

ВСЕРОССИЙСКИЙ
ЛЕКТОРИЙ



РНФ

Научные исследования в биологии и медицине: статистическая обработка данных на примере хронофизиологии

1

ЕЛЕНА БУДКЕВИЧ

канд. мед. наук, ведущий научный сотрудник

департамента науки

доцент кафедры физико-химической биологии и

иммунологии

$2+2=4?$

Влияние хронопсихологических особенностей девушек на эмоциональное питание и вкусовую чувствительность

Грант РНФ № 24-28-00131

Актуальность: рост нарушений, связанных с изменением пищевого поведения, сопровождаемых ростом массы тела.

Проблема: предпочтение активности в поздние вечерние часы (вечерний хронотип) связано с повышенным риском ожирения, а эмоциональное питание является фактором риска ожирения, но исследования проведены в большинстве случаев на популяции мужчин.

Решение: разные этапы циклических изменений в жизни женщины характеризуются различным уровнем гормонов и качеством сна, а следовательно, и выбором времени сна. Организация сна у женщин может быть ключевым фактором, способным повлиять на пищевое поведение.



Знай когда твои данные достоверны!

Experimental biologists, their reviewers and their publishers must grasp basic statistics, urges David L. Vaux, or sloppy science will continue to grow.

Генеральная совокупность = популяция – совокупность всех интересующих нас объектов



наблюдение

Статистика это инструмент для количественного анализа и интерпретации данных

Описательная статистика
(descriptive) – описываем выборку

Индуктивная статистика (inferential)
– на основе знаний о выборке
делаем заключение о свойствах
генеральной совокупности

Переменные

Качественные

nominal

(не выстраиваются в последовательность)



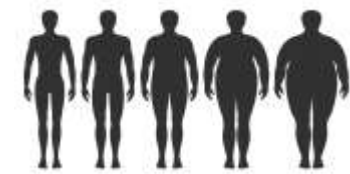
Ранговые

ordinal

(качественные, могут быть упорядочены; размер интервалов на шкале неодинаковый)



Количественные



шкала

отношений

ratio scale

интервальная

шкала

interval scale



Дискретные

discrete

Непрерывные

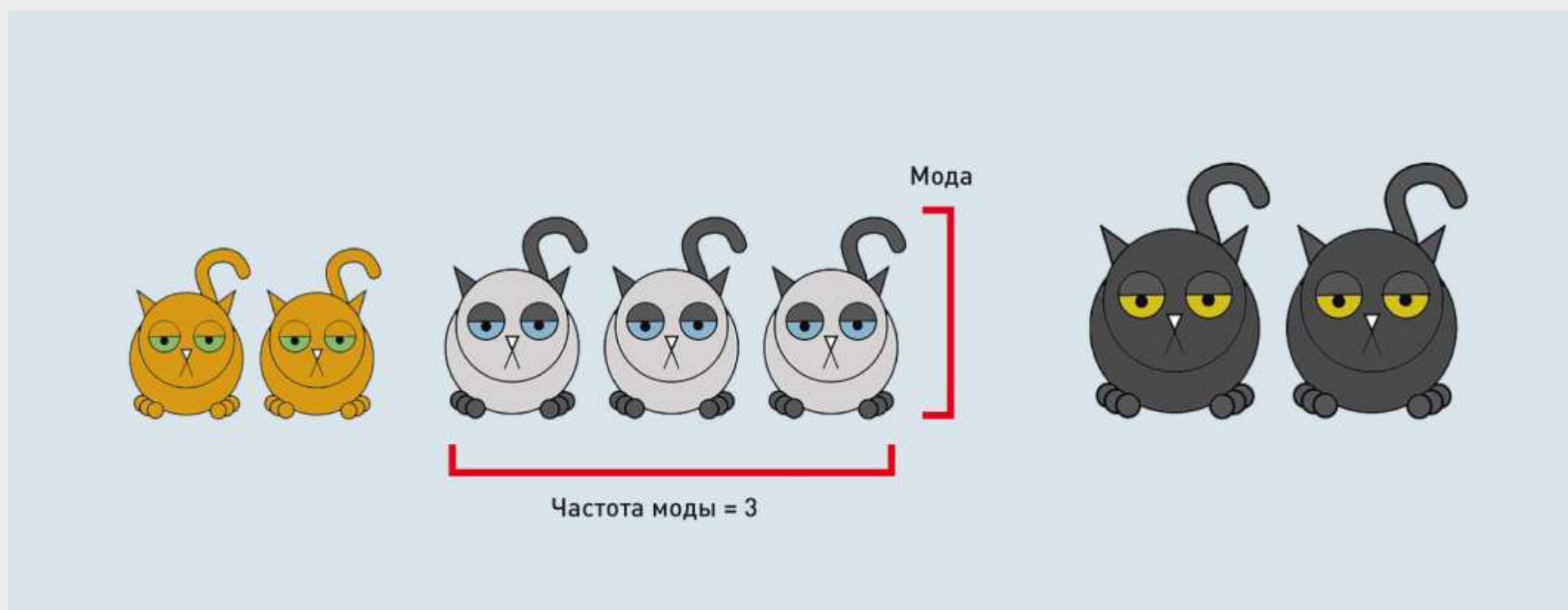
continuous



Потеря информации и точности

Оценка параметров популяции

«Середина» распределения
среднее, медиана, мода



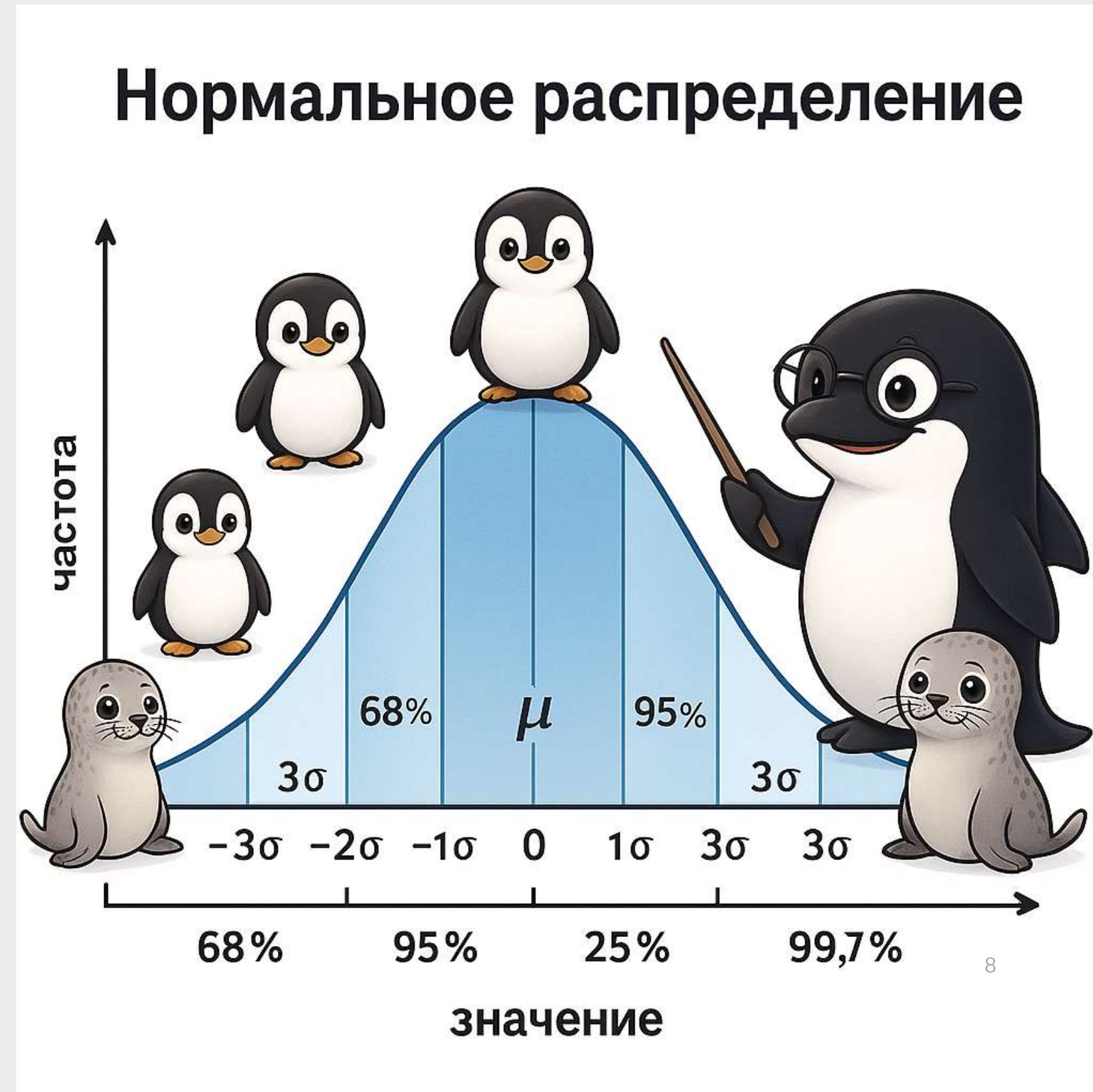
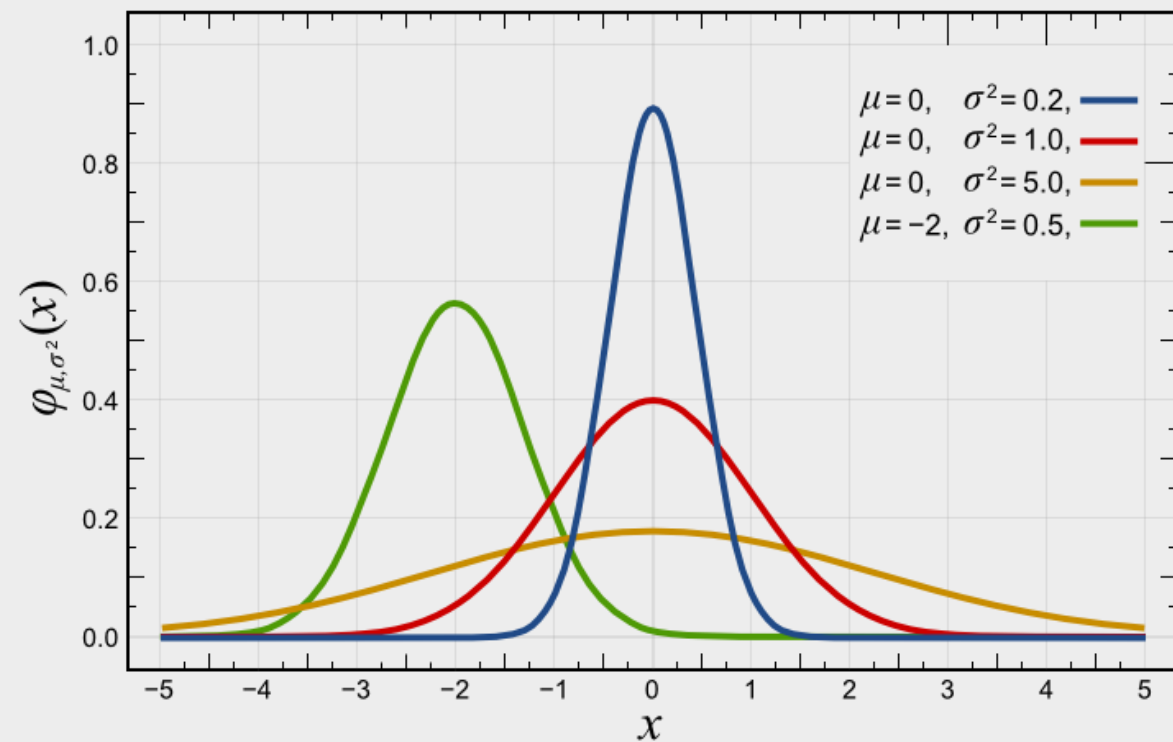
Оценка параметров популяции

«Ширина» распределения = разброс
размах, стандартное отклонение
дисперсия

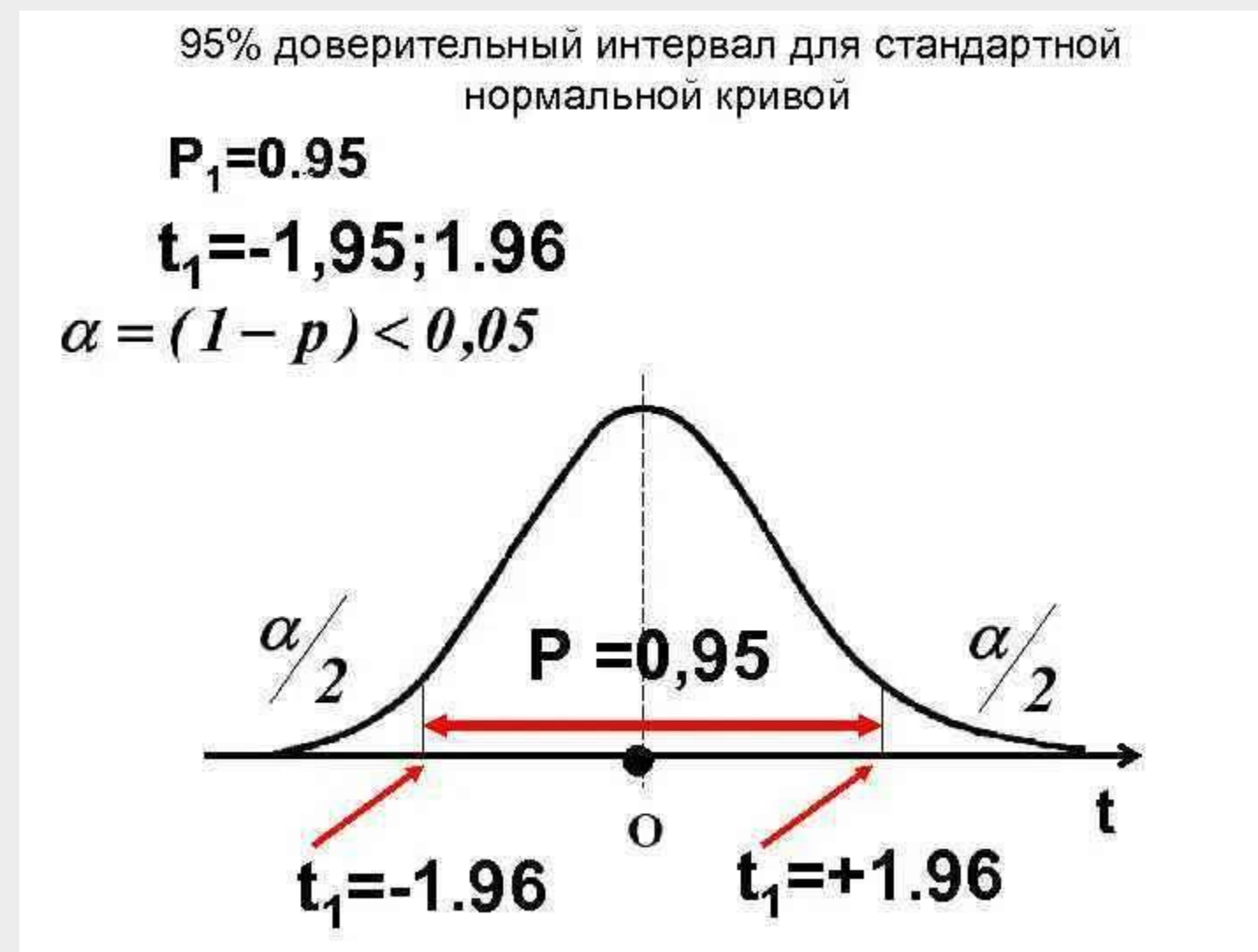


Оценка параметров популяции

«Форма» распределения
симметричное / скошенное
непрерывное / дискретное
асимптотическое / не асимптотическое



в публикациях часто используют доверительный интервал (95% CI), ведь он показывает местонахождение параметра в популяции

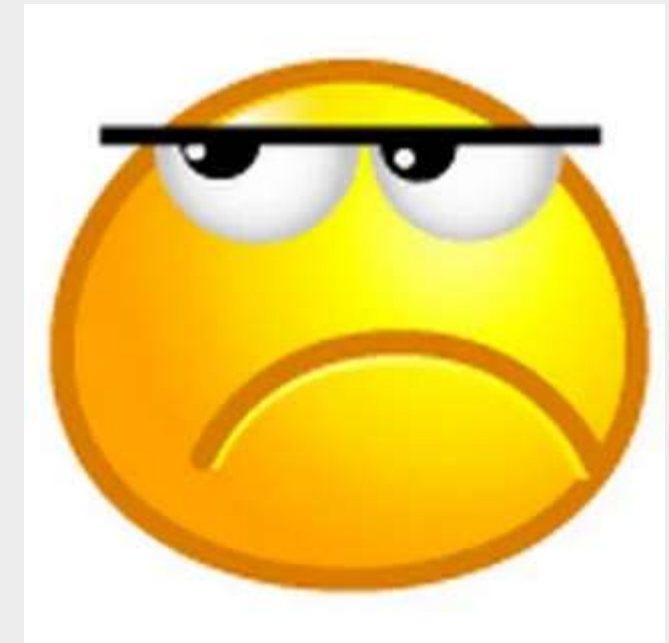


«...mean risk was 1.13, 95% CI = 0.99 to 1.29»

Как проверить предположения о взаимосвязях между переменными?

Бытовых впечатлений явно недостаточно.

То, что тетя Катя утверждает, что ей от боли в горле помог настой картофельной ботвы на молоке со скипидаром, не доказывает эффективность этого чудодейственного средства.



Тестирование статистических гипотез

Формулируем ДВЕ взаимоисключающие гипотезы:

H_0 (нулевая гипотеза, null hypothesis) – ее мы собираемся **опровергать**; обычно говорит, что нет различий, нет эффекта, нет изменений...

H_1 (альтернативная гипотеза, alternative hypothesis) – ее мы **примем**, если удастся отвергнуть H_0 .

Начинаем эксперимент?

НЕТ

- Формулируем причинную биологическую гипотезу.
- Операционализируем в измеримые исходы и параметры.
- Определяем направление и клинически значимую величину эффекта.
- Фиксируем критерии успеха до начала исследования.

G*Power

Statistical Power Analyses for Mac and Windows

G*Power — это инструмент для статистического анализа мощности различных t-критериев, F-критериев, χ^2 -критериев, z-критериев и некоторых точных критериев.

G*Power также можно использовать для вычисления величины эффекта и графического анализа за мощности.



ите, чтобы увеличить)



График мощности



График мощности (таблица)



Объем выборки

2) Супер-простой калькулятор (две группы, 80% мощность, $\alpha=0.05$, двусторонний)

Формула «на салфетке» для n на группу:

$$n \approx \frac{16}{d^2}, \quad d = \frac{\Delta}{\sigma}$$

где

- Δ — минимальный биологически важный эффект, который хотите обнаружить,
- σ — SD (стандартное отклонение) показателя,
- d — размер эффекта по Коэну.

То есть: чем меньше эффект Δ (или чем больше разброс σ), тем больше нужно животных.

0,2 — малый эффект, который трудно увидеть невооружённым глазом. Например, разница в росте между 15-летними и 16-летними девочками в США.

0,5 — средний эффект, который, вероятно, можно различить невооружённым глазом.

0,8 и более — большой эффект, который можно увидеть невооружённым глазом. Например, разница в росте между 13-летними и 18-летними девочками.

Объем выборки

3) Как взять σ и Δ без пилота (очень просто)

- Если знаете **CV%** (коэфф. вариации = $SD/\text{среднее} \cdot 100\%$):

$$d \approx \frac{\Delta\%}{CV\%}$$

Пример: $CV \approx 20\%$, хотим -20% эффекта $\Rightarrow d = 20/20 = 1 \Rightarrow n \approx 16/\text{группу}$.

- Если есть публикация/прошлый опыт с похожим протоколом — возьмите их SD как стартовую оценку.

Практические «минимумы» (ориентиры для старта)

Это не правила, а рабочие ориентиры, если нет пилотного исследования;

точное n всё равно считают по формуле выше!

Биохимия/простые физиологические метрики (низкая вариабельность): 8–12 животных/группу при ожидаемом крупном эффекте.

•Морфология/гистология (средняя вариабельность): 10–15/группу.

•Поведение (высокая вариабельность): 12–18/группу.

•Выживаемость/Каплан–Майер ($HR \approx 1.8-2.0$): 15–25/группу.



Контроль

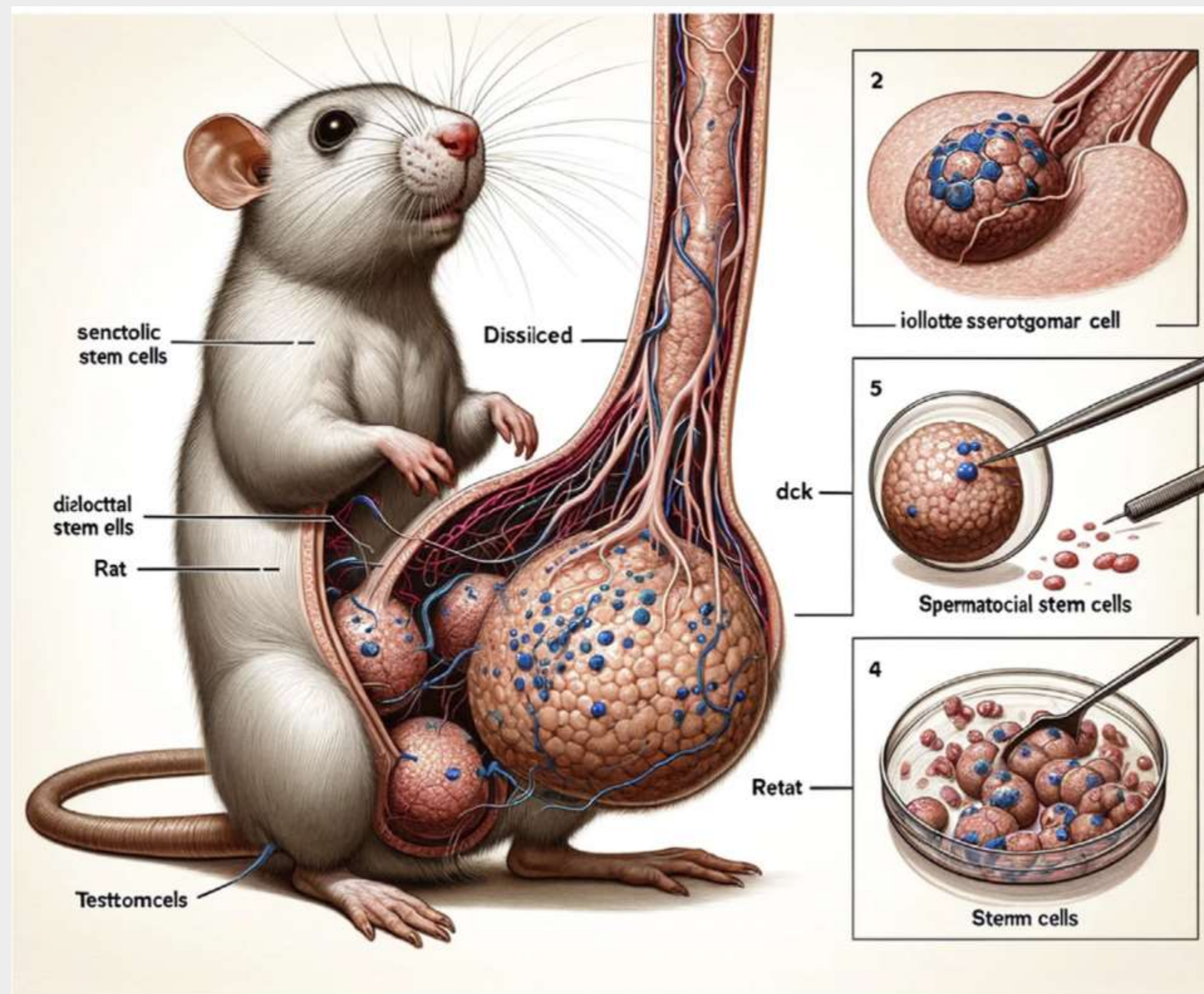
- **Виды контроля: отрицательный, положительный, плацебо.**
- **Конкурентный (в момент исследования) контроль > исторического (ранее полученные данные)**
- **единица контроля = единице анализа.**
- **Парные/матчированные контроли при высокой вариабельности.**
- **Стандартизация условий и процедур (время, оператор, партия, среда).**
- **Роль контроля в валидации модели и интерпретации нулевых результатов.**

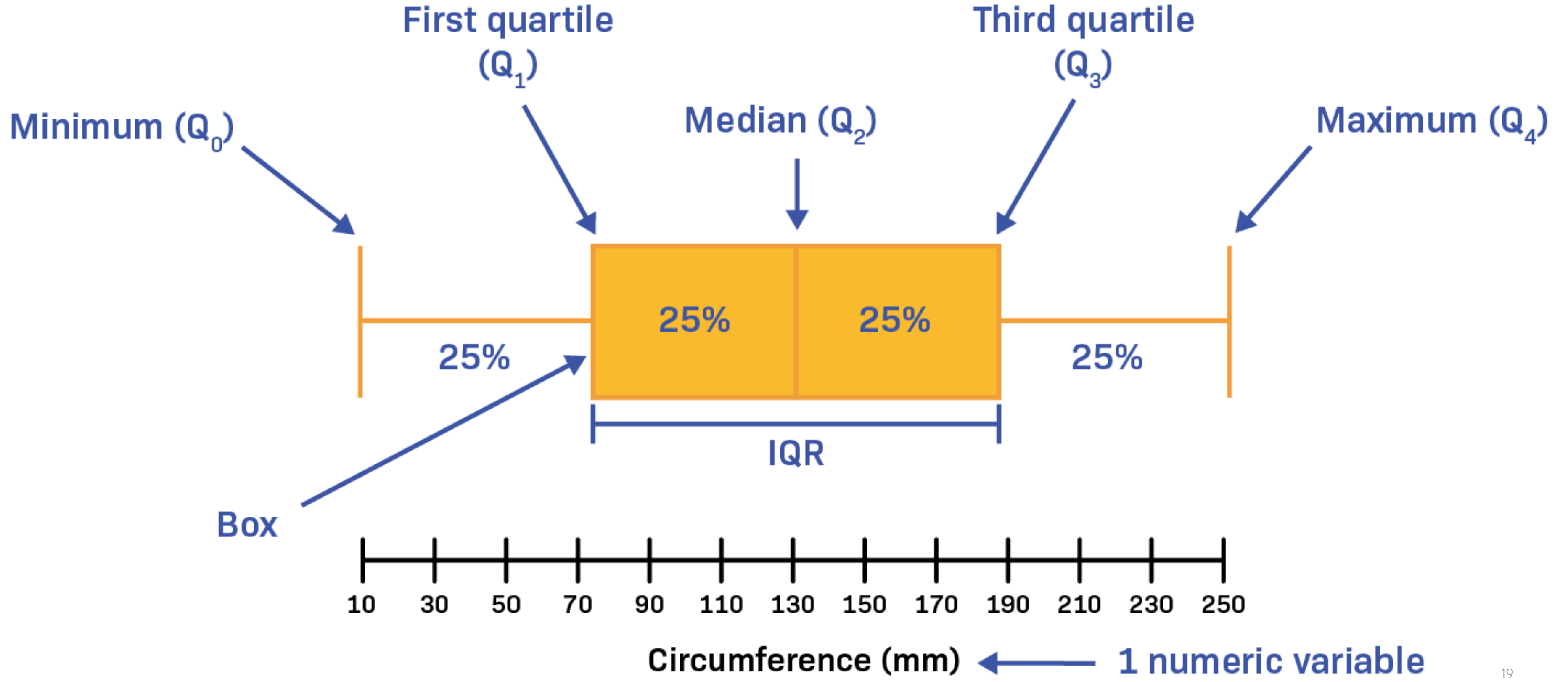
Рандомизация и сокрытие распределения

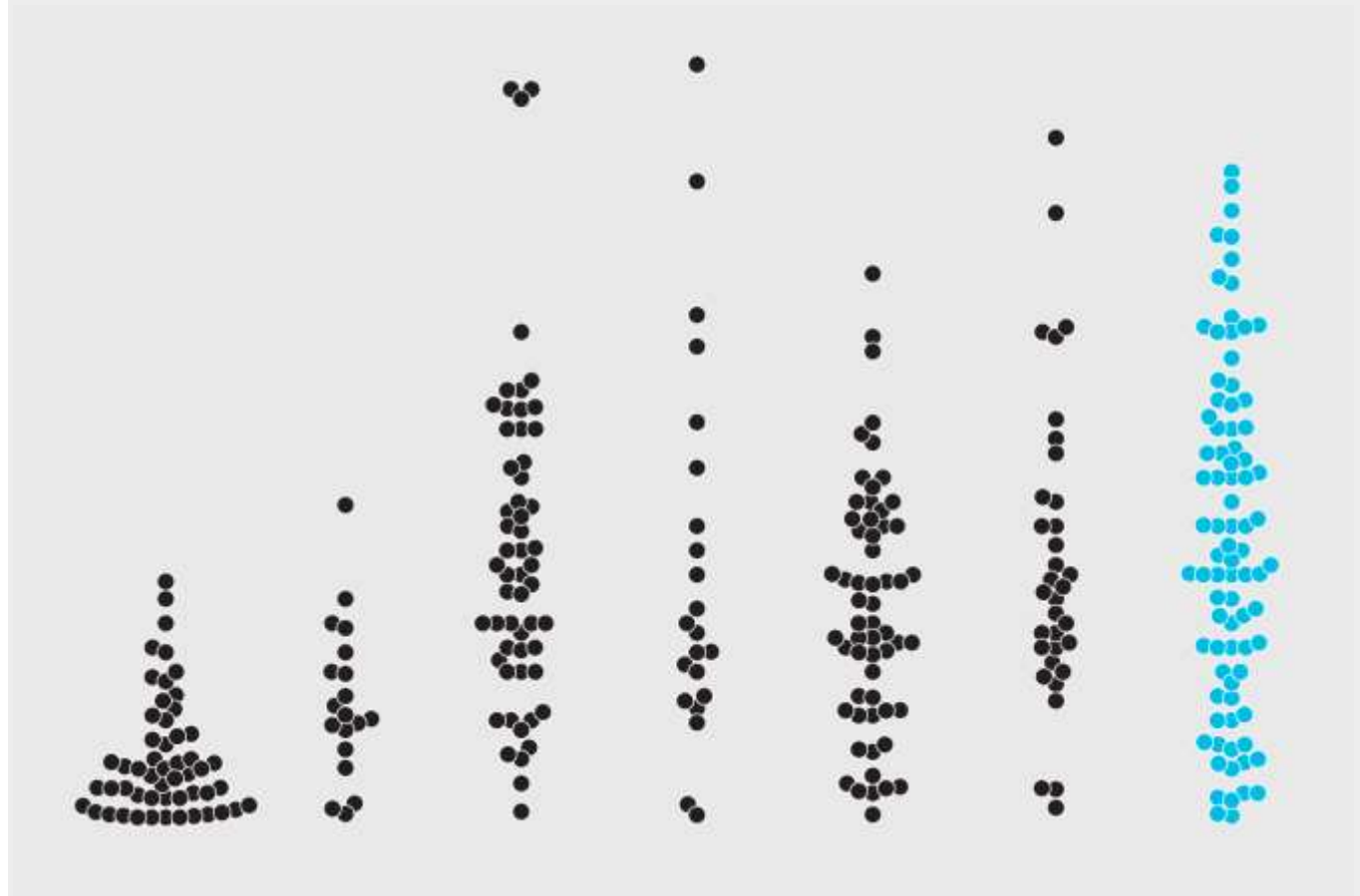
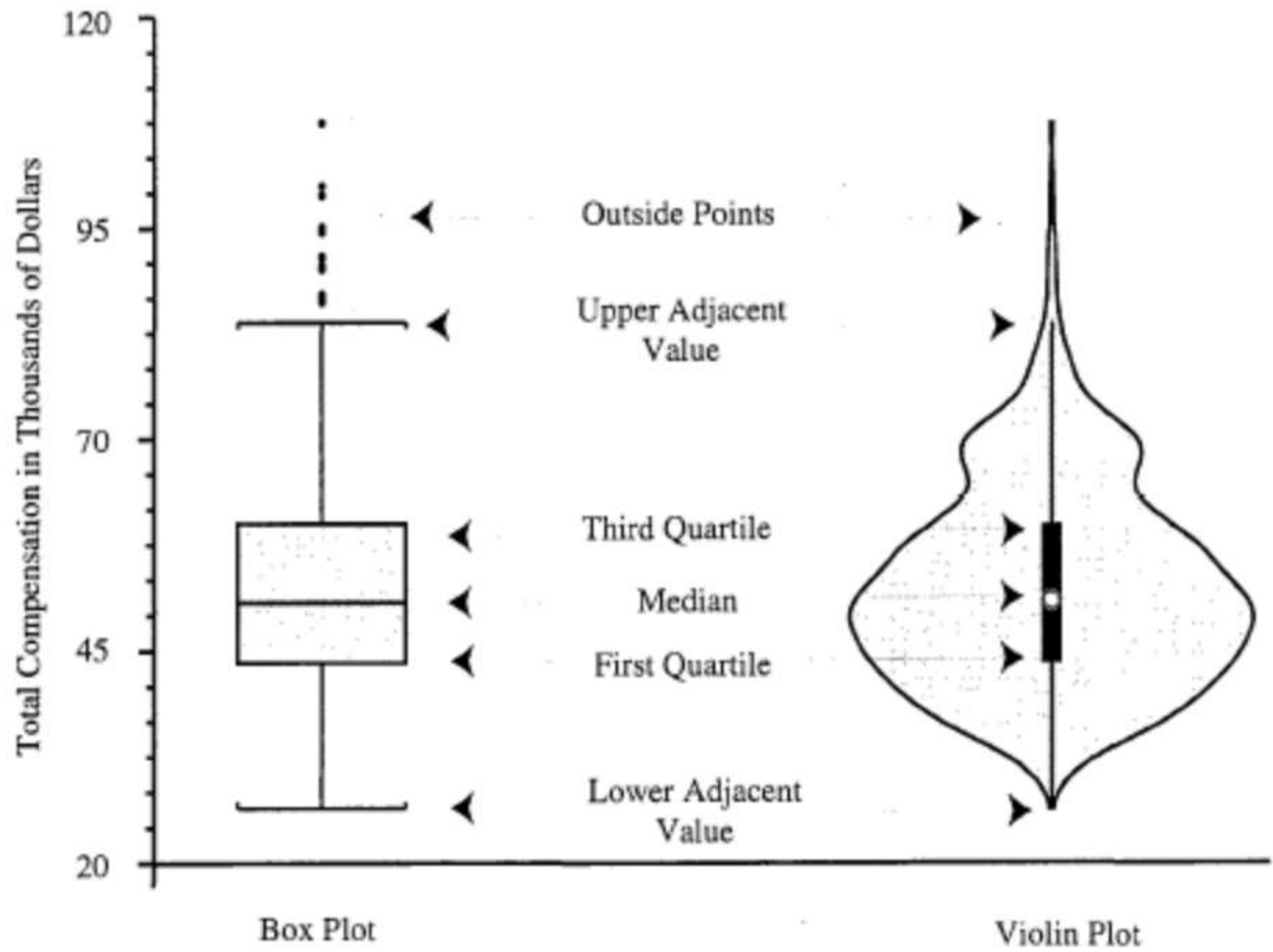
Цель: устранение систематического смещения; схемы – простая, блочная, стратифицированная, минимизация.

- **Кластерная рандомизация (помёты, планшеты) – учёт внутриклассовых корреляций и дифференциаций.**
- **Allocation concealment: независимая генерация, сокрытие до включения.**
- **Фиксация в SAP (План статистического анализа): алгоритм, блоки/страты, процедуры отклонений.**

Как мы показываем данные? Визуализация...





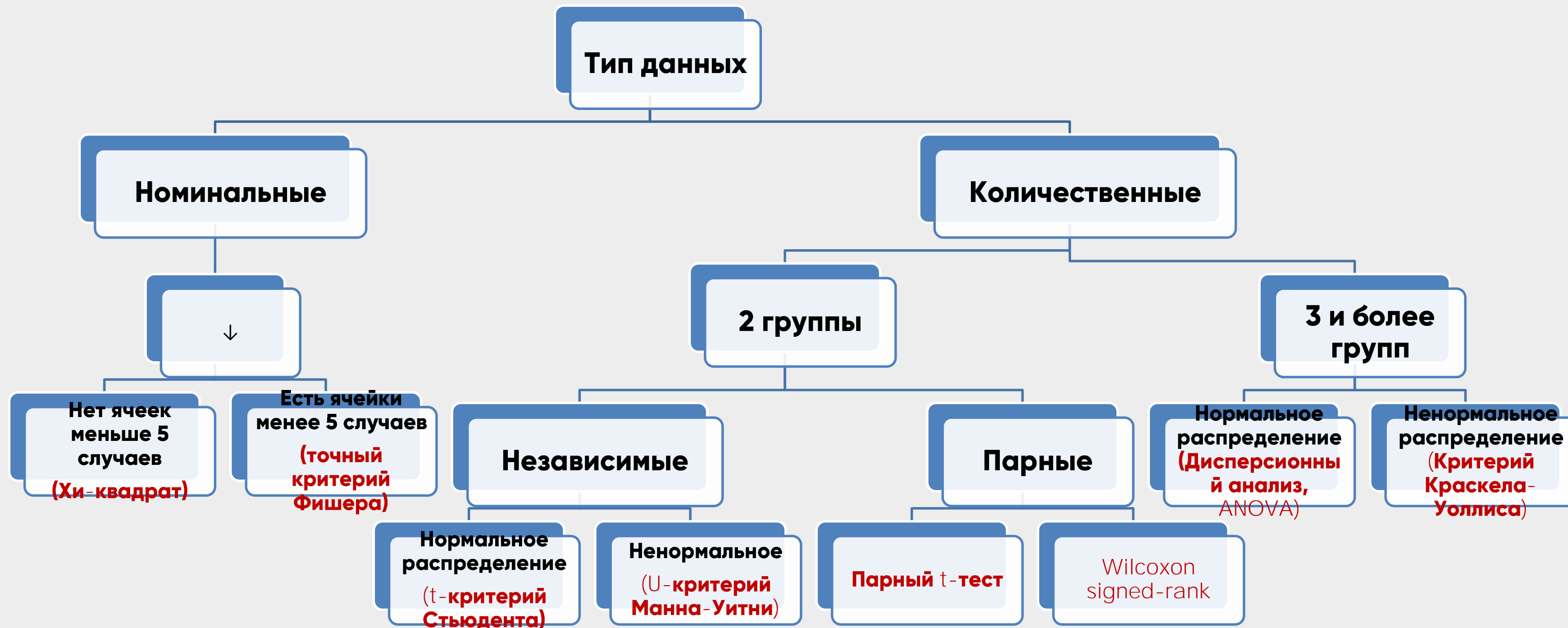


Описательная статистика и визуализация

- Центр: среднее vs медиана; разброс: SD, IQR, 95% CI.
- Показывайте сырые точки + сводку (violin/box + beeswarm).
- Малые выборки: медиана, IQR, бутстрэп-CI; избегать SE-«рожек».
- Проверка формы: гистограмма/плотность, Q-Q.



Выбор статистического критерия



**Чем мы можем воспользоваться,
чтобы облегчить себе жизнь?**



STATTECH

Размер выборки

Мои базы

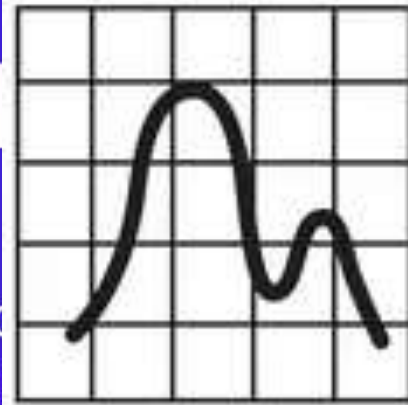
Мои опросы

Блог

Только 14 дней! 30% бонус для Вас и ваших

Программа для работы и анализа данных

с возможностью автоматического выбора
и оформлением результатов в формате



StatSoft[®]

STATISTICA

IBM

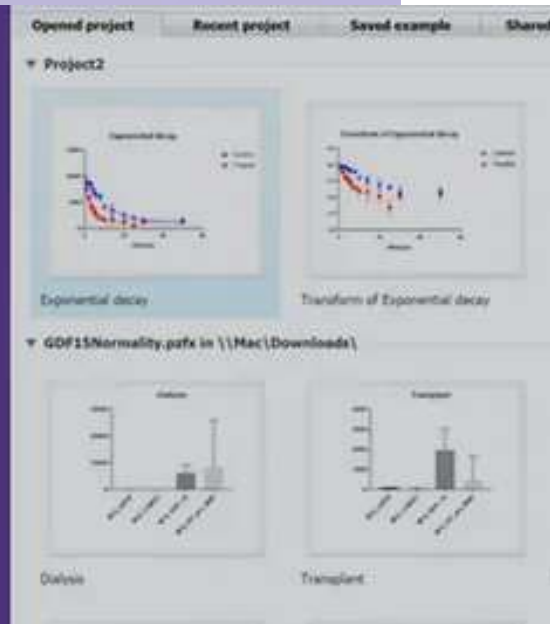


SPSS[®]

GraphPad



Prism



STATA[®]

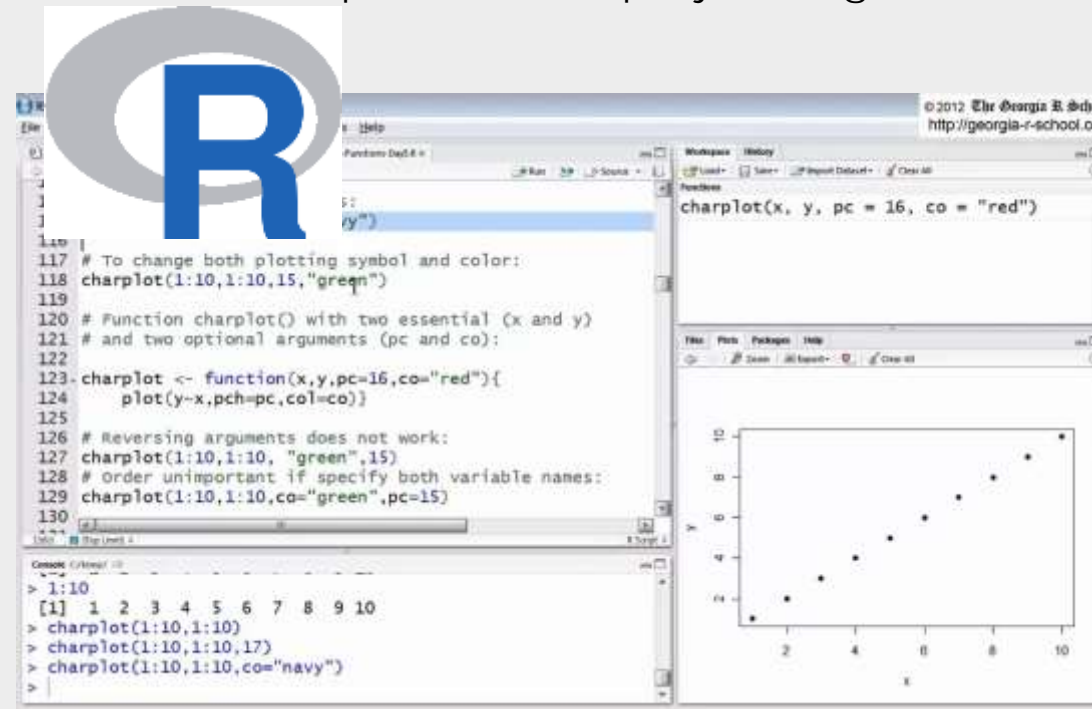
<https://past.en.lo4d.com/window>



Summary Statistics

PAST 4.06b Statistical Software

<https://www.r-project.org/>



```

117 # To change both plotting symbol and color:
118 charplot(1:10,1:10,15,"green")
119
120 # Function charplot() with two essential (x and y)
121 # and two optional arguments (pc and co):
122
123 charplot <- function(x,y,pc=16,co="red"){
124   plot(y~x,pch=pc,col=co)
125 }
126 # Reversing arguments does not work:
127 charplot(1:10,1:10, "green",15)
128 # order unimportant if specify both variable names:
129 charplot(1:10,1:10,co="green",pc=15)
130
> 1:10
[1] 1 2 3 4 5 6 7 8 9 10
> charplot(1:10,1:10)
> charplot(1:10,1:10,17)
> charplot(1:10,1:10,co="navy")

```

Бесплатные версии ПО для статистической обработки данных

G*Power

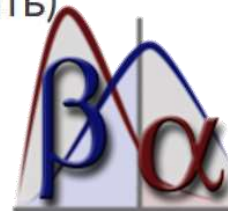
Statistical Power Analyses for Mac and Windows

G*Power — это инструмент для статистического анализа мощности различных t -критериев, F -критериев, χ^2 -критериев, z -критериев и некоторых точных критериев. G*Power также можно использовать для вычисления величины эффекта и графического отображения результатов анализа мощности.

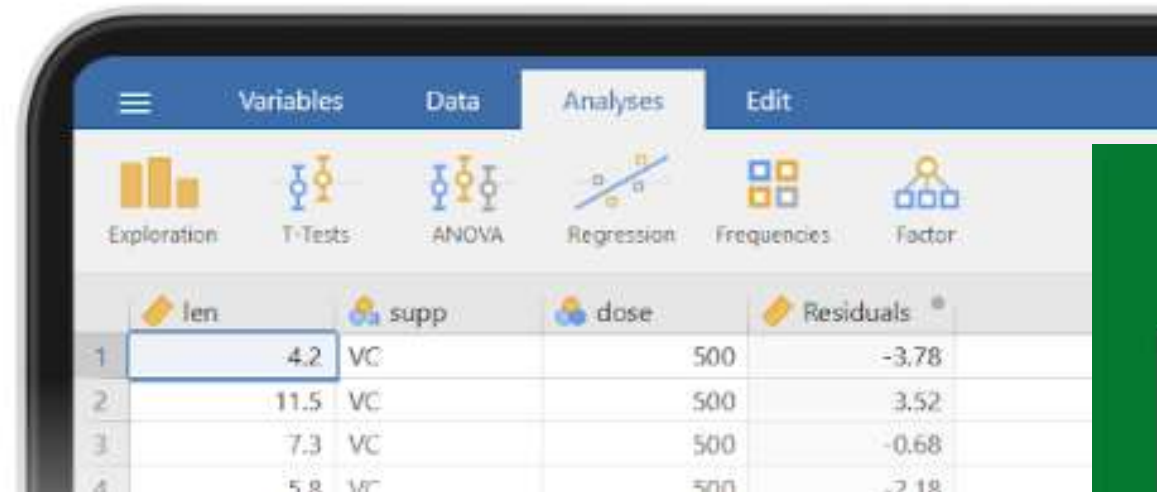
<https://www.jamovi.org/>



(нажмите, чтобы увеличить)



Главное окно
Главное окно (Таблица)
График мощности
График мощности (таблица)



Statulator

Conducts Statistical Analyses | Interprets the Results | Gives Reporting Advice

Sample Size | Descriptive Analysis | Statistical Tests



Excel

Где я могу научиться?



Курсы Лекторы Школьникам О проекте

Главная / Курсы / Математическая статистика. Лекции

Математика 20 лекции

Математическая статистика. Лекции

2022 Лекции Механико-Математический Факультет Математика V Семестр 3 Курс

Видеолекции Материалы О курсе

Анализ данных • С нуля

Бесплатный курс «Основы статистики и А/В-тестирования»

Научитесь применять статистику для решения рабочих задач и проводить корректные исследования

Начать учиться

★ 4,5

на Отзовике, IRecommend и TutorTop

10 000

выпускников
нашли работу
после обучения

Программа от опытных аналитиков

150+ практических задач и проект-симулятор

Объясняем простыми словами самые важные понятия

Учёба в удобном темпе, чтобы совмещать с другими делами

Алгоритм проведения А/В-тестов и проверки гипотез

Бессрочный доступ к теории и конспекты по курсу в подарок

Предмет математической статистики

Предмет изучения теории вероятностей можно охарактеризовать следующим образом:

математический анализ случайных явлений (неформально: зная природу некоторого явления, выяснить как будут вести себя характеристики, наблюдаемые в эксперименте).

teach-in

p-hacking

HARKing (выдвижение гипотез после получения результатов) — это аббревиатура, придуманная социальным психологом [Норбертом Керром^{\[1\]}](#), которая относится к сомнительной исследовательской практике «представления постфактум [гипотезы](#) во введении к отчёту об исследовании, как если бы это была [априорная](#) гипотеза»^{[1][2]} Следовательно, ключевой характеристикой HARKing является то, что выдвижение гипотез постфактум ошибочно воспринимается как выдвижение априорных гипотез.^[3] HARKing может возникнуть, когда исследователь проверяет априорную гипотезу, но затем исключает её из отчёта об исследовании после того, как узнаёт результаты проверки. [Постфактумный анализ](#) или [постфактумное теоретизирование](#) могут привести к появлению постфактумной гипотезы.

**Как это? Гипотеза формируется после просмотра результатов.
Иллюзия подтверждения и снижение воспроизводимости.**

Разделяйте исследовательский и подтверждающий анализ.

**Антидот: план до эксперимента и прозрачная отчётность.
Личная ответственность исследователя.**

Комитеты по биоэтике

В лекции были использованы Отдельные материалы и целые слайды из прекрасной лекции «Как правильно проанализировать результаты эксперимента на животных?» Д. Каримов Лекции по статистике Казанского федерального университета Статистика и котики, В. Савельев (идеально для начинающих) А так же множество изображений из сети Internet)

Сложные вопросы статистики в медико-биологических исследованиях хорошо отражены в работах Гржебовского А.М. (Elibrary AuthorID: 496304)

**Для интересующихся кодингом рекомендуем:
Р. Кабаков, 2014. R в действии, пер. П. Волковой**

**СПАСИБО
ЗА ВНИМАНИЕ!**