



10 АПРЕЛЯ 19:00 МСК
ОНЛАЙН

РАЗРУШИТЕЛИ СТАТИСТИЧЕСКИХ МИФОВ

МАТВЕЙ СЛАВЕНКО | МИФ №3:
НЕНОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ ТРЕБУЕТ НЕНОРМАЛЬНЫХ РЕШЕНИЙ

[HTTPS://T.ME/CHAT_BIOSTAT_R](https://t.me/chat_biostat_r)



План

1. Операционализация: конечные точки и эстиманды.
2. Статпроцедуры: из чего состоят, хорошие и плохие.
3. Z-тест, t-тест и допущение о нормальности.
4. Что не так с проверками на нормальность.
5. За(й)чатки разумного: чем заменить тесты на нормальность.
6. Дискотека.

Оговорки

Лекция для расширенного круга прикладных специалистов, использующих статистику. Изложение ставит в приоритет понятность перед полной технической корректностью.

Операционализация



Она же: endpoint / конечная точка, outcome / исход, (целевая) метрика и т. д.

То, что мы измеряем. Численное, измеримое выражение концепта.

Операционализация



Она же: endpoint / конечная точка, outcome, (целевая) метрика и т. д.

Пример 1.

Наталья — страховщик, она страхует недвижимое имущество. Например:

1. Дом клиента №1 она застраховала на три года за 3000 тугриков. Три года были спокойными, страховых выплат не было.
(+3000)
2. Дачу клиента №2 она застраховала на год за 2000 тугриков. Зимой дачу ограбили, поэтому пришлось возместить ущерб в размере 4000 тугриков.
(-2000)
3. Квартиру клиента №3 она застраховала на три года за 1500 тугриков. Квартиру затопили соседи сверху, и Наталья компенсировала расходы на ремонт в размере 500 тугриков.
(+1000)

Операционализация



Она же: endpoint / конечная точка, outcome, (целевая) метрика и т. д.

Пример 1.

Наталья — страховщик, она страхует недвижимое имущество

Наталья хочет научиться формировать резервы на покрытие страховых выплат, чтобы не разориться. В этом смысле:

- **наблюдение** — это конкретный страховой полис,
- **конечная точка** — доход / убыток от полиса.

Операционализация



Она же: endpoint / конечная точка, outcome, (целевая) метрика и т. д.

Пример 2.

Николай — психолог, он изучает силу воли у детей. Он оставляет ребенка в комнате с зефиркой на 15 минут и обещает по возвращении дать еще одну, если ребенок зефирку не съест в его отсутствие.

- **наблюдение** — поведение ребенка в конкретном опыте,
- **конечная точка** — съел или нет (1/0), как вариант сколько минут продержался (число от 0 до 15 минут).

Операционализация



Она же: endpoint / конечная точка, outcome, (целевая) метрика и т. д.

Пример 3.

Светлана — диabetолог, изучает полузакрытые системы менеджмента гликемии (помпа + монитор гликемии) для диабетиков первого типа. Она предлагает своим пациентам использовать систему и наблюдает их в течение трех месяцев. Здесь:

- **наблюдение** — измерения конкретного пациента после трех месяцев,
- **конечная точка** — концентрация гликированного гемоглобина в крови пациента (HbA1c) через три месяца ношения системы.

Операционализация



Конечные точки иногда делят на жесткие и суррогатные, метрики ключевые и прокси и т. д.

Можно сказать, что это деление описывает, насколько тесно операнты (конечные точки) связаны с исследуемым концептом.

- Страховые полисы — доход / убыток от полиса.
- Сила воли — да / нет или минуты до поедания зефирки.
- Диабет — концентрация гликированного гемоглобина через три месяца.

1. Структура социологической концептуализации на Постнауке, [ссылка](#)
2. Социологическая операционализация на Постнауке, [ссылка](#)

Операционализация



Насколько тесно операнты (конечные точки) связаны с исследуемым концептом?

- Страховые полисы — доход / убыток от полиса.
 - Тут все понятно.
- Сила воли — да / нет или минуты до поедания зефирки.
- Диабет — концентрация гликированного гемоглобина через три месяца.

Операционализация



Насколько тесно операнты (конечные точки) связаны с исследуемым концептом?

- Страховые полисы — доход / убыток от полиса.
- Сила воли — да / нет или минуты до поедания зефирки.
 - Почему зефирка? Может, печенье, чупа-чупс или пять долларов?
 - Почему 15 минут, а не 14 или 20?
 - Почему в комнате в одиночестве, а не на игровой площадке?
 - Схватывает ли эта операционализация именно силу воли? или что-то еще?
- Диабет — концентрация гликированного гемоглобина через три месяца.

Операционализация



Насколько тесно операнты (конечные точки) связаны с исследуемым концептом?

- Страховые полисы — доход / убыток от полиса.
- Сила воли — да / нет или минуты до поедания зефирки.
- Диабет — концентрация гликированного гемоглобина через три месяца.
 - Химическое и физиологическое обоснование (время жизни эритроцита, необратимая связь глюкозы и гемоглобина).
 - В дополнение или на замену: TIR, средняя (по времени) гликемия...

Операционализация



Нам нужно схлопнуть ряд чисел до какого-то одного показателя, который характеризует его —> *эстиманд*, также интересующий параметр распределения.

- Страховые полисы — доход / убыток от полиса.
- Сила воли — да / нет или минуты до поедания зефирки.
- Диабет — концентрация гликированного гемоглобина через три месяца.

Операционализация



Нам нужно схлопнуть ряд чисел до какого-то одного показателя, который характеризует его —> *эстиманд*.

- Страховые полисы — доход / убыток от полиса.
 - Наталье нужно контролировать возможные потери. Поэтому 99-й перцентиль расходов (value-at-risk модели, квантильная регрессия): с вероятностью 99% расходы от портфеля не превысят его.
- Сила воли — да / нет или минуты до поедания зефирки.
- Диабет — концентрация гликированного гемоглобина через три месяца.

Операционализация



Нам нужно схлопнуть ряд чисел до какого-то одного показателя, который характеризует его —> *эстиманд*.

- Страховые полисы — доход / убыток от полиса.
- Сила воли — да / нет или минуты до поедания зефирки.
- Диабет — концентрация гликированного гемоглобина через три месяца.
 - Гликемия должна находиться в оптимальном поясе, слишком высокая или низкая гликемия — это опасно. Поэтому средняя (матожидание) концентрация гликированного гемоглобина (систематическое поведение).
 - В дополнение — sd HbA1c как мера стабильности работы системы.

Операционализация



Нам нужно схлопнуть ряд чисел до какого-то одного показателя, который характеризует его —> *эстиманд*.

- Страховые полисы — доход / убыток от полиса.
- Сила воли — да / нет или минуты до поедания зефирки.
 - ???
- Диабет — концентрация гликированного гемоглобина через три месяца.

Свои версии присылайте в t.me/chat_biostat_R



Операционализация. Итог



Помним о разделении *статистического* и *научного* вывода (statistical vs. scientific inference), см. лекцию Ольги Мироненко.

Статвывод — инструмент и часть научного вывода.

Выбор конечной точки (операнта) внешний по отношению к статвыводу и подчинен научному выводу (доменным знаниям).

Выбор эстиманда — первый шаг статанализа. Выбор диктуется:

- целями научного вывода: насколько им соответствуют математические свойства эстиманда?
- дизайном эксперимента;
- нашими техническими возможностями (t-тесты и линейная регрессия сильно доступнее квантильной регрессии; с экзотическими эстимандами сложнее работать);
- см. лекцию Евгения Бакина, первую в серии.

Тесты и ДИ



1. Статистическая / вероятностная модель порождения данных.
2. Эстиманд.
3. Процедура: эстиматор, тест, ДИ.
4. Требования к процедуре при данном объеме выборки.
Для теста: вероятность ошибки первого рода ($\alpha = 0.05$ или другое).
Для ДИ: вероятность покрытия ($1 - \alpha = 0.95$).

Тесты и ДИ



	СТРОГОЕ	ОСЛАБЛЕННОЕ
Структура зависимости Наблюдения...	независимы	некоррелированы
Распределение Наблюдения происходят из...	нормального распределения	распределения с конечной ненулевой дисперсией
Точная vs. асимптотическая процедура Альфа обеспечивается...	даже на малых выборках	на достаточно большой выборке

Тесты и ДИ



	СТРОГОЕ	ОСЛАБЛЕННОЕ
Структура зависимости Наблюдения...	независимы	некоррелированы
Распределение Наблюдения происходят из...	нормального распределения	распределения с конечной ненулевой дисперсией
Точная vs. асимптотическая процедура Альфа обеспечивается...	даже на малых выборках	на достаточно большой выборке

ДОПУЩЕНИЯ НЕ ВСЕГДА МОЖНО ПРОВЕРИТЬ СТАТИСТИЧЕСКИ!

Тесты и ДИ



	СТРОГОЕ	ОСЛАБЛЕННОЕ
Структура зависимости Наблюдения...	независимы	некоррелированы
Распределение Наблюдения происходят из...	нормального распределения	распределения с конечной ненулевой дисперсией
Точная vs. асимптотическая процедура Альфа обеспечивается...	даже на малых выборках	на достаточно большой выборке

Какие допущения стоят за t-тестом?
Какие из допущений можно проверить?

Тесты и ДИ



Что такое хорошая процедура? Ответ на этот вопрос складывается из:

1. поведения процедуры при верности модели,
2. поведения процедуры при нарушениях допущений модели.

Тесты и ДИ



Как процедура ведет себя при условии, что статмодель верна?

Для тестов. Поведение при:

1. нулевой гипотезе, состоятельность теста: соблюдает ли тест предписанную альфу?
2. при альтернативе (альтернативной гипотезе), несмещенность теста: какова мощность теста против данной альтернативы, т. е. какова вероятность отвергнуть нулевую гипотезу при данной альтернативе? Эта вероятность больше альфы?

Эти вопросы можно задать себе в ситуации:

- “достаточно больших выборок” — асимптотическая состоятельность и несмещенность;
- “малых / конечных выборок” — точная процедура, (строгая) состоятельность и несмещенность.



Тесты и ДИ

Как процедура ведет себя при условии, что статмодель верна?
Для ДИ. При разных истинных значениях эстиманда:

1. соблюдает ли ДИ предписанную вероятность покрытия? (состоятельность);
2. какова ожидаемая ширина ДИ? (“эффективность”)

Эти вопросы можно задать себе в ситуации:

- “достаточно больших выборок” — асимптотическая состоятельность и эффективность;
- “малых / конечных выборок” — точная процедура, (строгая) состоятельность.

Тесты и ДИ



Поведение процедуры при неверно заданной модели, т. е. устойчивость (робастность) процедуры к нарушению допущений.

Модель предполагает, что...

1. наблюдения независимы. А если они коррелированы?
2. наблюдения происходят из нормального распределения. А если они из равномерного / логнормального / экспоненциального / произвольного распределения?
3. и т. д.

Ослабляя допущения, мы получаем более общую статмодель.

Оговорка



Дальнейшие рассуждения ведутся, в основном, на примере одновыборочного t -теста. Моя цель — показать ход мысли и основные идеи.

Все то же самое можно обобщить на двухвыборочные тесты, многовыборочные тесты (дисперсионный анализ) и линейные модели. Идеи остаются теми же, но их исполнение требует более сложных и продвинутых инструментов (теорем, методов теорвера и т. д.).

Асимптотический Z-тест



Асимптотический одновыборочный Z-тест

Допущения:

1. Наблюдения независимы, происходят из одного распределения (iid).
2. Наблюдения X_1, \dots, X_n обладают конечной ненулевой дисперсией, т. е.
 $0 < \text{var}(X_i) < \infty$.
3. Выборка достаточно велика.

$$H_0: E[X_i] = m_0$$

vs.

$$H_1: E[X_i] \neq m_0$$

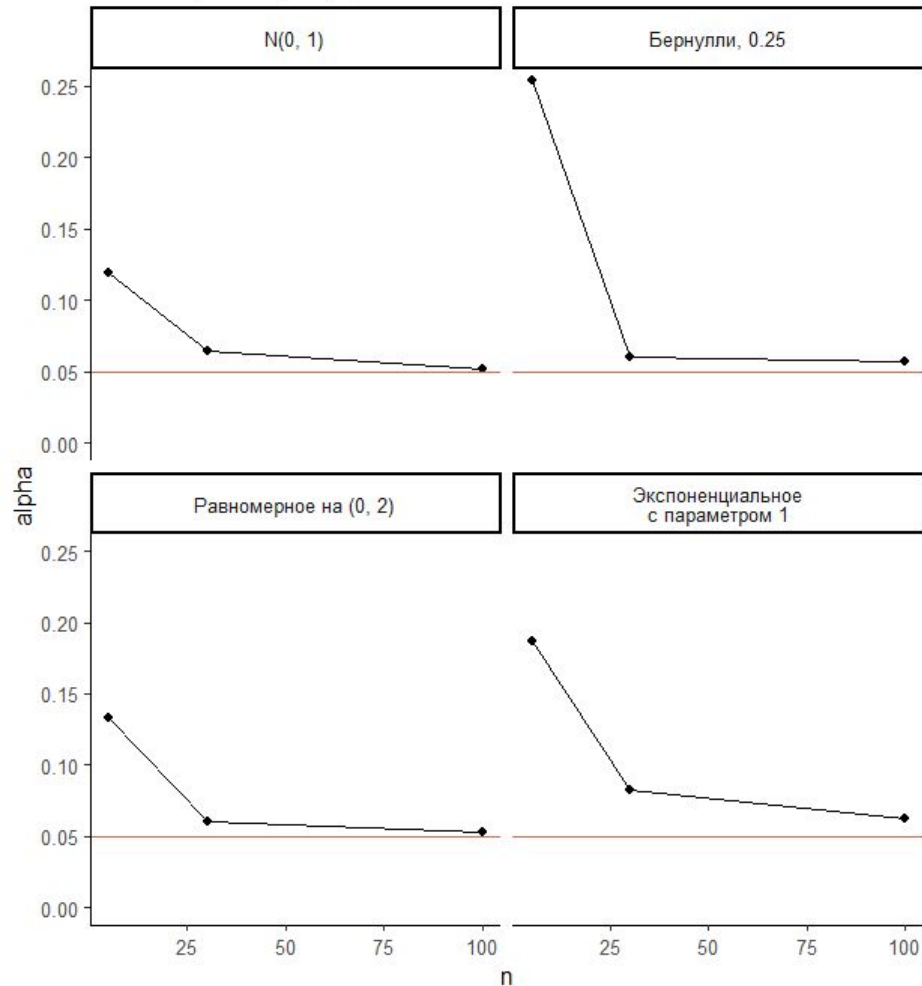
Тестовая статистика: $T_n := \sqrt{n} \frac{\bar{X}_n - m_0}{\sqrt{S_n^2}}$, при H_0 сходится к $N(0; 1)$.

Критическая область: отвергни H_0 , если $|T_n| > u(0.975)$.

$$\text{ДИ: } \bar{X}_n \mp u(0.975) \frac{\sqrt{S_n^2}}{\sqrt{n}}$$

Поведение Z-теста при H_0 в зависимости от распределения

Монте-Карло симуляции, $B = 10\,000$



Распределение	Объем выборки N		
	5	30	100
Экспоненциальное с параметром 1	0.1868	0.0827	0.0630
Равномерное на (0, 2)	0.1334	0.0606	0.0532
Бернулли, 0.25	0.2541	0.0596	0.0572
$N(0, 1)$	0.1192	0.0642	0.0516



Точный t-тест

Допущения:

1. Наблюдения независимы и происходят из одного распределения.
2. Наблюдения X_1, \dots, X_n происходят из нормального распределения.

$$H_0: E[X_i] = m_0$$

vs.

$$H_1: E[X_i] \neq m_0$$

$$T_n := \frac{\sqrt{n}(\bar{X}_n - m_0)}{\sqrt{S_n^2}}, \text{ распределено как } t_{n-1} \text{ при } H_0.$$

Критическая область: отвергни H_0 , если $|T_n| > t_{n-1}(0.975)$.

$$\text{ДИ: } \bar{X}_n \mp t_{n-1}(0.975) \frac{\sqrt{S_n^2}}{\sqrt{n}}$$

Точный t-тест



Допущения t-теста:

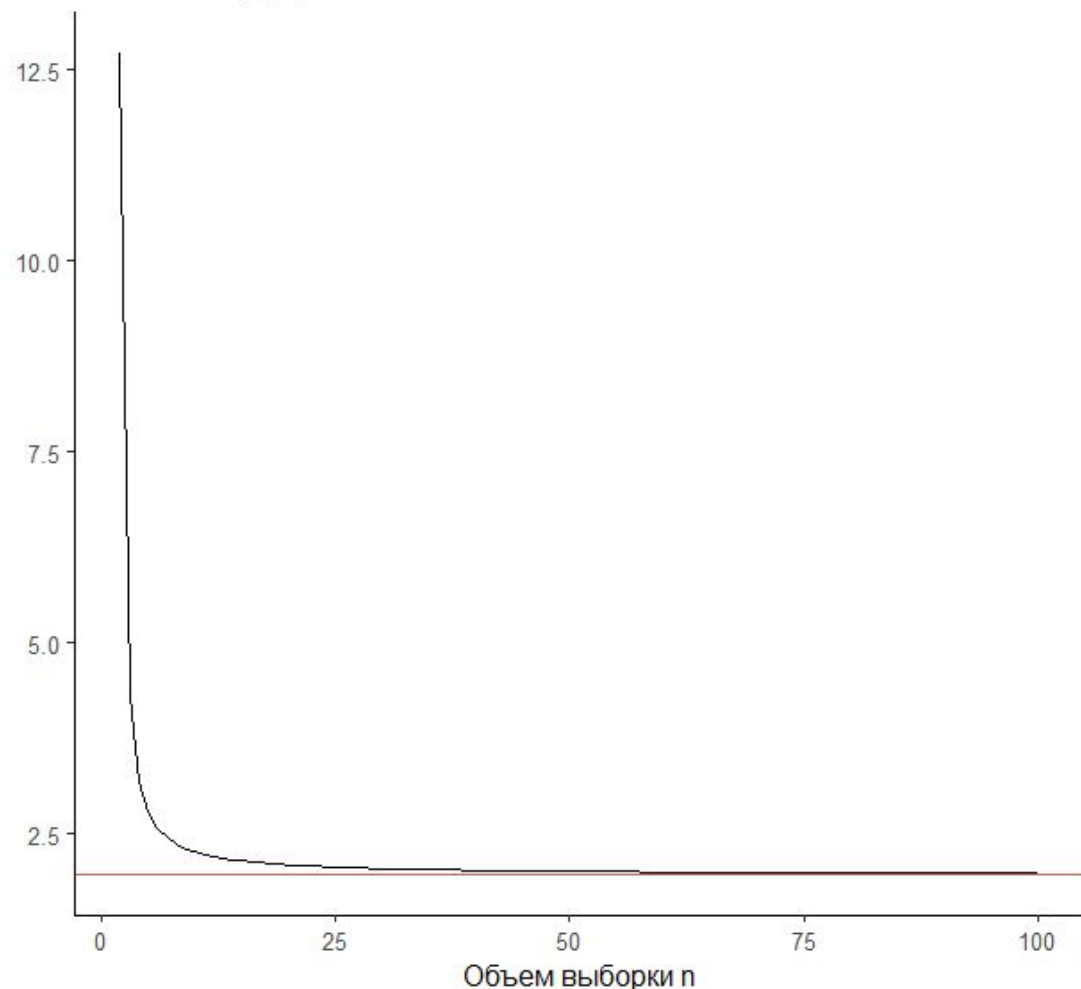
1. Наблюдения независимы и происходят из одного распределения.
2. Наблюдения X_1, \dots, X_n происходят из нормального распределения.

Допущения Z-теста:

1. Наблюдения независимы, происходят из одного распределения (iid).
2. Наблюдения X_1, \dots, X_n обладают конечной ненулевой дисперсией, т. е.
 $0 < \text{var}(X_i) < \infty$.
3. Выборка достаточно велика.

Проверим на нормальность. Если нормальность выполняется, то t-тест.
Если нет, то Z-тест.
Ведь так? Так ведь?...

Квантили t-распределения в зависимости от объема выборки и квантиль $N(0, 1)$



отвергни H_0 , если $|T_n| > t_{n-1}(0.975)$.

отвергни H_0 , если $|T_n| > u(0.975)$.



Асимптотический t-тест



Допущения:

1. Наблюдения независимы и происходят из одного распределения
2. Наблюдения X_1, \dots, X_n происходят из нормального распределения.
ЛИБО
1. Наблюдения X_1, \dots, X_n обладают конечной ненулевой дисперсией, т. е.
 $0 < \text{var}(X_i) < \infty$.

$$H_0: E[X_i] = m_0$$

vs.

$$H_1: E[X_i] \neq m_0$$

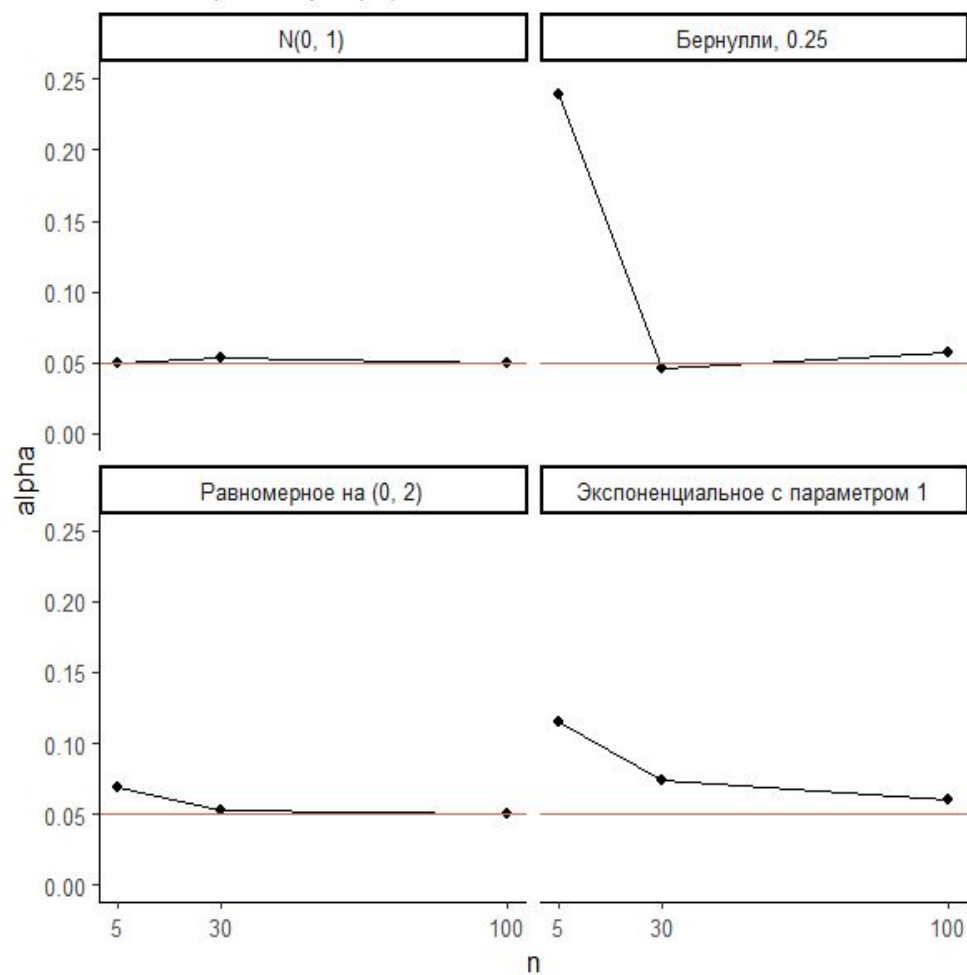
$T_n := \frac{\sqrt{n}(\bar{X}_n - m_0)}{\sqrt{S_n^2}}$, распределено приблизительно как t_{n-1} при H_0 .

Критическая область: отвергни H_0 , если $|T_n| > t_{n-1}(0.975)$.

$$\text{ДИ: } \bar{X}_n \mp t_{n-1}(0.975) \frac{\sqrt{S_n^2}}{\sqrt{n}}$$

Поведение t-теста при H_0 в зависимости от распределения

Монте-Карло симуляции, $B = 10\,000$



Распределение	Объем выборки N		
	5	30	100
Экспоненциальное с параметром 1	0.1868	0.0827	0.0630
Равномерное на $(0, 2)$	0.1334	0.0606	0.0532
Бернулли, 0.25	0.2541	0.0596	0.0572
$N(0, 1)$	0.1192	0.0642	0.0516

Распределение	Объем выборки N		
	5	30	100
Экспоненциальное с параметром 1	0.1147	0.0733	0.0605
Равномерное на $(0, 2)$	0.0690	0.0522	0.0508
Бернулли, 0.25	0.2395	0.0458	0.0572
$N(0, 1)$	0.0502	0.0538	0.0493



Итог: допущения и зачем они

Не все допущения можно проверить статистическими инструментами. Часто они обеспечиваются природой данных и дизайном эксперимента.

Если допущения выполнены, то у нас есть основания считать, что процедура оперирует как следует.

Если допущения не выполнены, то из этого не следует автоматически, что все плохо. Скорее это повод предложить новую модель и проводить дополнительные исследования того, как ведет себя процедура в новых условиях.



Итог: допущения и зачем они

Если допущения не выполнены, то из этого не следует автоматически, что все плохо. Скорее это повод предложить новую модель и проводить дополнительные исследования того, как ведет себя процедура в новых условиях.

Мы можем найти более широкие и разумные условия, в которых теоретически покажем, что процедура работает хорошо (см. реформулировка t-теста).

Может оказаться так, что теоретических обоснований найти не получается, но симуляции показывают, что процедура ведет себя хорошо (см. перестановочные тесты).

Может действительно поломаться.

Что не так с проверками?



Обычно в шпаргалках предлагают подход: протестируем на нормальность.

Если $p < 0.05$, то используем манна-уитни. Если нет, то t-тест.

Проблема: это бессмысленная проверка. Едва ли вы найдете данные, которые действительно распределены по нормальному закону. Т. е. при достаточно большой выборке тест на нормальность отвергнет гипотезу.



5.4. ONE-SAMPLE WILCOXON TEST (WILCOXON SIGNED-RANK TEST)

This test assumes a **symmetric** distribution and it compares **the center of the symmetry** with a given constant.

Model: $\mathcal{F} = \{\text{continuous distribution with the density } f \text{ that satisfies } \exists \delta \in \mathbb{R} : f(\delta - x) = f(\delta + x) \forall x \in \mathbb{R}\}$

The parameter being tested: the center of the symmetry δ_X

Remark. The model requires the **density** of X_i being **symmetric** around the point δ_X . Then it holds that $m_X = \delta_X$ and if moreover $X_i \in \mathcal{L}^1$, then also $E X_i \equiv \mu_X = \delta_X$.

The hypothesis and the alternative:

$$H_0 : \delta_X = \delta_0, \quad H_1 : \delta_X \neq \delta_0,$$

where δ_0 is a given constant.

Remark. Provided that model \mathcal{F} holds then the hypothesis H_0 is equivalent to the hypothesis $H_0^* : m_X = \delta_0$ (i.e. we are resting the median). Further if $X_i \in \mathcal{L}^1$, then the hypothesis H_0 is also equivalent to the hypothesis $H_0^{**} : \mu_X = \delta_0$ (i.e. we are testing the expected value).

Asymmetry. When the density f is not symmetric, then the parameter being tested is not the median of X_i but the so called *pseudo-the median* that is the median of the random variable $\frac{X_1 + X_2}{2}$. The problem of the pseudo-median is that it is difficult to interpret. Generally one can also say that its value lies between the median m_X and the expected value $E X_i$ (provided that this expectation exists).

The next unpleasant consequence of the asymmetry of our observations is that even if view the one-sample Wilcoxon test as the test of the pseudo-median than its actual/true level (exact as well as asymptotic) is different from the prescribed level α . Nevertheless the simulation experiments show that the difference of the true level from the prescribed level is not large even for rather asymmetric distributions. Thus when the data are not obviously asymmetric then the main problem of the one-sample Wilcoxon test is the interpretation of the pseudo-median.

Что не так с проверками?



Обычно в шпаргалках предлагают подход: протестируем на нормальность.

Если $p < 0.05$, то используем манна-уитни. Если нет, то t-тест.

Проблема философская. Вообще-то мы должны выбирать эстиманд и формулировать гипотезу до сбора данных, исходя из сути задачи. Здесь же телега ставится впереди лошади: основная гипотеза становится зависимой от данных (помним, что Манн-Уитни не сравнивает ни средние, ни медианы) и выбирается случайным образом.

Пример: таблица 1 (mean and sd vs. median and IQR)

Что не так с проверками?



Обычно в шпаргалках предлагают подход: протестируем на нормальность.

Если $p < 0.05$, то используем манна-уитни. Если нет, то t-тест.

Проблема интерпретации. Каждая процедура исходит из конкретного эстиманда.

Какой эстиманд “держит в уме” двухфазовый подход? Какую гипотезу он тестирует?

Что не так с проверками?



Обычно в шпаргалках предлагают подход: протестируем на нормальность.

Если $p < 0.05$, то используем манна-уитни. Если нет, то t-тест.

Проблема множественного тестирования. Вы проводите два теста. Более того, результаты этих тестов не являются независимыми. Какие свойства у такой процедуры?

Что не так с проверками?



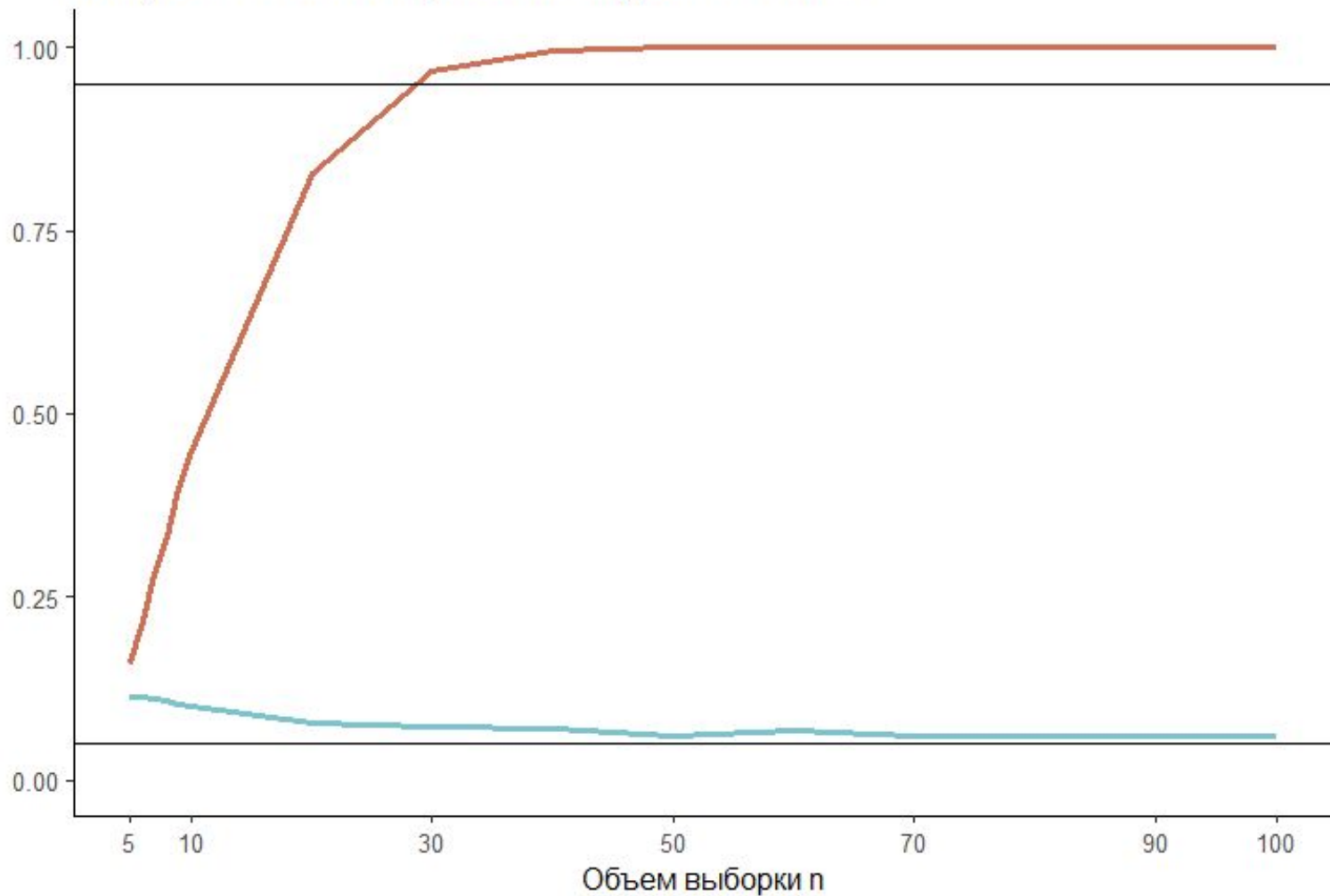
Проблема теоретическая. Вспомним, что без дополнительных телодвижений тест дает два ответа: “отвергаем гипотезу о том, что выборка происходит из нормального распределения” или “не могу отвергнуть гипотезу”.

То есть, если $p < 0.05$, у нас есть основания усомниться в нормальности и, возможно, посмотреть в сторону других инструментов.

Однако если $p > 0.05$, у нас нет оснований считать, что условие нормальности выполнено.

Вдобавок, тесты на нормальность имеют малую мощность: т. е. не слишком успешно детектируют отклонение от нормальности, особенно на малых выборках.

Мощность теста Шапиро-Уилка и уровень t-теста



На примере
 $\text{Exp}(1)$

Что не так с проверками?



Вдобавок, тесты на нормальность имеют малую мощность: т. е. не слишком успешно детектируют отклонение от нормальности, особенно на малых выборках.

В сумме имеем парадокс:

- На малых выборках нам “нужна” нормальность, но мы не можем ее проверить.
- На больших выборках мы можем проверить допущение о нормальности, но оно нас уже не интересует.

Что не так с проверками?

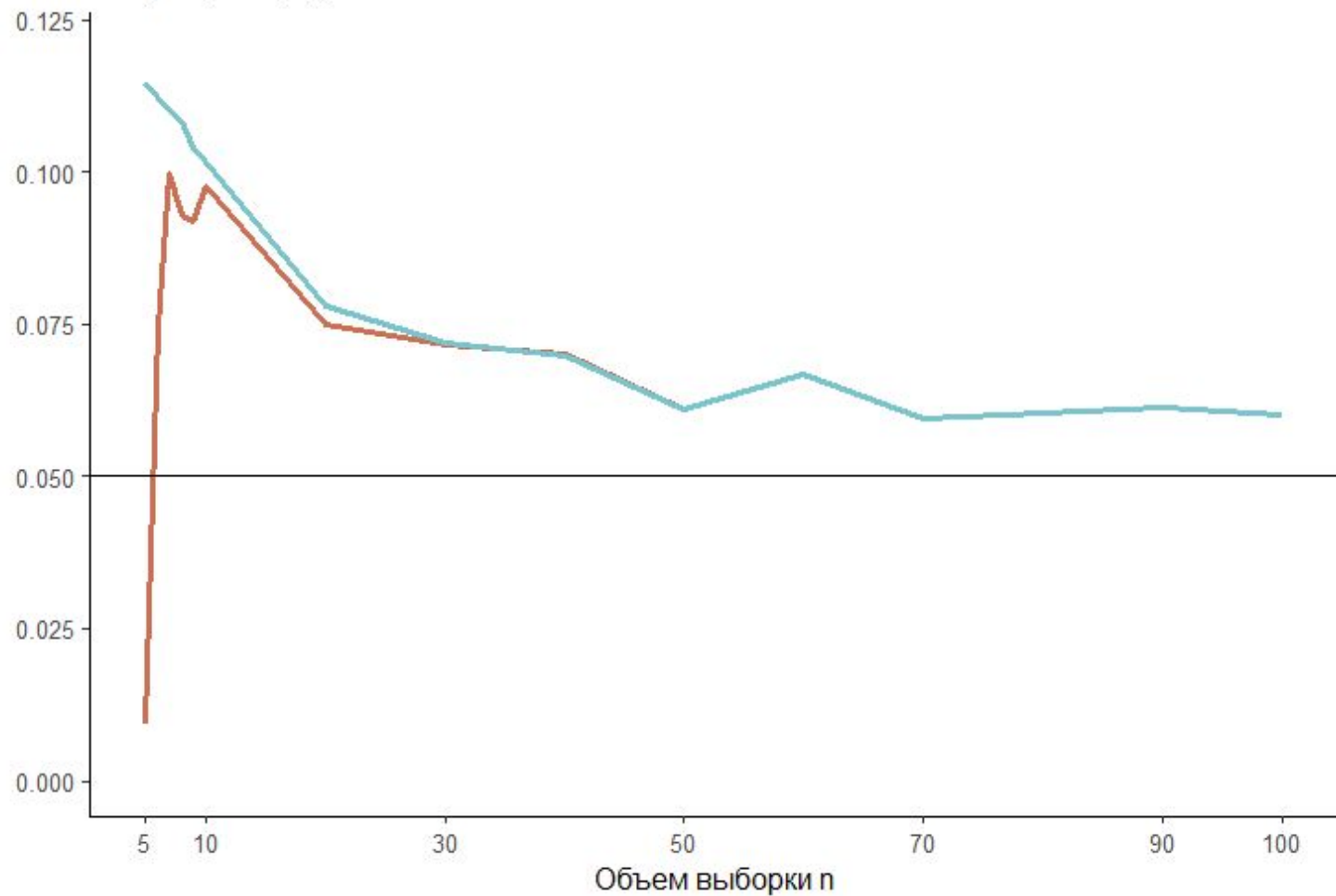


Обычно в шпаргалках предлагают подход: протестируем на нормальность.

Если $p < 0.05$, то используем манна-уитни. Если нет, то t-тест.

Проблема прикладная. У такой процедуры проблемы с соблюдением предписанного уровня и мощностью

Уровень двухфазовой процедуры и t-теста на примере Exp(1)



Что не так с проверками?



Обычно в шпаргалках предлагают подход: протестируем на нормальность.

Если $p < 0.05$, то используем манна-уитни. Если нет, то t-тест.

Проблема “туннельного зрения”. Какие допущения у “непараметрических” тестов? Почему допущение о нормальности надо проверять, а допущения “непараметрики” — нет:

- симметричность функции плотности для одновыборочного Вилкоксона?
- модель сдвига для Манна-Уитни?

Что не так с проверками?



Обычно в шпаргалках предлагают подход: протестируем на нормальность.

Если $p < 0.05$, то используем манна-уитни. Если нет, то t-тест.

Проблема воспроизводимости. Ваша гипотеза выбирается случайным образом на основе данных. В репликации эксперимента вам может выпасть другой метод.



Зерно разумного и что (не) делать

- Бутстреп и трансформации данных (логарифмирование, Бокс-Кокс) не являются панацеей.
- Know your tools.
 - У процедур есть разные формулировки с разными допущениями.
 - Знайте, как ведут себя процедуры в разных условиях.
- Know your data.
- Существуют другие полностью параметрические тесты, не только основанные на нормальности (см. тесты в Пуассоновских или экспоненциальных моделях, например; сюда же тесты для бинарных данных).
- Вы можете отказаться анализировать малые выборки. Или анализировать их, но держать в голове, что могут быть проблемы.



Ждём вас на втором сезоне лектория
«Разрушители статистических мифов» в мае!

Анонс появится на сайте: bioinf.me/stat_myths
и в социальных сетях Института биоинформатики!



Институт биоинформатики в социальных сетях

Разрушители статистических мифов: bioinf.me/stat_myths

Чат по биостатистике и R: https://t.me/chat_biostat_R

По всем вопросам: biostat@bioinf.me

Сайт Института: bioinf.me

Институт в VK: vk.com/bioinf

Telegram-канал Института: t.me/bioinforussia

Чат про образование и карьеру: t.me/bioinf_career

YouTube-канал: www.youtube.com/bioinforussia



Бонус 1



Remark. Sometimes it is recommended to test the equality of variances of our samples before using the two-sample t -test; this can be done for example by using the test from Chapter 6.5 or so called Levene's test (not presented in these lecture notes). If the equality of variances is rejected, we use Welch test, otherwise we use two-sample t -test. However, we advise against using this kind of approach. It is so called *two-phase test*, where the result depends on three different test statistics that are not independent. It is not guaranteed that the significance level of this test is equal to the required level α . If we are not sure about the assumption of normality or equal variances, we should use the Welch test. Then we do not have to verify either one of the assumptions of the two-sample t -test.

Бонус 2

1.5. Implications for Data Analysis



The asymptotic results we have just summarized indicate that linear regression model with ordinary least squares estimation of regression parameters can be used to obtain asymptotically correct statistical inference even if the response is not normal and the error terms do not have equal variance. We only need to have enough observations available for analysis so that the asymptotic results provide a reasonable approximation of the true distribution of the parameter estimator and other quantities of interest.

In this aspect, linear regression is actually a robust nonparametric statistical procedure.

- (a) If the responses are *normal* and possess *equal variances* we can perform exact statistical inference based on Proposition 1.1 regardless of the size of the dataset (for any $n > p$).
- (b) If the responses are *not normal* but have *equal variances* we can perform asymptotic inference based on Proposition 1.2 for large enough number of observations.
- (c) If the responses are *not normal* and have *unequal variances* we can perform asymptotic inference based on Proposition 1.3 with sandwich variance estimator for large enough number of observations.

What number of observations is large enough to trust the asymptotic approaches (b) and (c) depends on the complexity of the linear model.

Furthermore, if the error variances are unequal but are known up to a proportionality constant, i.e., $\text{var } Y_i = \sigma^2 w_i$ with known w_i , weighted least squares estimation can be used instead of the sandwich.