обладает набором относящихся к задаче признаков (свойств)
- В набор признаков должно быть включено **все, что может влиять на главный анализируемый признак.**
- Если признак ровно один, то это – выборка **одномерных** данных (содержат только главный признак), если более одного – **многомерных**.

Когда данные могут быть одномерными...

- Выборка 1: Трансплантация костного мозга (ТКМ) по протоколу №1
- Выборка 2: ТКМ по протоколу №2
- Главный признак — продолжительность жизни после трансплантации костного мозга.
- Сопутствующие признаки:
 - Возраст
 - Основное заболевание
 - Стадия заболевания
 - Рецидивы

Когда данные могут быть одномерными...

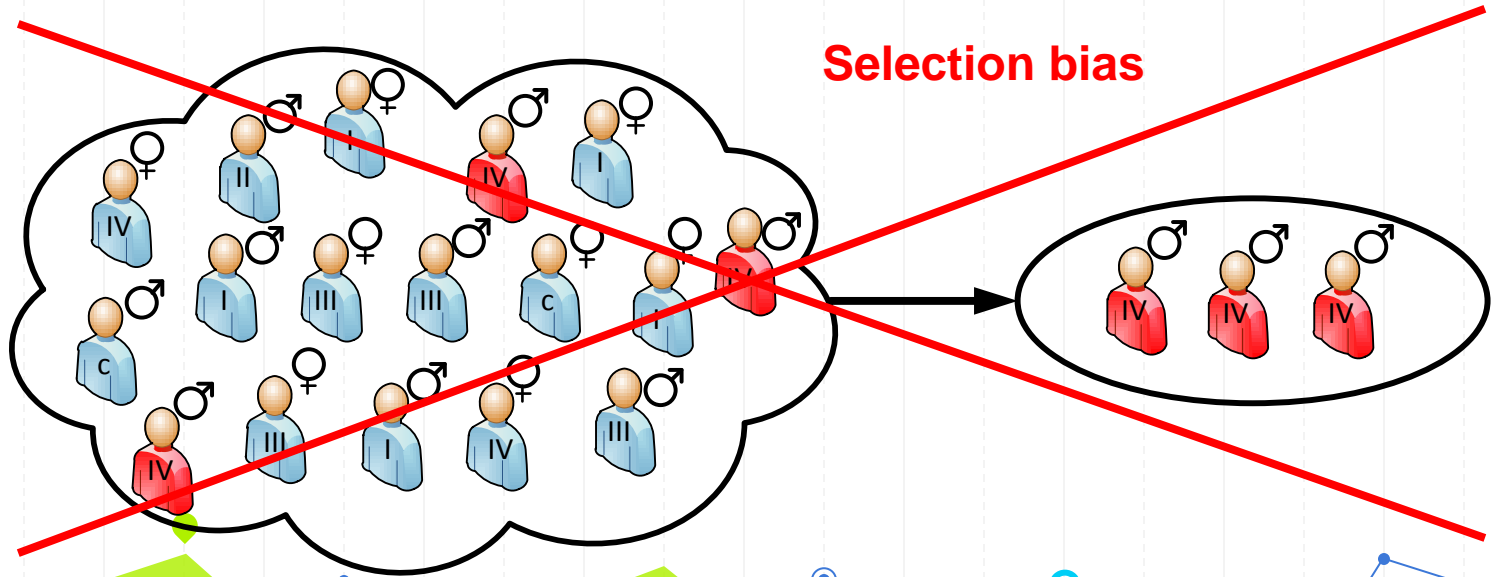
- Выборка 1: ТКМ по протоколу №1, возраст 18-25 лет, ХМЛ, фаза акселерации, первичная ТКМ
- Выборка 2: ТКМ по протоколу №2, возраст 18-25 лет, ХМЛ, фаза акселерации, первичная ТКМ
- Главный признак — продолжительность жизни после трансплантации костного мозга.
- Сопутствующие признаки:
 - **НЕТ**

Не надо бояться сузить область исследования!



Репрезентативность

Выборка должна отражать основные свойства генеральной совокупности (**репрезентативность**).



Чем больше независимых признаков анализируется, тем сложнее сформировать выборку с соблюдением репрезентативности!

Не надо бояться сузить область исследования!



Пример

- Решили написать статью **«Эффективность X в лечении лимфомы Ходжкина на стадии I»**.
- Набрали для статьи 54 пациента, но рецензент написал, что этого мало.
- Коллега предложил 6 своих пациентов, но у них ЛХ на стадии II.
- Можно ли написать статью с названием **«Эффективность X в лечении лимфомы Ходжкина на ранних стадиях»?**

Независимость

- Признаки одного элемента выборки были сформированы независимо от признаков другого элемента выборки.
- Каждый элемент выборки не испытывал влияния других элементов.
- Один из частых источников ошибок — несоблюдение принципа независимости в данных.



Пример № 0

- Выборка состоит из школьников, пишущих контрольную работу.
- Главный признак – набранные за контрольную баллы.
- Являются ли баллы разных школьников независимыми?

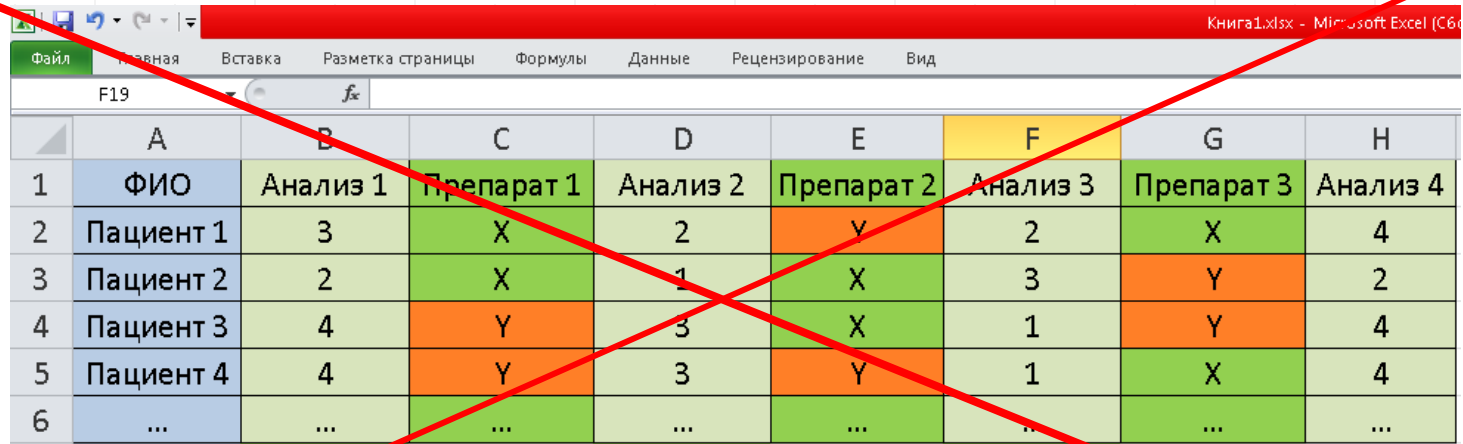


Пример №1

- Исследовалось время приобретения внутрибольничной инфекции у пациентов неинфекционного профиля.
- Сформированная выборка пациентов была размещена в отдельном крыле здания больницы, в палатах по 2-4 человека.
- Что не так?



Пример №2



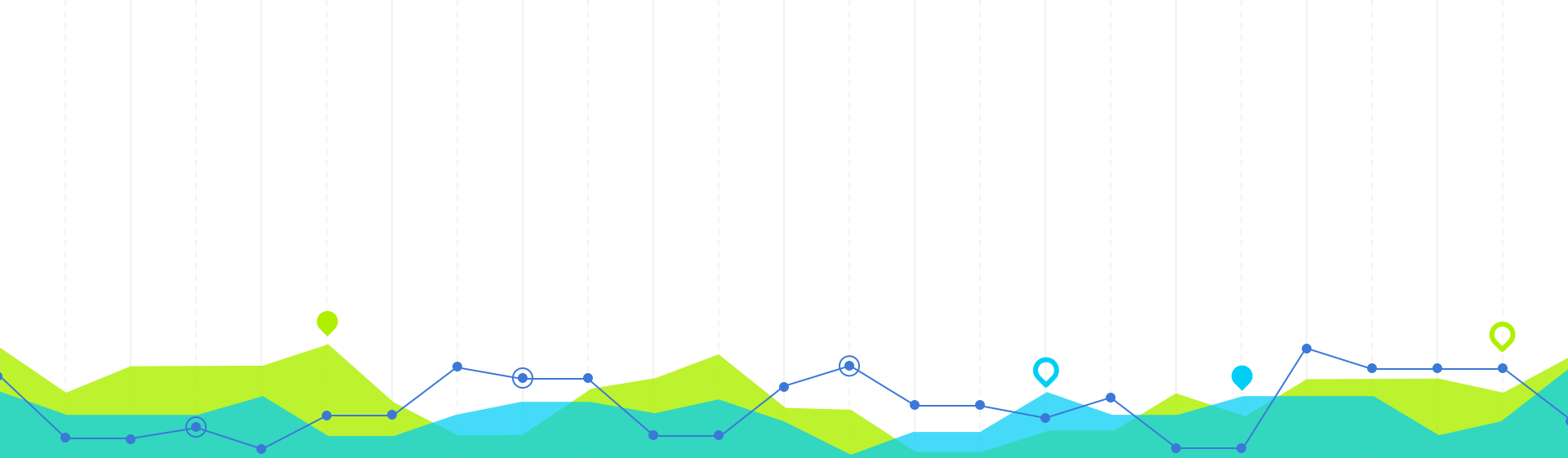
	A	B	C	D	E	F	G	H
1	ФИО	Анализ 1	Препарат 1	Анализ 2	Препарат 2	Анализ 3	Препарат 3	Анализ 4
2	Пациент 1	3	X	2	Y	2	X	4
3	Пациент 2	2	X	1	X	3	Y	2
4	Пациент 3	4	Y	3	X	1	Y	4
5	Пациент 4	4	Y	3	Y	1	X	4
6

Выборка 1 (по препарату X): -1, +2, -1, +2, -2, +3

Выборка 2 (по препарату Y): 0, -1, -1, +3, -1, -2

Хорошая выборка – это репрезентативная выборка из независимых случайных элементов.





Виды клинических исследований

4

Два ключевых понятия

Воздействие (англ. exposure) — признак, который описывает некий фактор, влияющий на пациента.

Исход (англ. outcome) — признак, который отражает результат воздействия.



Существует много видов клинических исследований

Параллельное
Диагностическое
Рандомизированное Ретроспективное
Случай-контроль Скрининговое
Обсервационное Платцебо-контролируемое
Профилактическое Многоцентровое Пилотное
Перекрестное Одноцентровое Слепое
Одномоментное Интервенционное
Лонгитудальное Наблюдательное
Двойное Открытое
Проспективное Неинтервенционное
Когортное

3 основных критерия к выбору вида клинического исследования

- Располагаемое время и доступный бюджет
- Возможность вмешательства в процессы
- Точность получаемых результатов

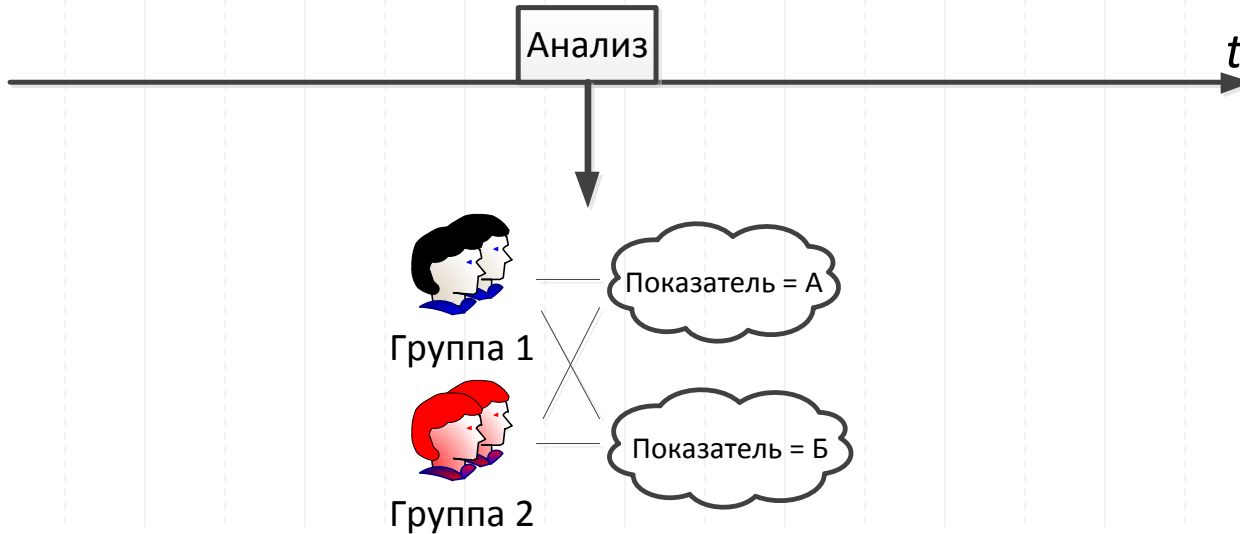
1. Максимально сжатые сроки

Поперечное (одномоментное, кросс-секционное) исследование.

Англ. – cross-sectional study.



Поперечное исследование



1. Набираем группу добровольцев, подходящих под критерии включения/невключения.
2. Проводим анкетирование (разовое обследование/забор анализов).
3. Находим взаимосвязи.

Поперечное исследование (пример)

- Хотим понять, с какими факторами ассоциирована бессонница
- Проводим опрос, в котором предлагаем ответить на ряд вопросов:
 - Пол, возраст, ИМТ
 - Доход, особенности рабочего дня
 - Страдает ли респондент бессонницей
- Сравниваем частоту бессонницы между полами, возрастными группами, уровнями доходов и пр.



Поперечное исследование (реальные примеры)

Вчера, 20:06 • 5 минут • Денис Гордеев

2 677

**Ученые доказали, что
обед с коллегами повышает
работоспособность**

09 декабря 2015, 17:50

**Ученые доказали, что женщины справедливее
мужчин, когда речь идет о заботе и
беспристрастности** 3.1

Интересно, что в России — о
различия.

04:18, 28 сентября 2020

Главная / Пресс-релизы / Ученые доказали, что фитнес повышает когнитивные функции

**Ученые доказали, что фитнес повышает
когнитивные функции**

Спортивный клуб АТЛЕТик 29 Сентября 2020 13:23

☆☆☆☆

**Ученые доказали влияние женитьбы
на снижение веса**



1



3



Добавить в «Мою Ленту»

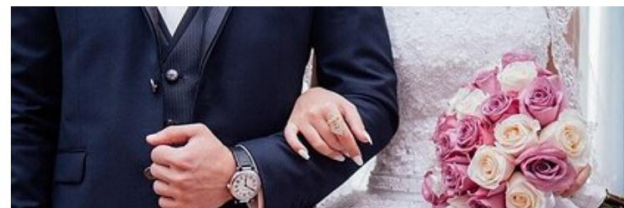


Фото: Pixabay.com

Итого

Плюсы:

- Дешевизна и скорость
- Можно использовать для формирования первичных идей для дальнейшей проработки

Минусы:

- Невозможность проанализировать временные характеристики
- Невозможность установить причинно-следственную связь



2. Данные собирались какое-то время в прошлом

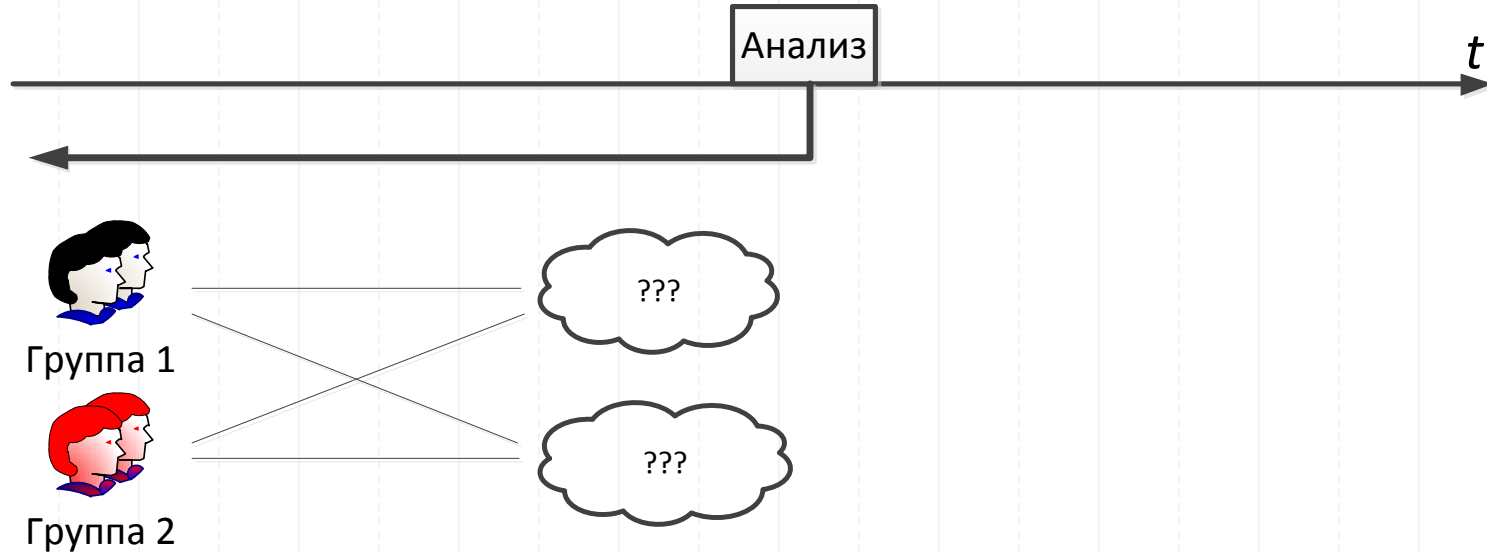
Ретроспективное исследование



Ретроспективное исследование

1. Набор группы и наблюдение за ней происходило в прошлом
2. Как правило, данные собирались ситуативно, качество БД невысокое
3. Много времени уходит на фильтрацию и предобработку данных
4. Два принципа формирования выборки:
 1. Когортное
 2. Случай-контроль

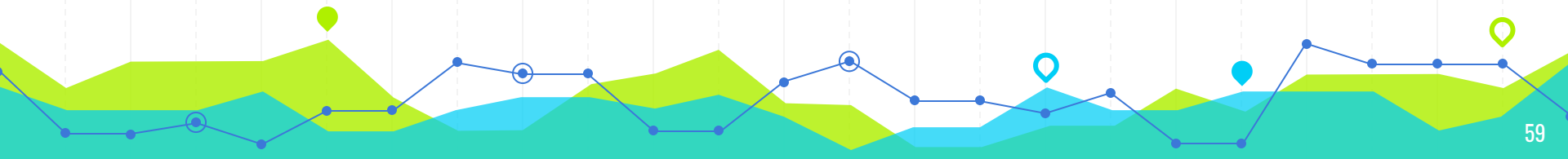
Ретроспективное когортное исследование



1. Некогда были **естественным образом** осуществлены некие воздействия (сформированы группы).
2. За группами проводились наблюдения, регистрировались показатели.
3. В момент анализа производится сравнение **исходов** между **группами воздействия**.

Ретроспективное когортное исследование (пример)

- К детской поликлинике приписаны дети, среди которых есть переболевшие в младенчестве рахитом.
- В ходе наблюдения за ними регистрируется ИМТ, набранный к моменту совершеннолетия.
- В процессе анализа между группами сравнивается вероятность приобрести дефицит массы тела.



Итого

Плюсы:

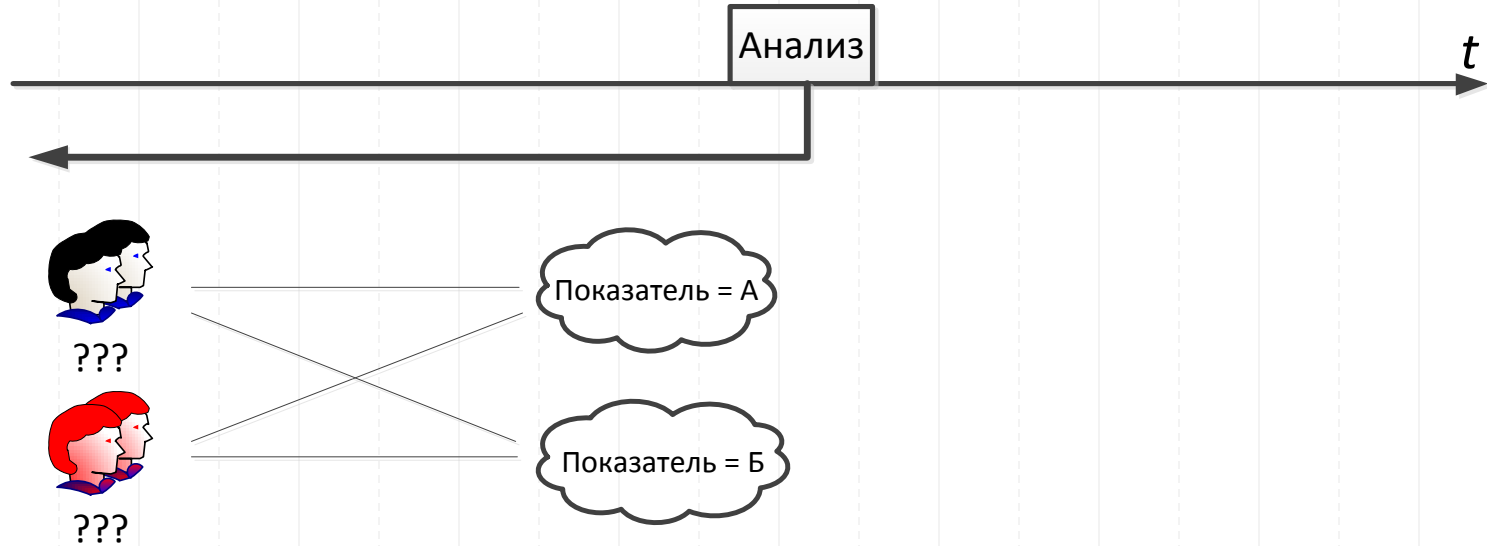
- Работа с уже собранными данными
- Можно формировать различные группы воздействия и сравнивать различные исходы
- Можно строить более обоснованные предположения о причинно-следственной связи, нежели в одномоментном КИ

Минусы:

- Группы могут оказаться сильно несбалансированными
- В процессе наблюдения на группы влияют различные факторы
- Если исход редкий, то он может не возникнуть ни в одной из групп



Ретроспективное исследование типа «случай-контроль»



1. Мы заинтересовались неким редким исходом (заболеванием)
2. Обращаясь во времени назад, описываются воздействия, при которых событие произошло и не произошло.
3. В момент анализа производится сравнение **воздействий** между **исходами**.

Ретроспективное исследование типа «случай-контроль» (пример) *

- Цель - поиска заболеваний-предшественников неходжкинской лимфомы
- Было проведено сравнение анамнезов группы больных лимфомой и группой здоровых
- Выявлено, что больные в прошлом чаще болели целиакией и ревматоидным артритом

* <https://pubmed.ncbi.nlm.nih.gov/16391371/>

Итого

Плюсы:

- Не надо дожидаться редких событий
- Не нужно длительно наблюдать за группами испытуемых

Минусы:

- Иногда трудно восстановить предшествующие события
- Не всегда возможно сформировать корректно группу сравнения



Главное различие между областями применения когортного метода и метода «случай-контроль»

Когортный – для анализа редких воздействий
Случай-контроль – для анализа редких исходов

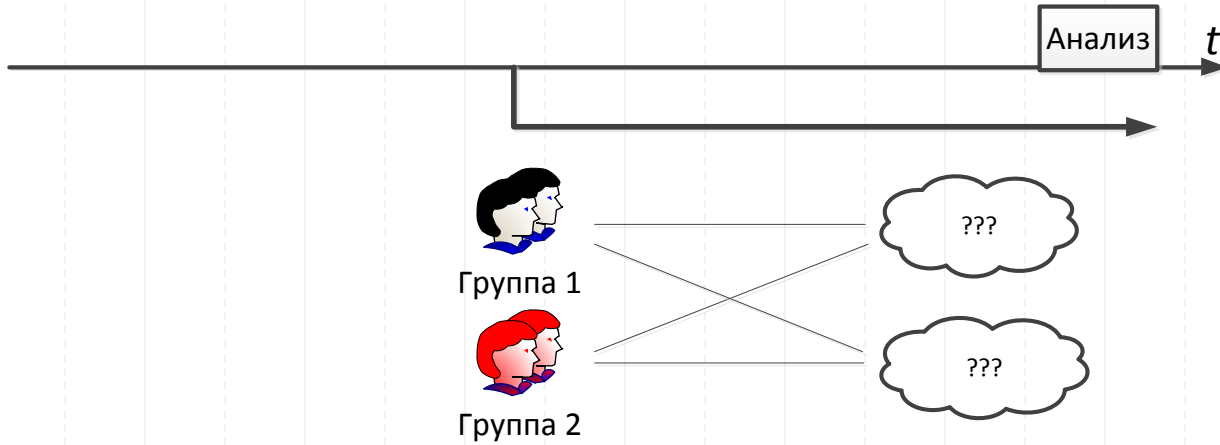


3. Есть много времени на проведение исследования

Проспективное исследование



Проспективное исследование



1. Мы заинтересовались сравнением исходов, возникающих в неких группах.
2. Набираются группы интереса.
3. **А**нализ проводится по прошествии времени наблюдения.

Проспективное исследование (пример)

- 3 фаза клинических испытаний
- Цель — испытание нового препарата
- Формируется две группы добровольцев с заданной нозологией: первой назначается препарат, второй — плацебо
- По итогам исследования принимается решение об эффективности препарата

Проспективное исследование (пример) *

- Цель — исследование влияния хромовой пыли на здоровье работников металлургического завода (МЗ)
- Были сформированы группы наблюдения среди работников МЗ и среди схожих по возрасту жителей Балтимора
- Группы наблюдались 20 лет
- Частота возникновения рака легких была в 2 раза выше у работников МЗ

* <https://pubmed.ncbi.nlm.nih.gov/541160/>

Такие длинные исследования
называются лонгитудальными

Итого

Плюсы:

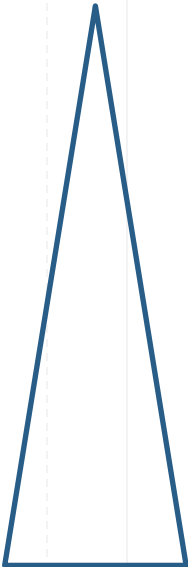
- Позволяют устанавливать причинно-следственные связи
- Можно формировать группы испытуемых с заданными показателями

Минусы:

- Время получения результатов может быть весьма продолжительным
- Обладают наибольшей дороговизной среди всех видов КИ



Иерархия степени вмешательства



Обсервационное (наблюдательное) *: просто описать текущую практику лечения выбранного заболевания

Неинтервенционное *: наблюдение за назначением и эффектом известного препарата в рутинной практике

Интервенционное: исследование нового препарата (или метода лечения) с недоказанной эффективностью путем его назначения группе добровольцев

* Часто отождествляются

Факторы влияющие на точность исследования

- Наличие правильно подобранной контрольной группы
- Вероятность вмешательства субъективных факторов
- Объем выборки участников исследования

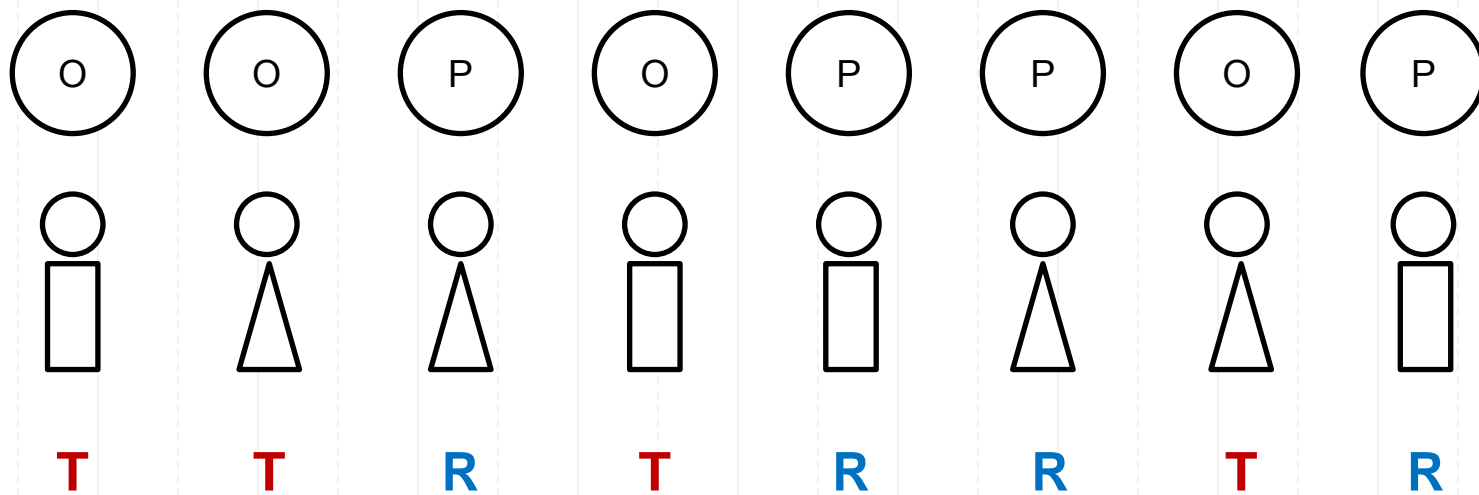


Контрольная группа

- Любой эффект познается только в сравнении
- Контрольная группа должна быть сбалансирована по основным показателям с основной группой
- Балансировка может достигаться за счет рандомизации



Рандомизация



Treatment – 2:2
Reference – 2:2

Всегда ли возможно наличие контрольной группы?

**Если не возможно, то это
неконтролируемое исследование**

Субъективные факторы

- Пациент знает, что принимает плацебо и волнуется

Слепое исследование

- Врач знает, что пациент принимает плацебо и волнуется за него

Двойное слепое исследование

- Биостатистик знает, что препарат не работает и может исказить результат

Тройное слепое исследование



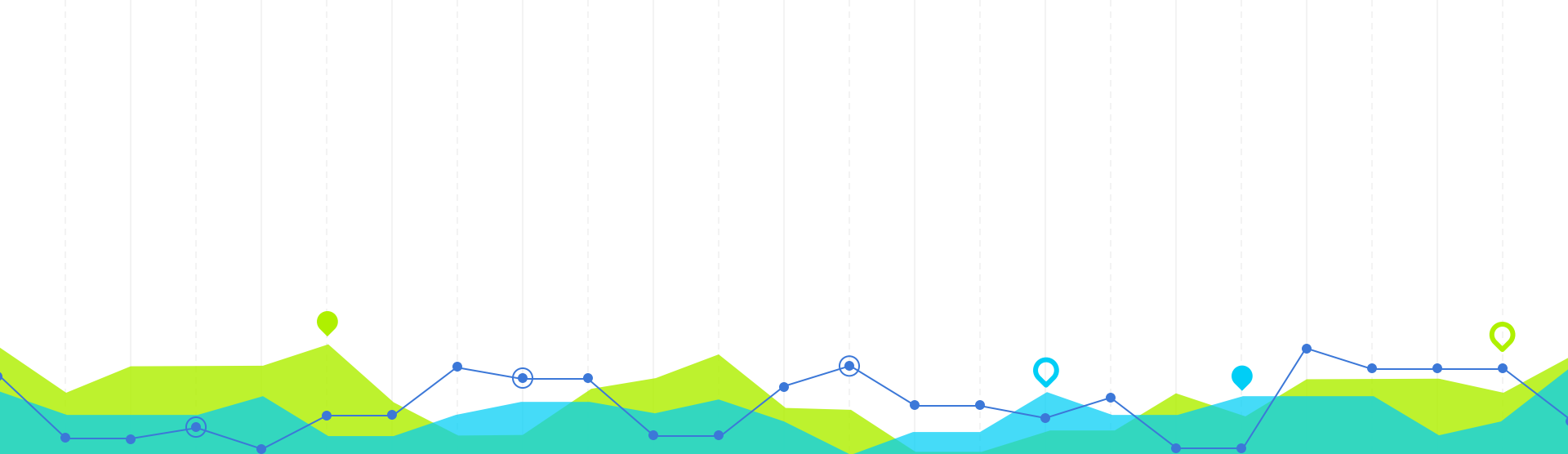
Всегда ли возможно заслепление?

**Если не возможно, то это
открытое исследование**

Объем выборки

- Одноцентровое
- Многоцентровое
- Мета-анализ





Типы признаков

5

Основные типы

качественные

(номинативные, категориальные, факторные)

- цвет
- пол
- название заболевания
- принимаемый препарат

Над такими данными нельзя проводить арифметические операции, отсутствует разумная сравнимость (больше/меньше, хуже/лучше, сильнее/слабее)

количественные

(числовые, численные)

- возраст
- вес
- уровень гемоглобина
- количество рецидивов

Арифметические операции допустимы и разумны (+, -, *, /, >, <)

Основные типы

качественные

(номинативные, категориальные, факторные)

количественные

(числовые, численные)

порядковые

(ординальные, ранговые)

- стадия заболевания (1, 2, 3, 4...)
- степень согласия с утверждением (шкала Лайкерта):
нет (-2), скорее нет (-1), не знаю (0), скорее да (+1), да (+2)
- оценка за прохождение теста: неуд. (2), удовл. (3), хор. (4), отл. (5)

Арифметические операции технически допустимы, но не всегда разумны. Имеют смысл только операции сравнения.

Схема оценки типа признака

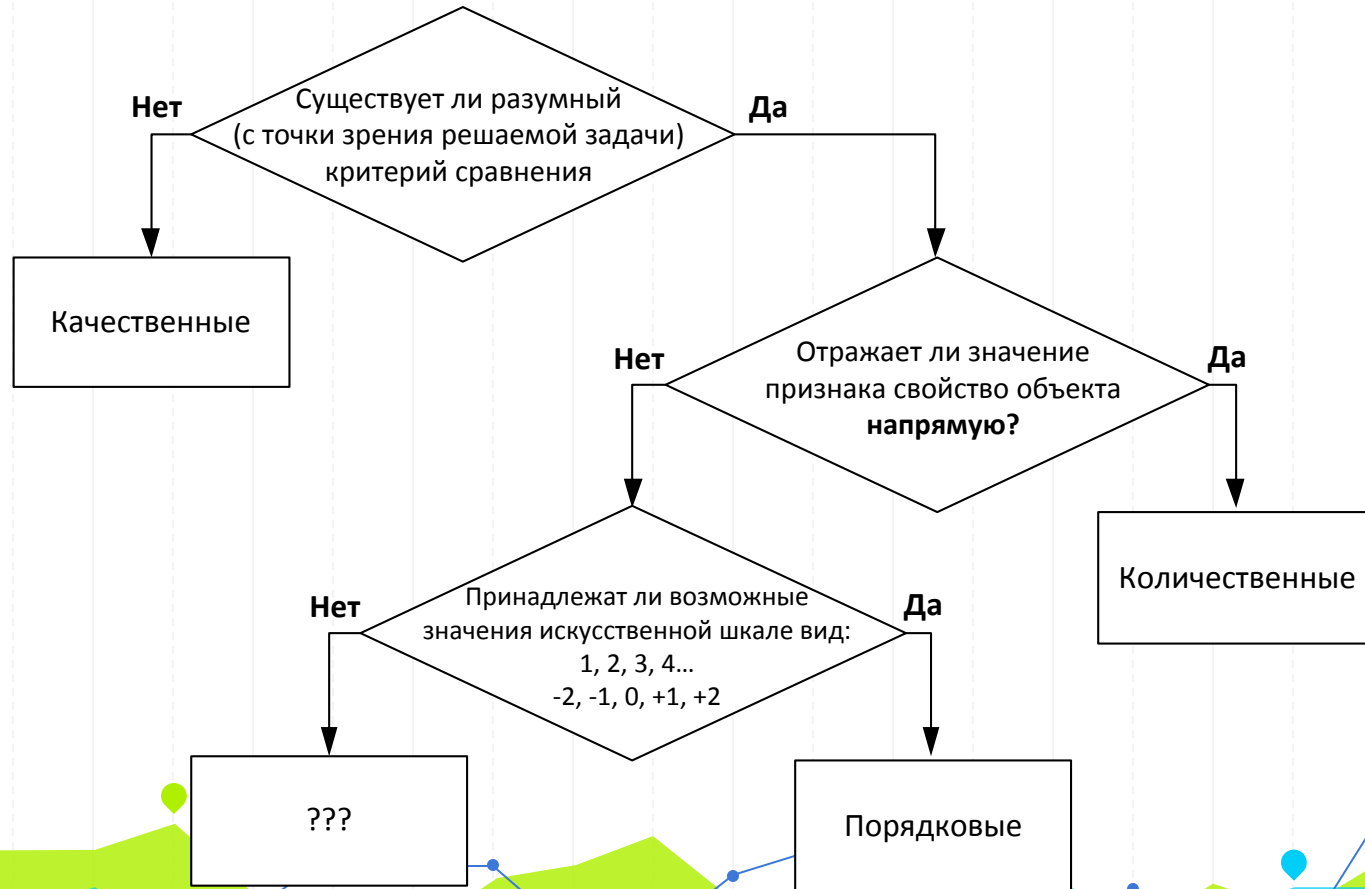
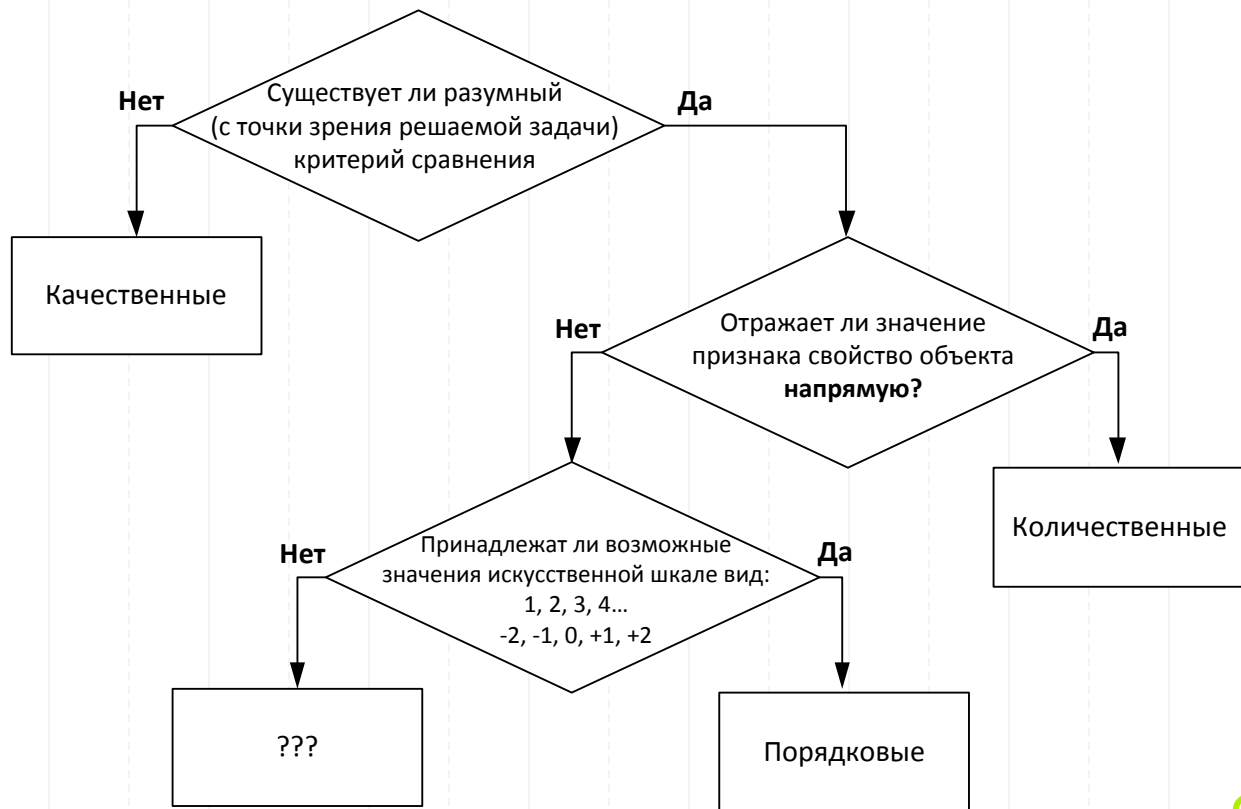


Схема оценки типа признака

- пол (м/ж)
- количество эпизодов диареи в сутки
- продолжительность жизни после операции
- степень инвалидности
- дозировка препарата

Не все так просто

- номер отделения, на котором лежит пациент
- год поступления пациента
- группа крови



Два ключевых понятия

Воздействие (англ. exposure) — признак, который описывает некий фактор, влияющий на пациента.

Исход (англ. outcome) — признак, который отражает результат воздействия.



Воздействие (exposure)

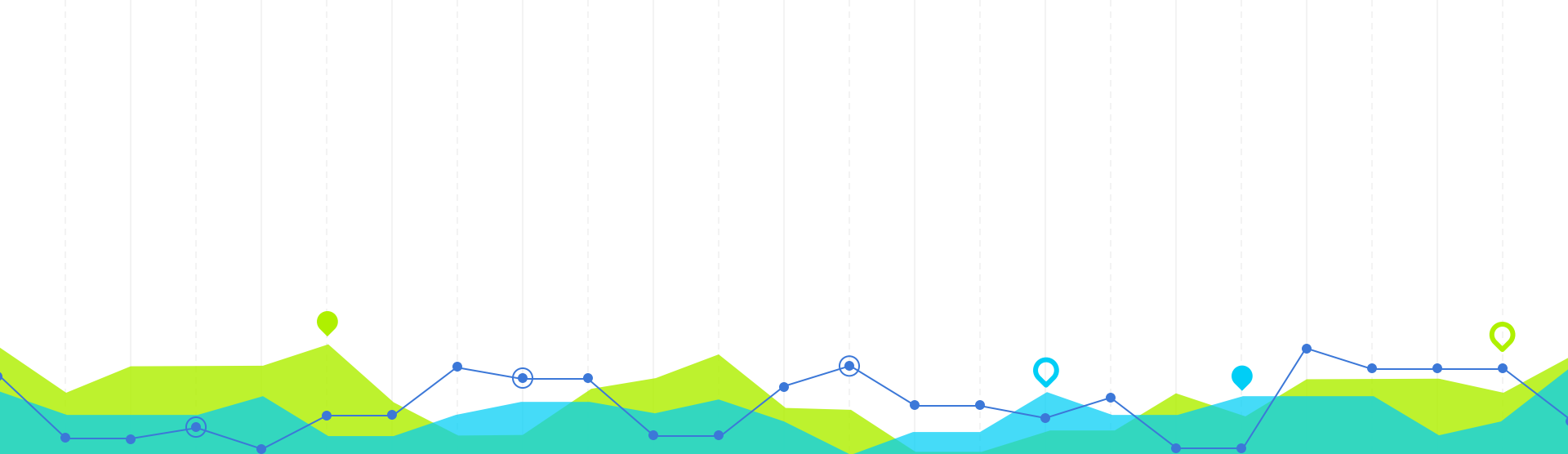
- Категориальное: препарат **A**/препарат Б, курит/не курит и пр.
- Количественный: доза лекарства, возраст, продолжительность работы в шахте и пр.



Исход (outcome)

- Категориальный: умер/выжил, госпитализация/амбулаторное лечение и пр.
- Количественный: продолжительность жизни, уровень гемоглобина, частота сердечных сокращений и пр.





Меры взаимосвязи

6

“

Мера взаимосвязи (ассоциации, association), число, показывающее силу зависимости между признаками.

Ассоциация \neq Причинно-следственная связь



Пример

В ходе кросс-секционного исследования выяснили, что у участников с повышенным индексом массы тела — более низкий доход (*отрицательная ассоциация*).

Интерпретация 1:

Повышенный ИМТ приводит к снижению трудоспособности и, как следствие, снижению дохода.

Интерпретация 2:

Низкий уровень доходов приводит к нездоровому питанию, и, как следствие, повышению ИМТ.

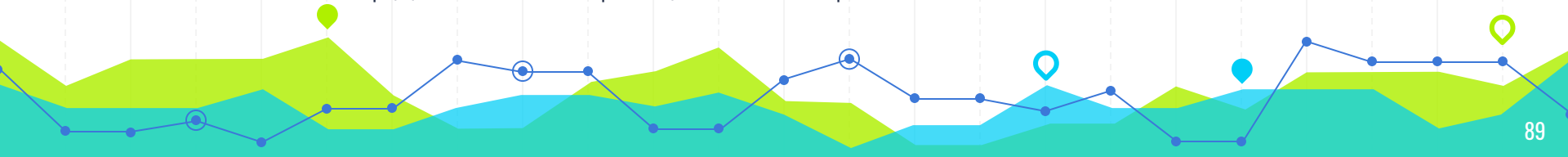


Воздействие (exposure)

- Категориальное: препарат А/препарат Б, курит/не курит и пр.
- Количественный: доза лекарства, возраст, продолжительность работы в шахте и пр.

Исход (outcome)

- Категориальный: умер/выжил, госпитализация/амбулаторное лечение и пр.
- Количественный: продолжительность жизни, уровень гемоглобина, частота сердечных сокращений и пр.

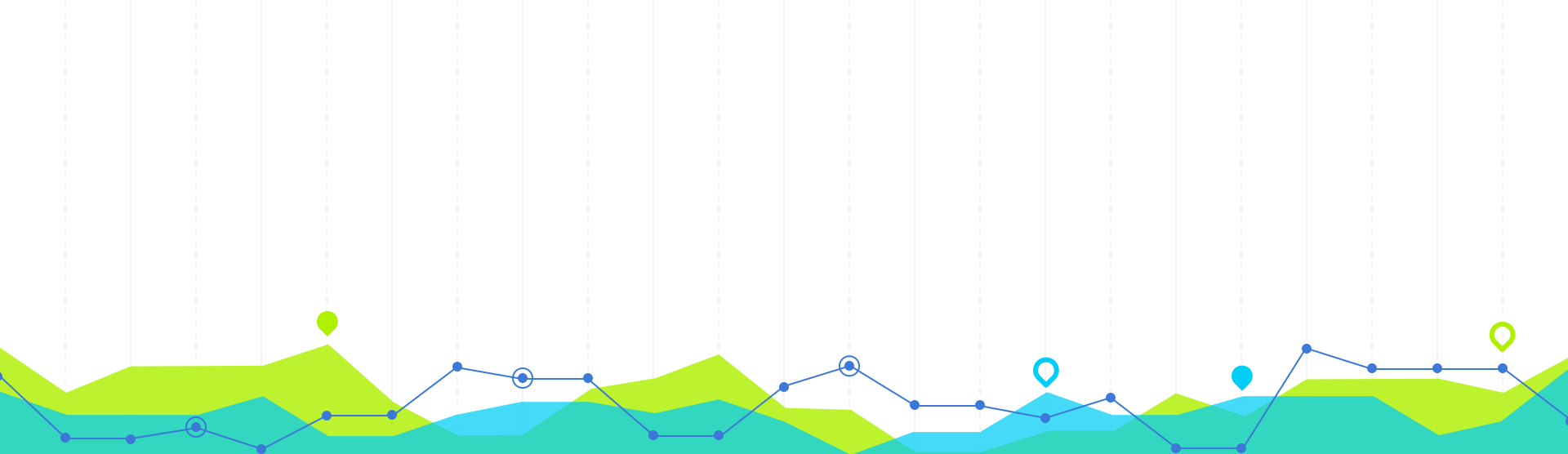


Возможные комбинации

- Воздействие — категориальный, исход — категориальный:
 - Препарат **A**/Препарат **B** → **З**доров/**Б**олен
- Воздействие — категориальный, исход — количественный:
 - Препарат **A**/Препарат **B** → **У**ровень гемоглобина
- Воздействие — количественный, исход — категориальный:
 - **Д**озировка препарата → **З**доров/**Б**олен
- Воздействие — количественный, исход — количественный:
 - **Д**озировка препарата → **У**ровень гемоглобина

Для каждой комбинации необходима своя мера взаимосвязи!





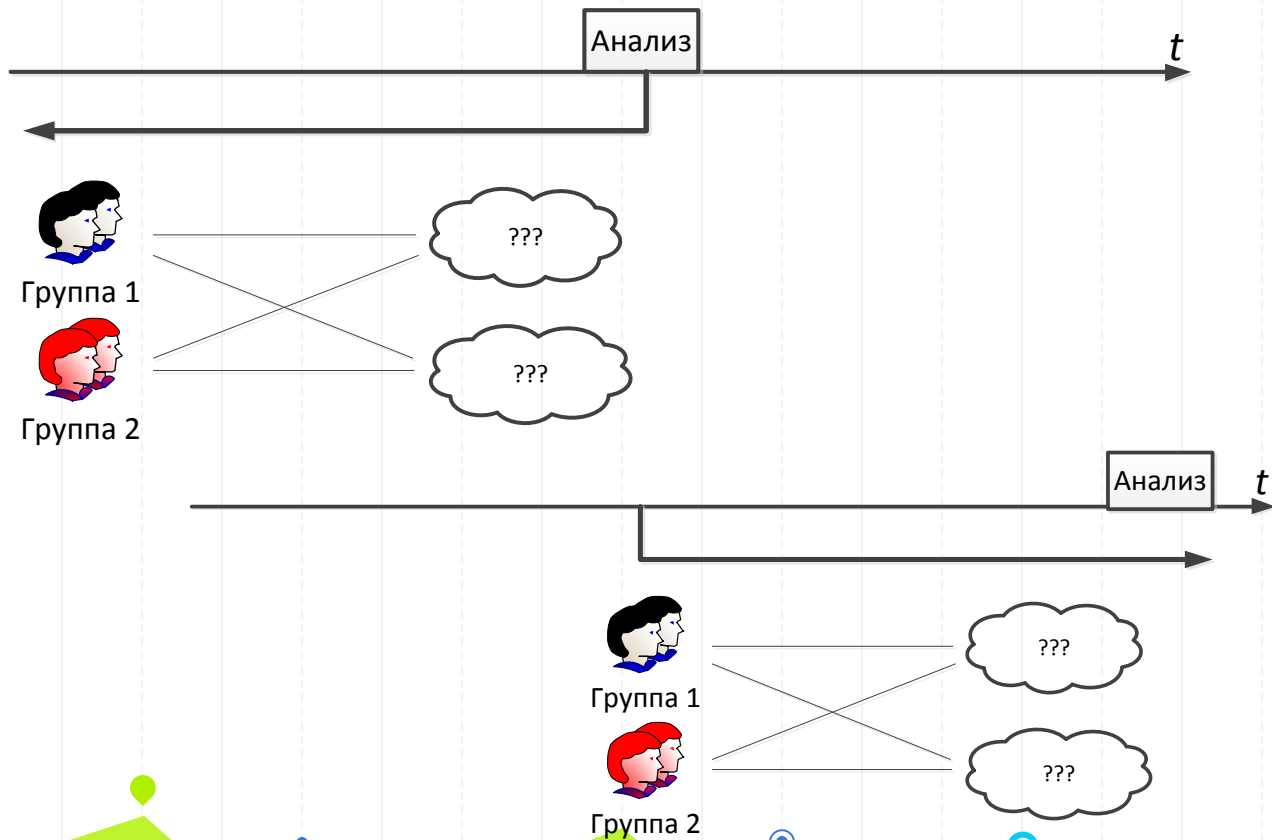
Взаимосвязь категориальных признаков 6.1

Бинарные категориальные признаки

Признак принимает только **два** возможных значения:

- Пол: м/ж
- Трудовой статус: работает/не работает
- Лечение: препарат **A**/препарат Б
- и т. д.

Когортные исследования



Двумерные таблицы для **пар** признаков в когортном исследовании

Лечение\Состояние через неделю	Болен	Здоров
Терапия - группа А (n = 300)	100	200
Плацебо - группа Б (n = 180)	80	100

Риск наличия болезни в группе А: $100/300 = 0.33$

Риск наличия болезни в группе Б: $80/180 = 0.44$

Отношение рисков (Risk ratio, RR): $0.33/0.44 = 0.75$

Двумерные таблицы для **пар** признаков в когортном исследовании

Лечение\Состояние через неделю	Болен	Здоров
Терапия - группа А	a	b
Плацебо - группа Б	c	d

$$RR = \frac{a(c + d)}{c(a + b)}$$

Двумерные таблицы для **пар** признаков в когортном исследовании

Лечение\Состояние через неделю	Болен	Здоров
Терапия - группа А (n = 100000)	30	99970
Плацебо - группа Б (n = 100000)	40	99960

Риск наличия болезни в группе А: $30/100000 = 0.0003$

Риск наличия болезни в группе Б: $40/100000 = 0.0004$

Отношение рисков (Risk ratio, RR): $0.0003 / 0.0004 = 0.75$

Двумерные таблицы для **пар** признаков в когортном исследовании

Лечение\Состояние через неделю	Болен	Здоров
Терапия - группа А (n = 100000)	30	99970
Плацебо - группа Б (n = 100000)	40	99960

Риск наличия болезни в группе А: $30/100000 = 0.0003$

Риск наличия болезни в группе Б: $40/100000 = 0.0004$

Разница рисков (Risk difference, RD): $0.0004 - 0.0003 = 0.0001$

Двумерные таблицы для **пар** признаков в когортном исследовании

Лечение\Состояние через неделю	Болен	Здоров
Терапия - группа А	a	b
Плацебо - группа Б	c	d

$$R D = \frac{a}{a + b} - \frac{c}{c + d}$$

Двумерные таблицы для пар признаков в когортном исследовании

Лечение\Состояние через неделю	Болен	Здоров
Терапия - группа А (n = 300)	100	200
Плацебо - группа Б (n = 180)	80	100

Отношение рисков (Risk ratio, RR): $0.33 / 0.44 = 0.75$

Разница рисков (Risk difference, RD): $0.44 - 0.33 = 0.11$

Лечение\Состояние через неделю	Болен	Здоров
Терапия - группа А (n = 100000)	30	99970
Плацебо - группа Б (n = 100000)	40	99960

Отношение рисков (Risk ratio, RR): $0.0003 / 0.0004 = 0.75$

Разница рисков (Risk difference, RD): $0.0004 - 0.0003 = 0.0001$

Промежуточный вывод

В результатах работы нужно фиксировать не только отношение рисков (RR), но и разницу рисков (RD)



Двумерные таблицы для **пар** признаков в когортном исследовании

Лечение\Состояние через неделю	Болен	Здоров
Терапия - группа А (n = 300)	100	200
Плацебо - группа Б (n = 180)	80	100

Риск наличия болезни в группе А: $100/300 = 0.33$

Риск наличия болезни в группе Б: $80/180 = 0.44$

Отношение рисков (Risk ratio, RR): $0.33/0.44 = 0.75$

Двумерные таблицы для **пар** признаков в когортном исследовании (пусть в группе Б было в 10 раз больше участников)

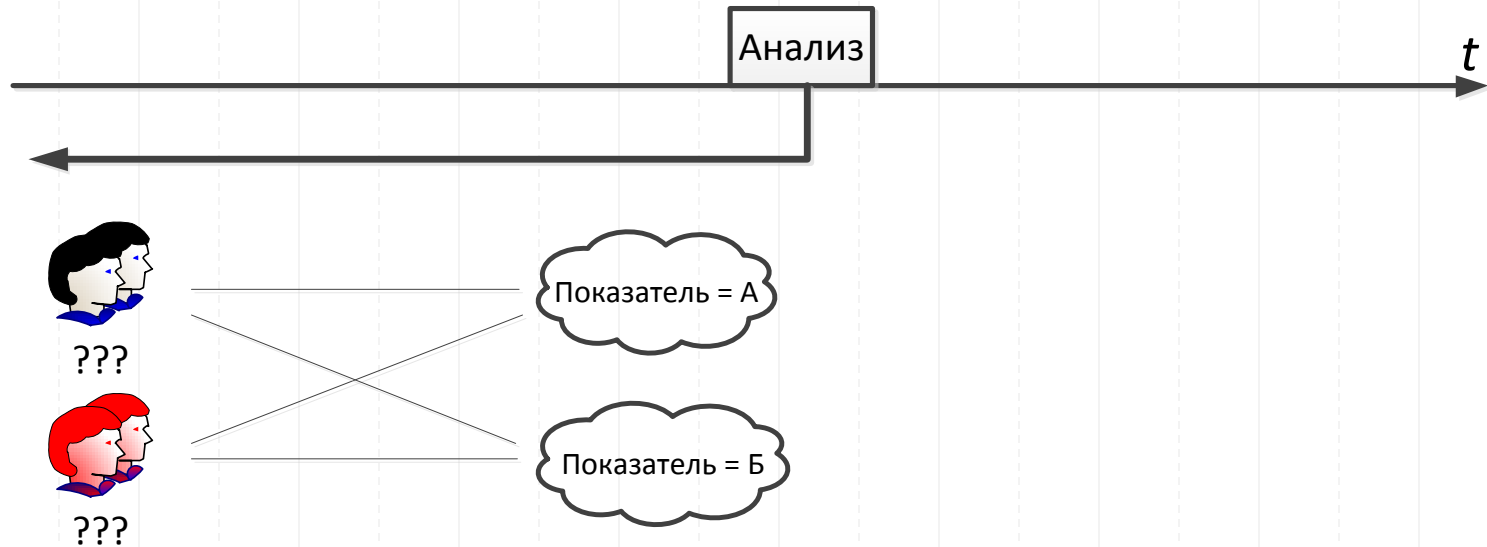
Лечение\Состояние через неделю	Болен	Здоров
Терапия - группа А (n = 300)	100	200
Плацебо - группа Б (n = 1800)	800	1000

Риск наличия болезни в группе А: $100/300 = 0.33$

Риск наличия болезни в группе Б: $800/1800 = 0.44$

Отношение рисков (Risk ratio, RR): $0.33/0.44 = 0.75$

Ретроспективное исследование типа «случай-контроль»



1. Мы заинтересовались неким исходом (заболеванием)
2. Обращаясь во времени назад, описываются воздействия которые были оказаны на пациентов с тем или иным исходом.

Двумерные таблицы для **пар** признаков в исследовании типа «случай-контроль»

Лечение\Состояние через неделю	Болен (n = 180)	Здоров (n = 300)
Терапия - группа А (n = 300)	100	200
Плацебо - группа Б (n = 180)	80	100

Риск наличия болезни в группе А: $100/300 = 0.33$

Риск наличия болезни в группе Б: $80/180 = 0.44$

Отношение рисков (Risk ratio, RR): $0.33/0.44 = 0.75$

Двумерные таблицы для **пар** признаков в исследовании типа «случай-контроль» (пусть набрали в 10 раз больше здоровых участников)

Лечение\Состояние через неделю	Болен (n = 180)	Здоров (n = 3000)
Терапия - группа А (n = 300)	100	2000
Плацебо - группа Б (n = 180)	80	1000

Риск наличия болезни в группе А: $100/2100 = 0.0476$

Риск наличия болезни в группе Б: $80/1080 = 0.074$

Отношение рисков (Risk ratio, RR): $0.0476 / 0.074 = 0.675$

Все ещё хуже с разницей рисков

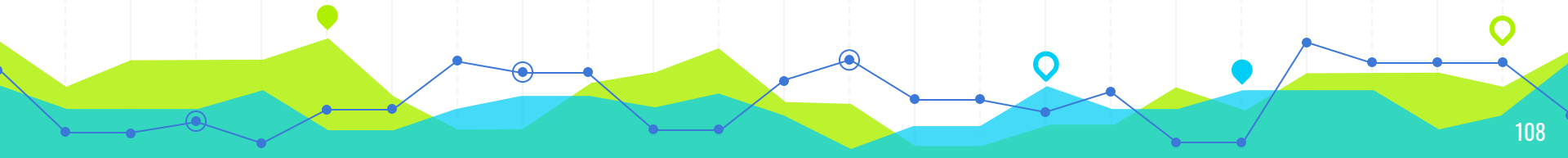
1. $RD = 0.44 - 0.33 = 0.11$

2. $RD = 0.074 - 0.05 = 0.024$



Промежуточный вывод

Стандартный способ расчет отношения рисков в исследованиях типа «случай-контроль» дает искаженный результат.



Двумерные таблицы для **пар** признаков в когортном исследовании

Лечение\Состояние через неделю	Болен	Здоров
Терапия - группа А (n = 300)	100	200
Плацебо - группа Б (n = 180)	80	100

Шанс наличия болезни в группе А: $100/200 = 0.5$

Шанс наличия болезни в группе Б: $80/100 = 0.8$

Отношение шансов (Odds ratio, OR): $0.5/0.8 = 0.625$

Двумерные таблицы для **пар** признаков в когортном исследовании

Лечение\Состояние через неделю	Болен	Здоров
Терапия - группа А	a	b
Плацебо - группа Б	c	d

$$OR = \frac{ad}{bc}$$

Двумерные таблицы для **пар** признаков в исследовании типа «случай-контроль»

Лечение\Состояние через неделю	Болен (n = 180)	Здоров (n = 300)
Терапия - группа А	100	200
Плацебо - группа Б	80	100

$$O R = \frac{a d}{b c} = \frac{100 \cdot 100}{80 \cdot 200} = 0.625$$

Двумерные таблицы для **пар** признаков в исследовании типа «случай-контроль» (пусть набрали в 10 раз больше здоровых участников)

Лечение\Состояние через неделю	Болен (n = 180)	Здоров (n = 3000)
Терапия - группа А	100	2000
Плацебо - группа Б	80	1000

$$OR = \frac{ad}{bc} = \frac{100 \cdot 1000}{80 \cdot 2000} = 0.625$$

Выводы

- Риски и все с ними связанное (отношение, разница) можно использовать только в когортных исследованиях.
- **Отношение шансов** можно использовать и в когортных и в случай-контроль.

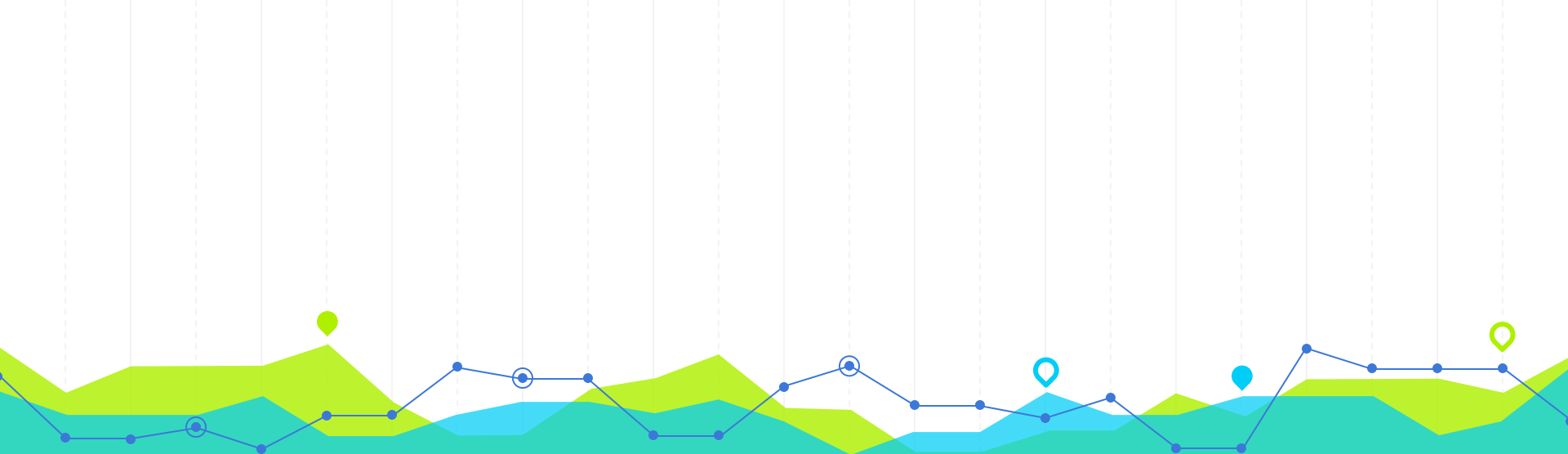


И **OR**, и **RR** отражают соотношение частот возникновения событий (болезни, смерти, попадания в реанимацию и пр.) между двумя группами пациентов **но немного по-разному**.

OR = 1 или RR = 1 или RD = 0: группы эквивалентны

OR > 1 или RR > 1 или RD > 0: в одной группе событие возникает чаще, чем в контрольной





Взаимосвязь количественного и категориального признаков

6.2

Оба случая рассматриваются совместно

- Воздействие — категориальный, исход — количественный:
 - Препарат А/Препарат В → Уровень гемоглобина
- Воздействие — количественный, исход — категориальный:
 - Дозировка препарата → Здоров/Болен

Мера ассоциации

- Воздействие — категориальный, исход — количественный:
 - Препарат **A**/Препарат **B** → Уровень гемоглобина

Рассчитывается средний уровень гемоглобина в группе, принимающей препарат **A**: обозначается через $\hat{\mu}_A$

Рассчитывается средний уровень гемоглобина в группе, принимающей препарат **B**: обозначается через $\hat{\mu}_B$

Разница в средних (means difference, MD): $\hat{\mu}_A - \hat{\mu}_B$



Мера ассоциации

- Воздействие — количественный, исход — категориальный:
 - Дозировка препарата → Здоров/Болен

Рассчитывается средняя дозировка препарат в группе выздоровевших: обозначается через $\hat{\mu}_З$

Рассчитывается средняя дозировка препарат в группе не выздоровевших: обозначается через $\hat{\mu}_Б$

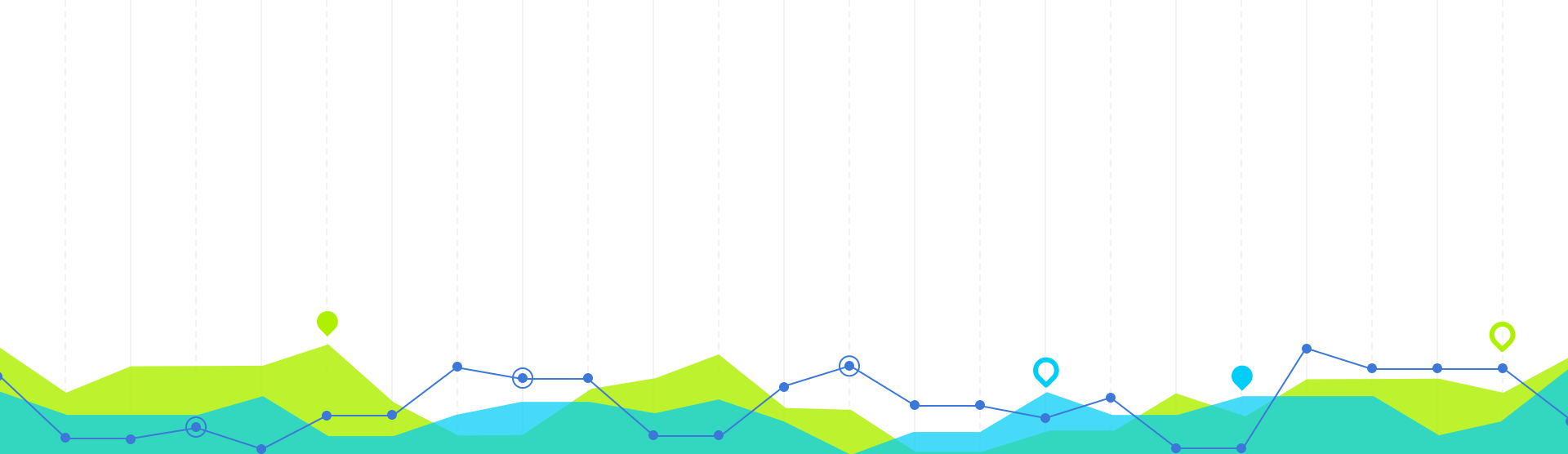
Разница в средних (means difference, MD): $\hat{\mu}_З - \hat{\mu}_Б$

Мера ассоциации, лучше подходящая для ординальных переменных

Целесообразно ли сравнивать средние группы инвалидности?

$$\Pr\{X_A > X_B\} = ?$$

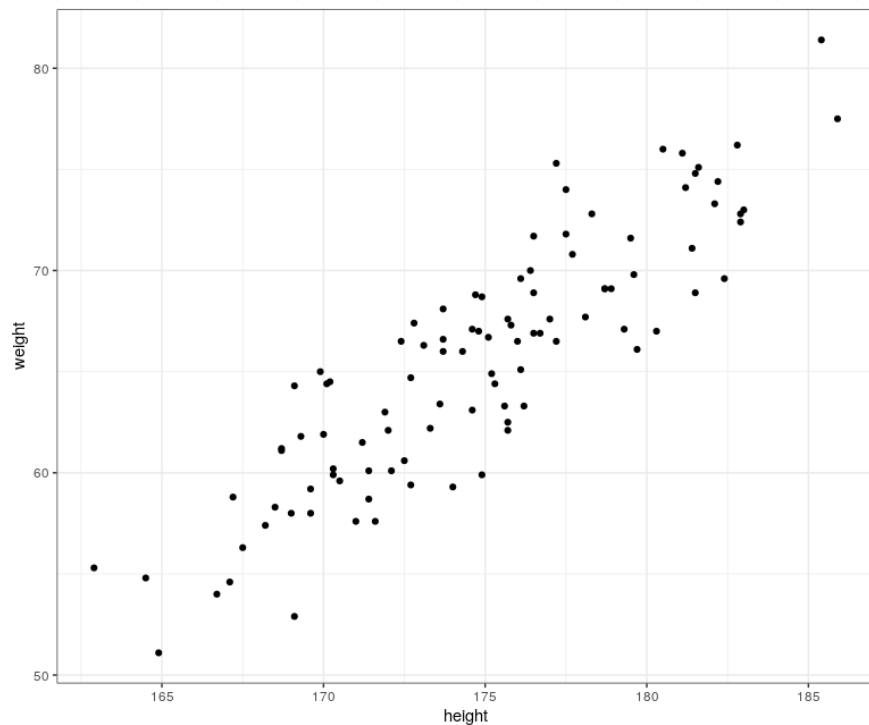




Взаимосвязь количественных признаков

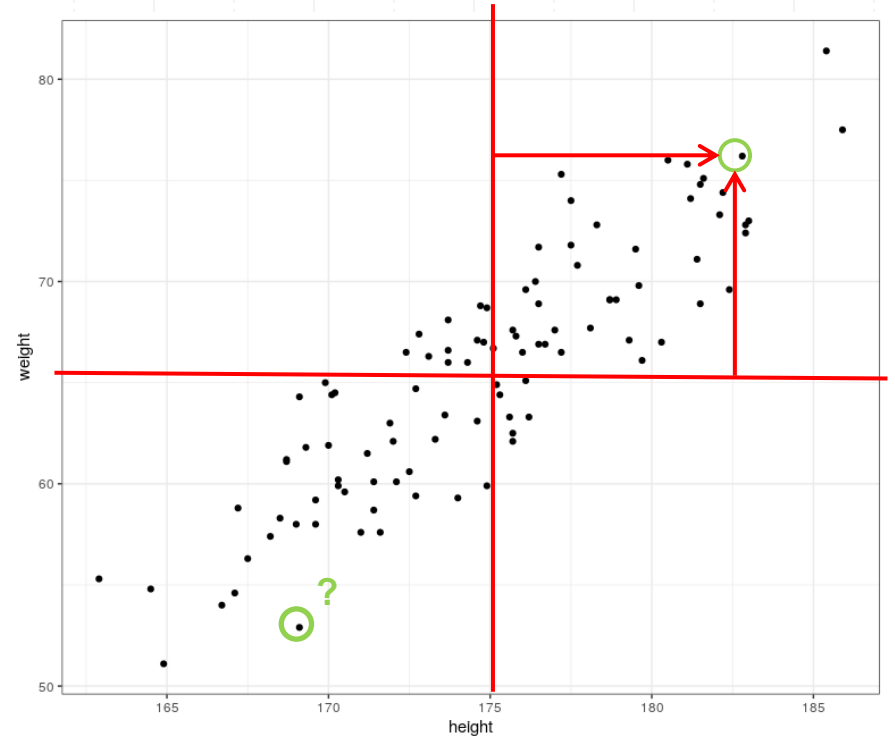
6.3

Визуализация пары количественных признаков (диаграмма рассеяния)



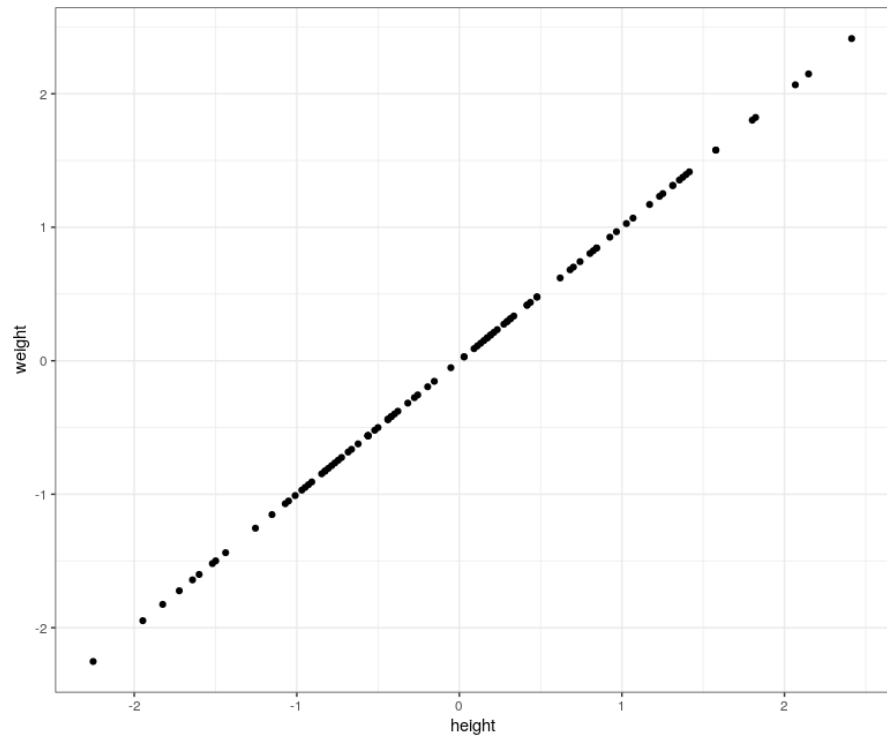
Корреляция (по Пирсону)

Меряет, на сколько отклонение одной величины относительно среднего сопоставимо с отклонением второй величины. Сдвиг вверх одной, как правило сопровождается сдвигом вверх другой и наоборот.

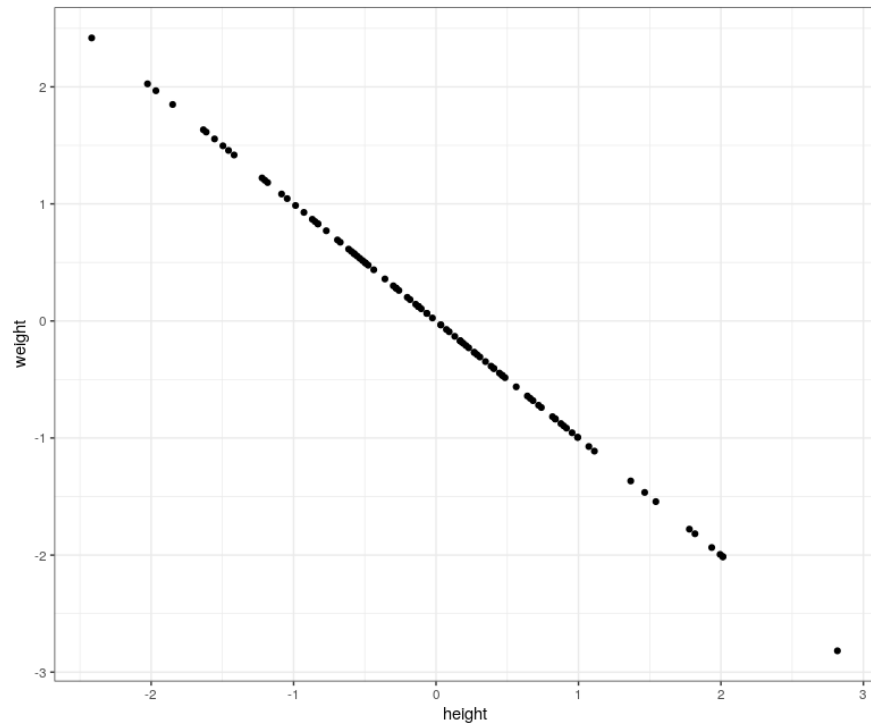


Как это выглядит?

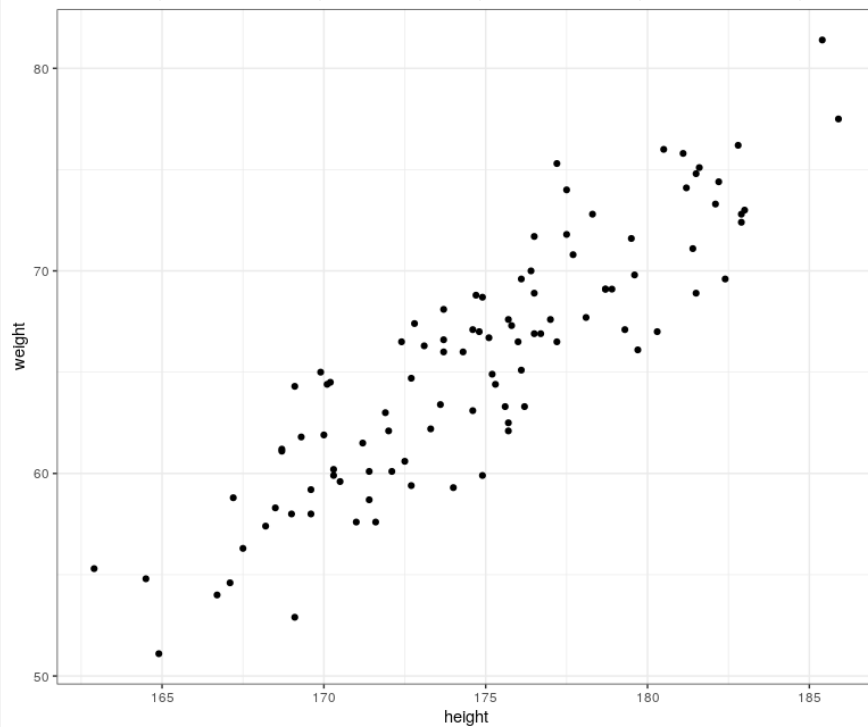
+1



-1

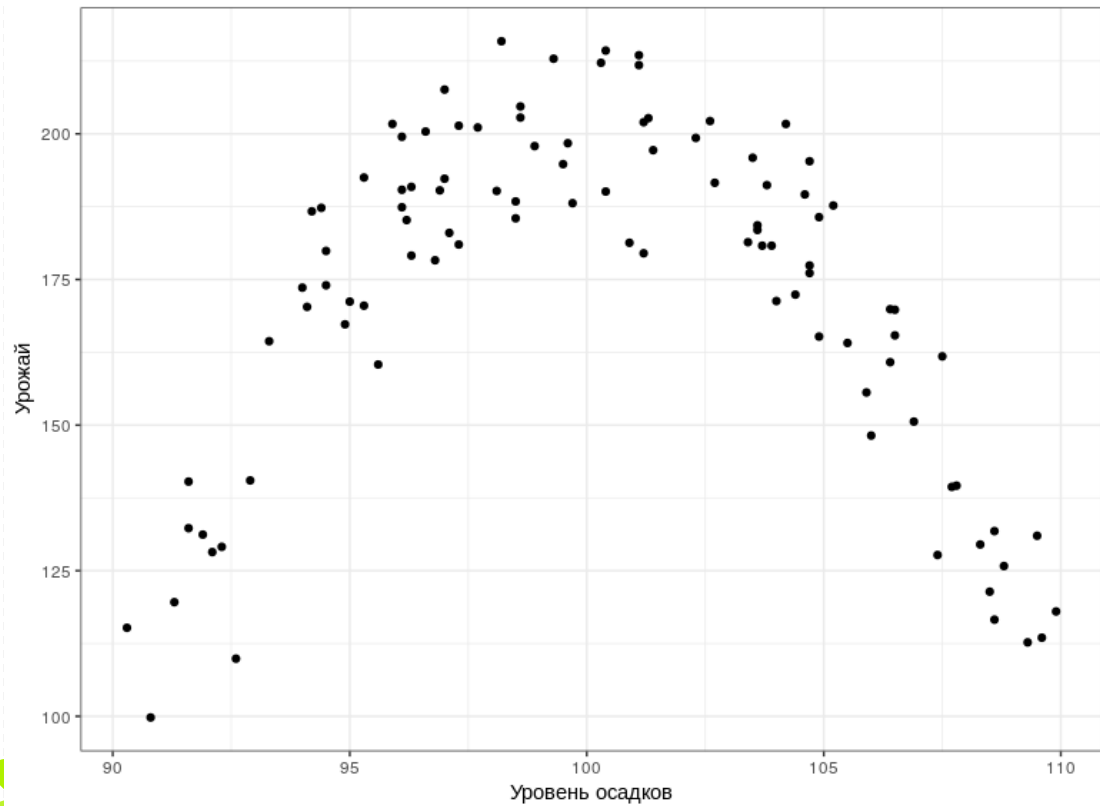


Промежуточный уровень корреляции



$$r_{XY} = 0,85$$

Сложный вариант взаимосвязи



!!! ВАЖНО !!!

О чем говорит наличие корреляции?

Ⓐ о наличии взаимосвязи (ассоциации) между признаками

О чем **не** говорит наличие корреляции?

Ⓐ о наличии причинно-следственной связи

О чем говорит отсутствие корреляции?

Ⓐ об отсутствии **линейной** зависимости между признаками

О чем **не** говорит отсутствие корреляции?

Ⓐ об отсутствии зависимости между признаками



Спасибо за внимание!

