Ограничения статистических методов. Непараметрические методы.

Основы биостатистики, осень 2022

Марина Варфоломеева

- Трансформации данных
- Непараметрические методы
- Альтернативы одновыборочному и парному t-тестам
    - Тест знаков и тест Вилкоксона
- Альтернативы двухвыборочному t-тесту
    - Тест U Манна-Уитни и тест Колмогорова-Смирнова
- Пермутационные тесты

Если нарушаются условия применимости метода

Если нарушаются условия применимости метода

Подход

- параметрический метод с другими условиями применимости

Если нарушаются условия применимости метода

Подход

- параметрический метод с другими условиями применимости

Пример использования

- обобщенная линейная модель, вместо простой линейной регрессии

Если нарушаются условия применимости метода

Подход

- параметрический метод с другими условиями применимости

- подходящая трансформация данных *

Пример использования

- обобщенная линейная модель, вместо простой линейной регрессии

- логарифмирование счетных данных для нормализации формы распределения

Если нарушаются условия применимости метода

Подход	Пример использования
параметрический метод с другими условиями применимости	обобщенная линейная модель, вместо простой линейной регрессии
подходящая трансформация данных *	логарифмирование счетных данных для нормализации формы распределения
непараметрические методы *	тест Вилкоксона вместо t-теста

Если нарушаются условия применимости метода

Подход	Пример использования
• параметрический метод с другими условиями применимости	• обобщенная линейная модель, вместо простой линейной регрессии
• подходящая трансформация данных *	• логарифмирование счетных данных для нормализации формы распределения
• непараметрические методы *	• тест Вилкоксона вместо t-теста
• пермутационные тесты *	• распределения любых статистик для H_0 генерируются из самих данных

Трансформация данных

Трансформация данных

Трансформация данных — математическая операция, которую применяют к значениям переменной (т.е. к ряду данных целиком, а не к единичным наблюдениям)

- для коррекции формы распределения
- для выравнивания дисперсий

Трансформация данных

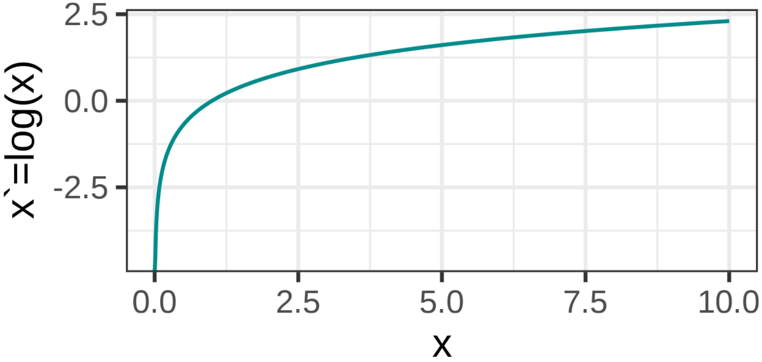
Трансформация данных — математическая операция, которую применяют к значениям переменной (т.е. к ряду данных целиком, а не к единичным наблюдениям)

- для коррекции формы распределения
- для выравнивания дисперсий

Самые распространенные трансформации

- логарифм
- квадратный корень
- арксинус-преобразование

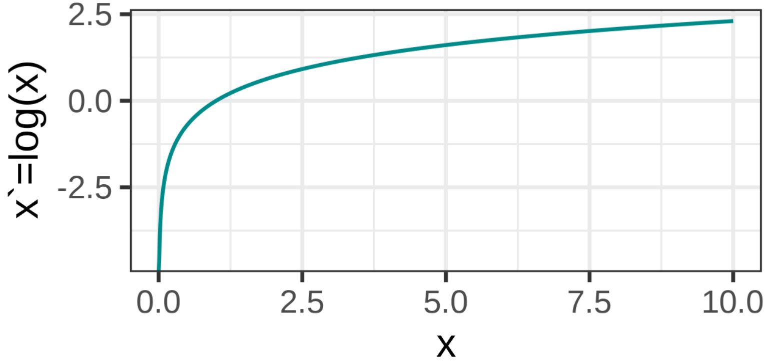
Логарифмирование



$$x' = \log(x)$$

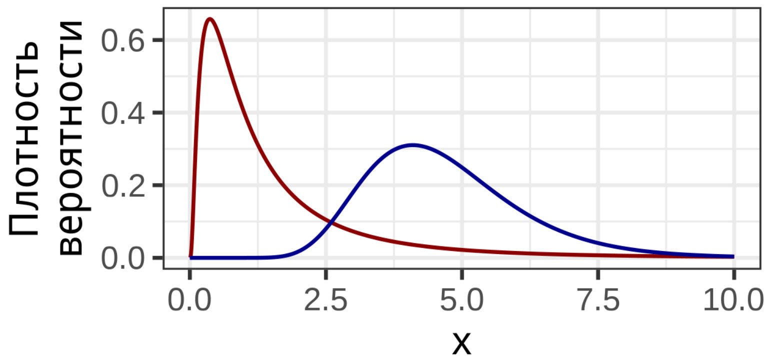
Если в данных есть нули, то прибавляют маленькую константу $x' = \log(x + 1)$.

Логарифмирование

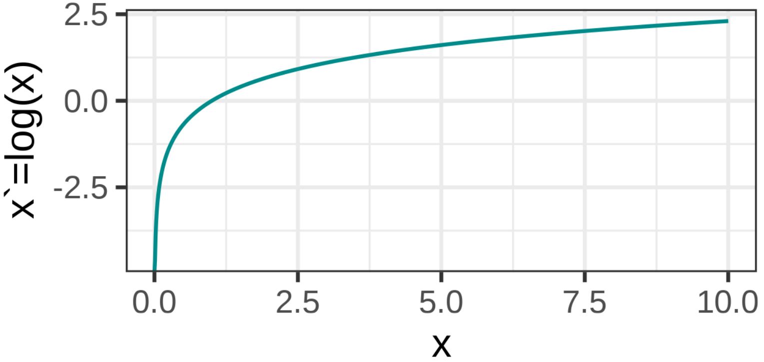


$$x' = \log(x)$$

Если в данных есть нули, то прибавляют маленькую константу $x' = \log(x + 1)$.

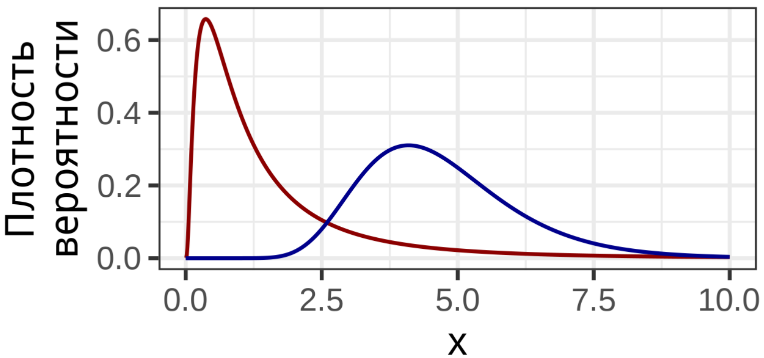


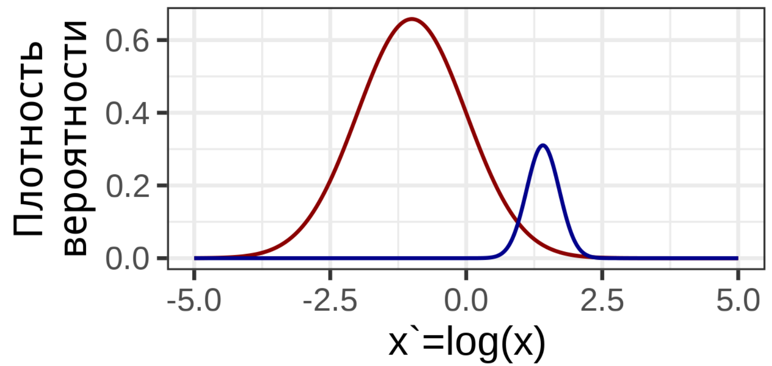
Логарифмирование



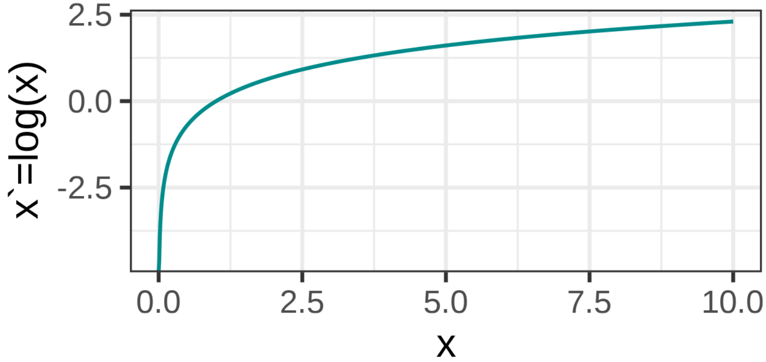
$$x' = \log(x)$$

Если в данных есть нули, то прибавляют маленькую константу $x' = \log(x + 1)$.



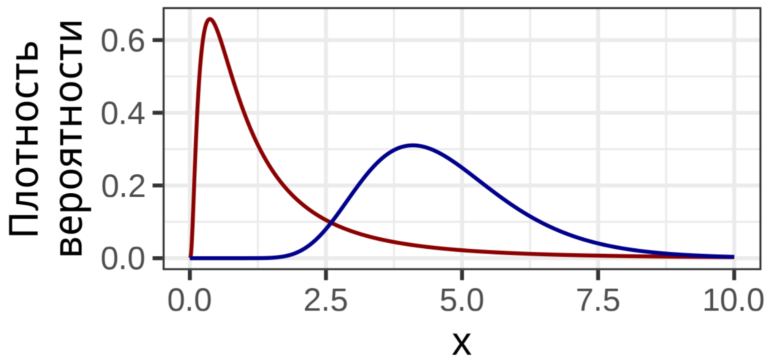


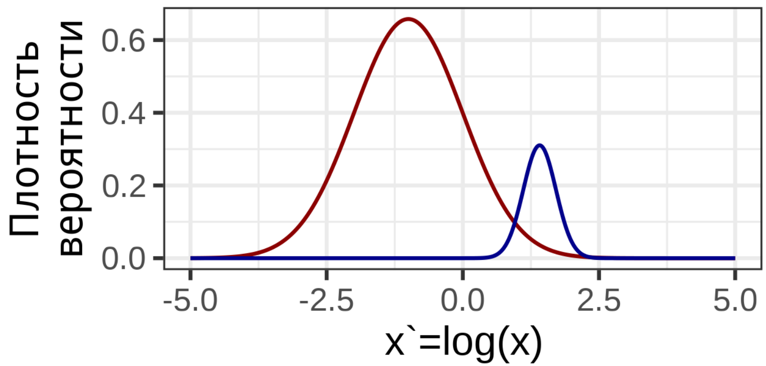
Логарифмирование



$$x' = \log(x)$$

Если в данных есть нули, то прибавляют маленькую константу $x' = \log(x+1)$.

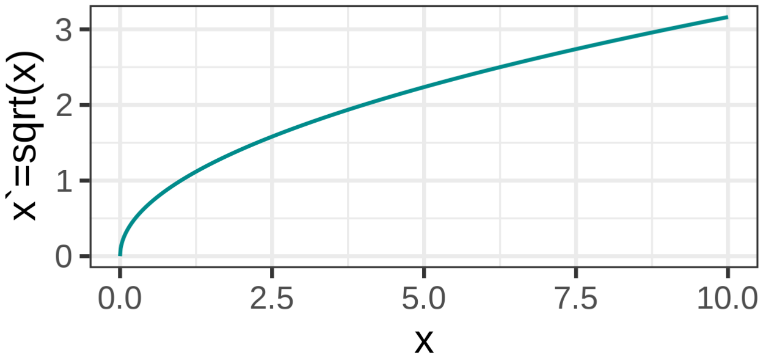




Подойдет, если

- распределение имеет длинный правый хвост
- измерения — это отношения, произведения, подсчеты численности
- в группах с большим средним большое стандартное отклонение
- диапазон данных несколько порядков

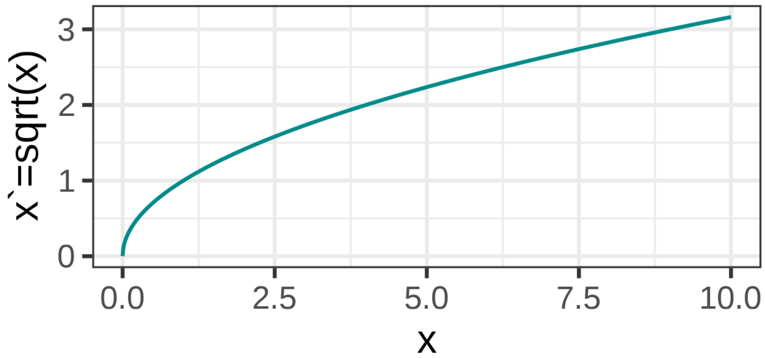
Извлечение корня

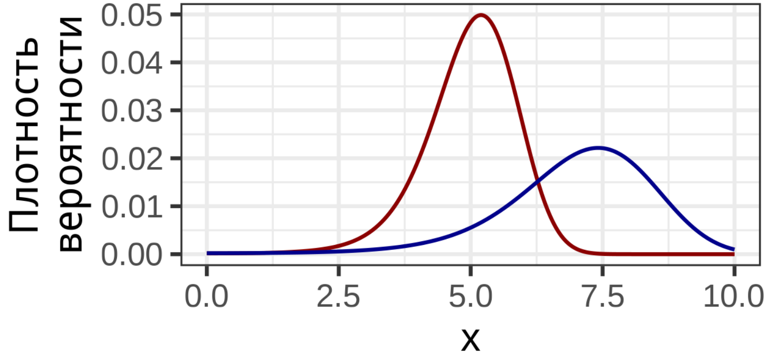


$$x' = \sqrt{x}$$

Иногда используют корни других степеней.

Извлечение корня

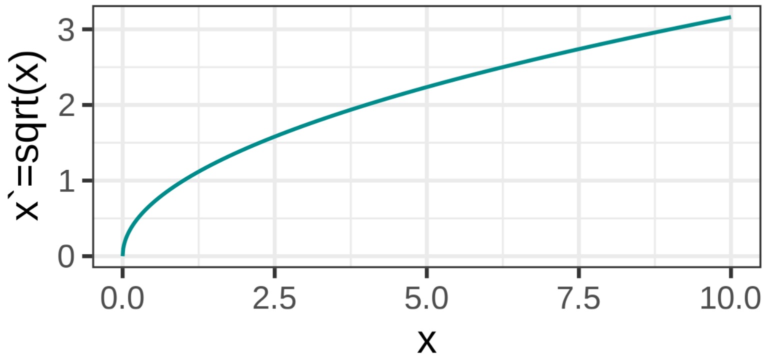




$$x' = \sqrt{x}$$

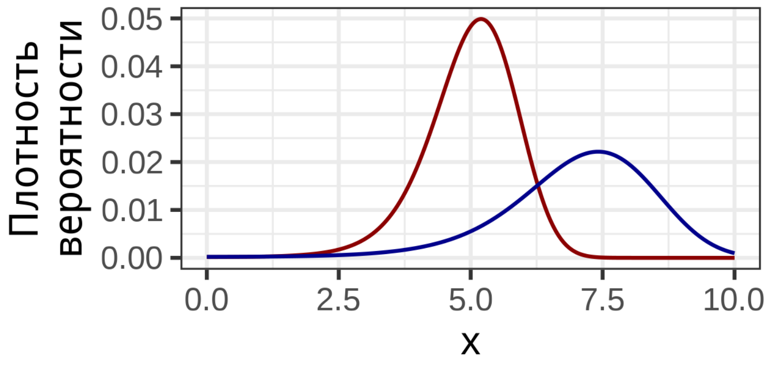
Иногда используют корни других степеней.

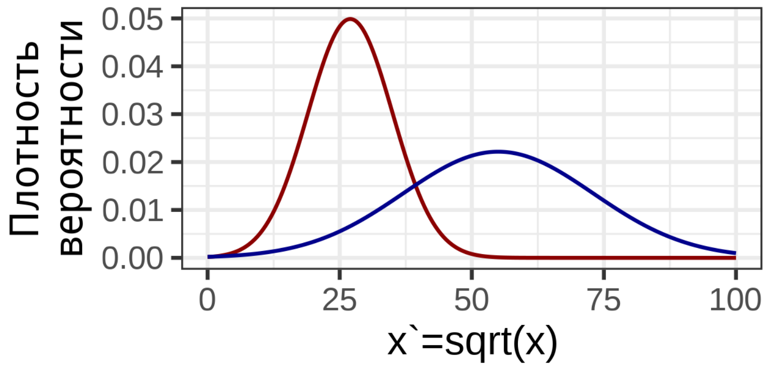
Извлечение корня



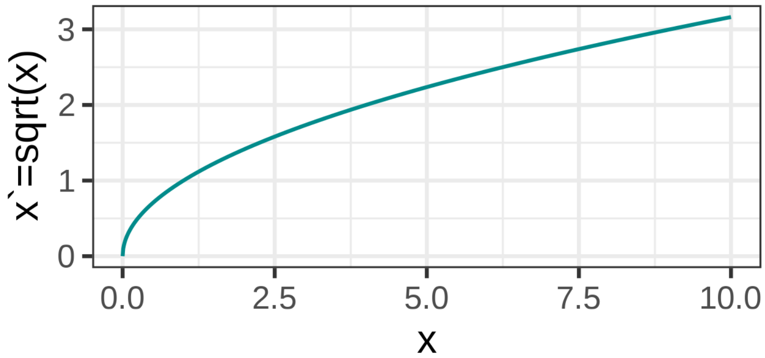
$$x' = \sqrt{x}$$

Иногда используют корни других степеней.



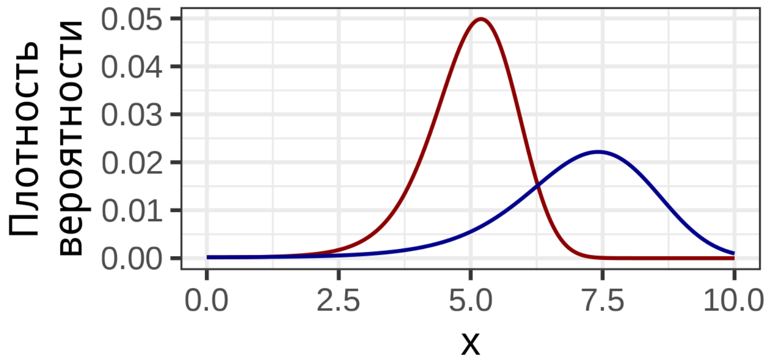


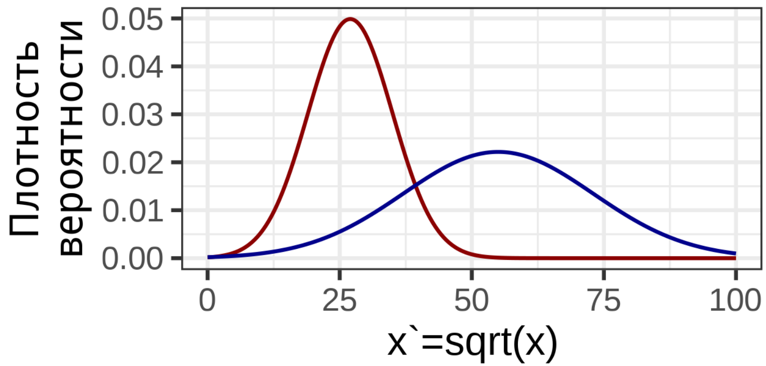
Извлечение корня



$$x' = \sqrt{x}$$

Иногда используют корни других степеней.

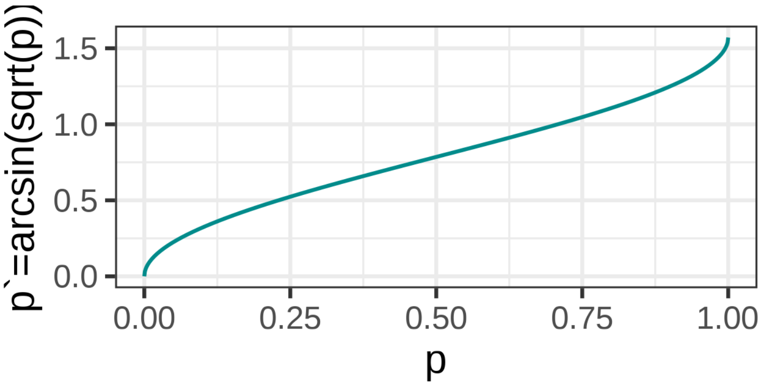




Подойдет, если

- распределение имеет длинный левый хвост
- данные - это подсчеты численности
- в группах с большим средним большое стандартное отклонение

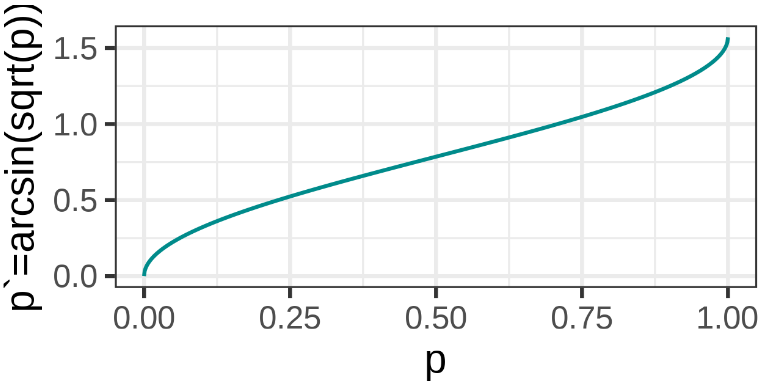
Арксинус-преобразование для долей



$$p' = \arcsin \sqrt{p}$$

Значения долей ограничены $0 \leq p \leq 1$.
Если много долей, близких к 0 или 1, трансформация “растянет” концы шкалы.

Арксинус-преобразование для долей



$$p' = \arcsin\sqrt{p}$$

Значения долей ограничены $0 \leq p \leq 1$.
Если много долей, близких к 0 или 1,
трансформация "растянет" концы шкалы.

Усовершенствованные варианты:

- $p' = \arcsin\sqrt{\dfrac{x+3/8}{n+3/4}}$ (Johnson, Kotz, 1969)
- $p' = 2\sqrt{n}\cdot\left(\arcsin\sqrt{\dfrac{x+3/8}{n+3/4}} - \arcsin\sqrt{p}\right)$ (Freeman, Tukey, 1950)

Основные трансформации данных

Степенные

Название	Трансформация	Обратная трансформация
степень -2	$x' = 1/x^2$	$x = \sqrt{1/x'}$
степень -1	$x' = 1/x$	$x = 1/x'$
степень -0.5	$x' = 1/\sqrt{x}$	$x = 1/x'^{\,2}$
степень 0.5	$x' = \sqrt{x}$	$x = x'^{\,2}$
логарифмирование	$x' = \log(x)$	$x = e^x$

Для долей

Название	Трансформация	Обратная трансформация
Арксинус-трансформация	$p' = \arcsin\sqrt{p}$	$p = \sin(p')^2$
Логит-преобразование	$p' = \log\left(\frac{p}{1-p}\right)$	$p = \frac{e^{p'}}{1+e^{p'}} = \frac{1}{1+e^{-p'}}$

Подбор подходящей трансформации

Нечестно специально выбирать трансформацию, приводящую к наиболее значимому результату теста.

Выбирайте трансформацию, которая лучше всего приближает к выполнению условий применимости.

Пример: морские заповедники

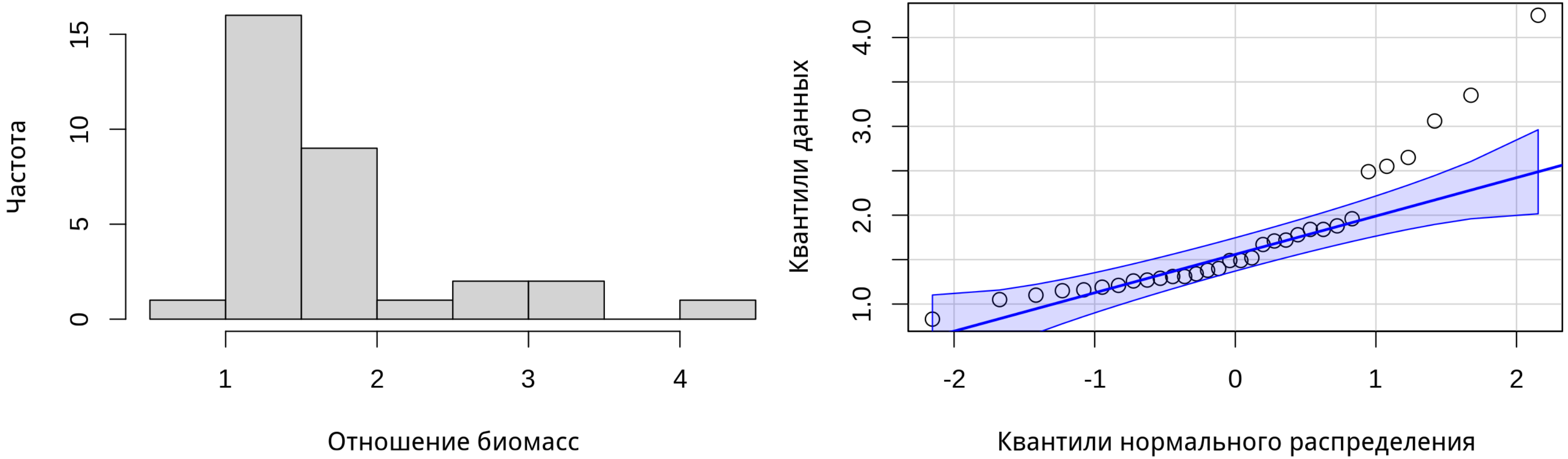
Эффективны ли морские заповедники для сохранения природы (Halpern, 2003)?

32 пары заповедник—контрольная точка (до заповедника или рядом с ним)

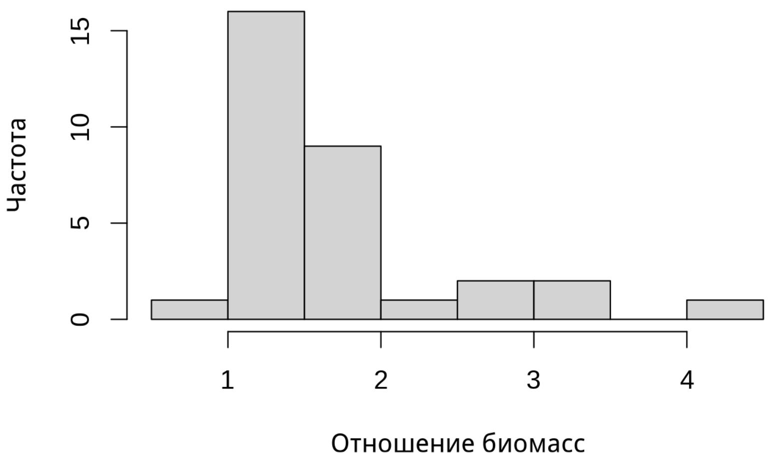
Показатель успеха защиты $I = \dfrac{B_{\text{заповедник}}}{B_{\text{контроль}}}$.

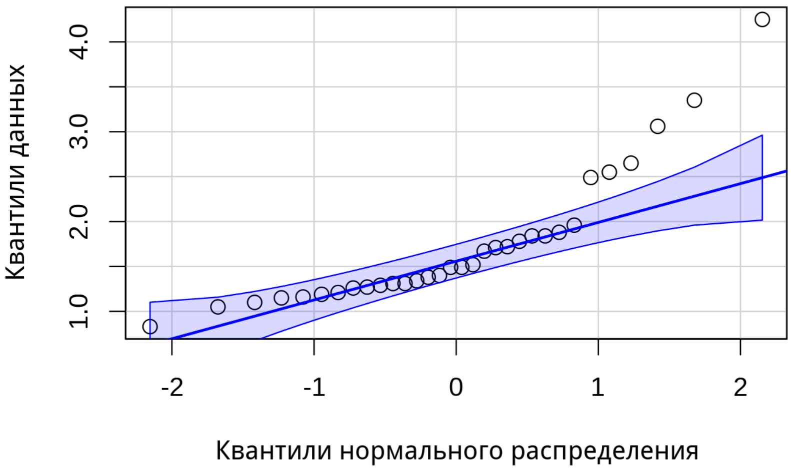
$H_0: \mu = 1$ — среднее отношение биомасс не зависит от статуса акватории $H_1: \mu \neq 1$ — среднее отношение биомасс различается в охраняемых и не охраняемых акваториях

Проверяем условия для одновыборочного t-теста



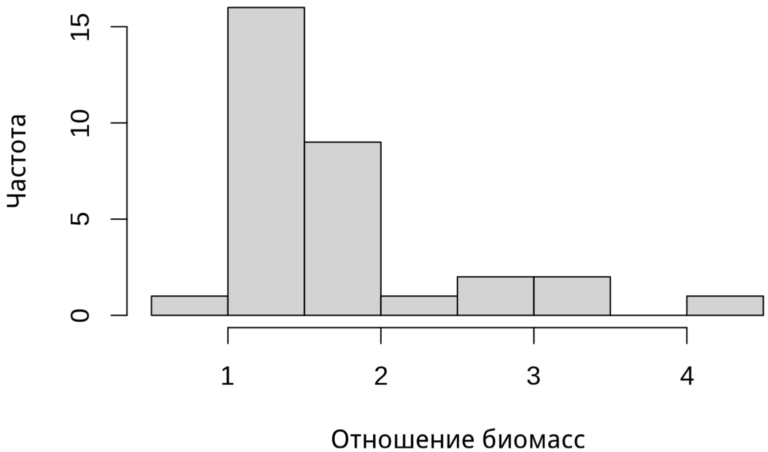
Проверяем условия для одновыборочного t-теста

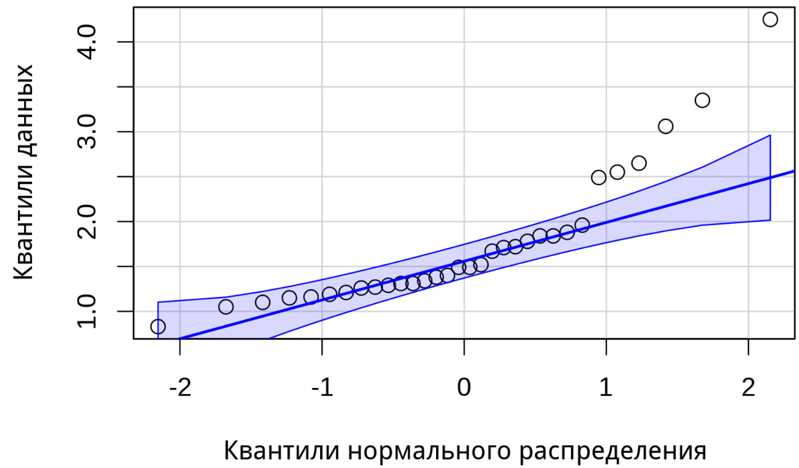




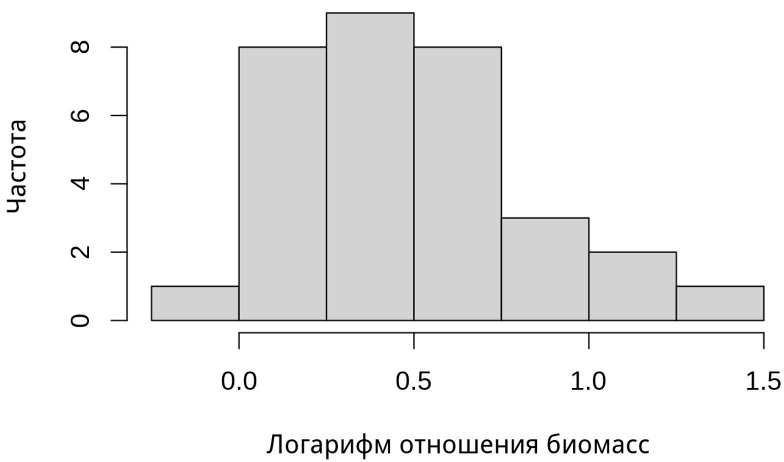
Асимметрия вправо. Попробуем логарифмировать данные.

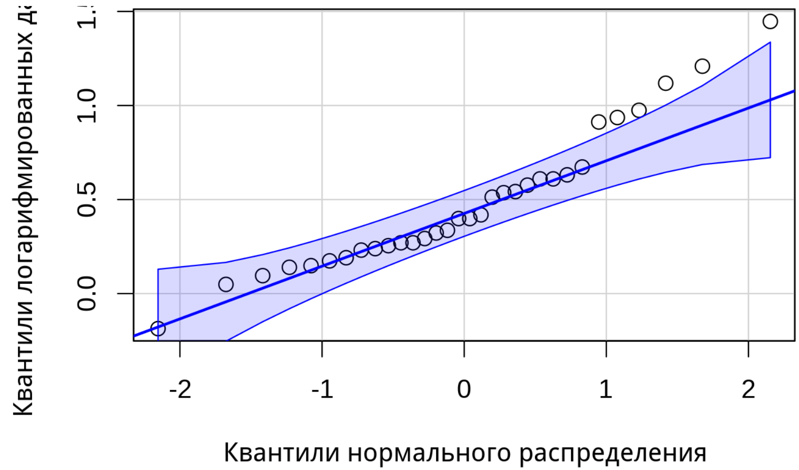
Проверяем условия для одновыборочного t-теста



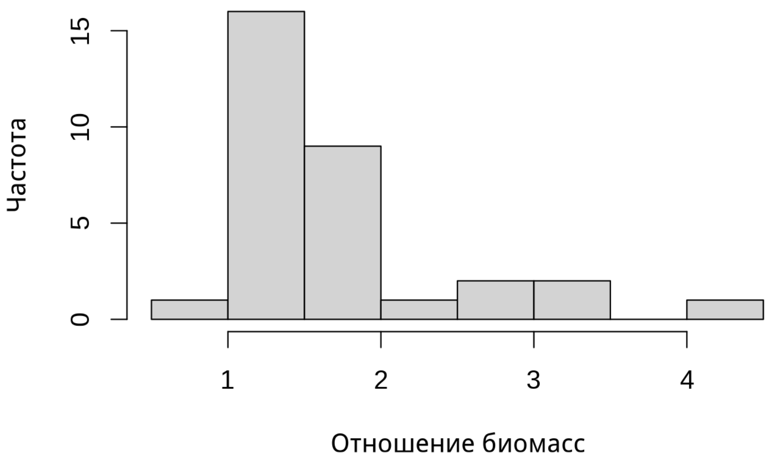


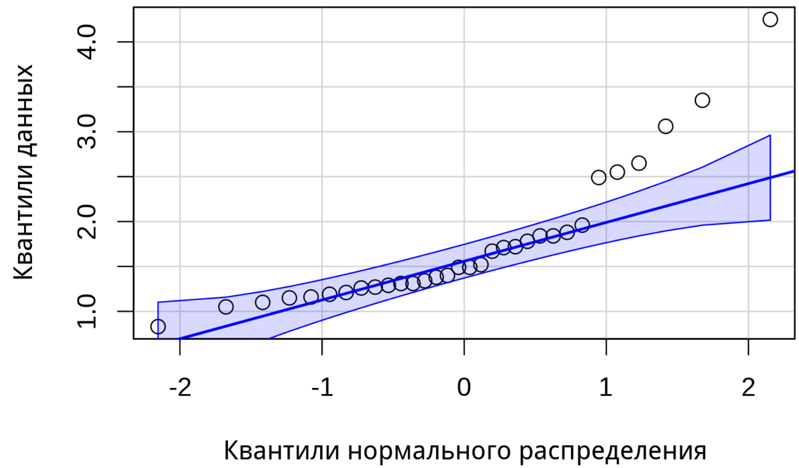
Асимметрия вправо. Попробуем логарифмировать данные.



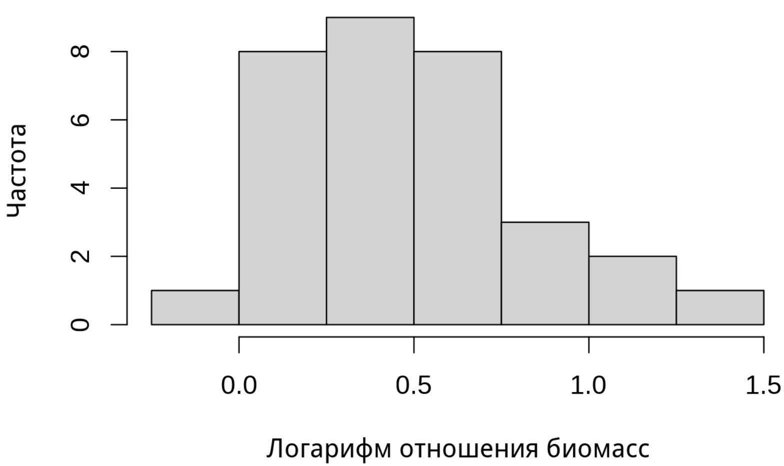


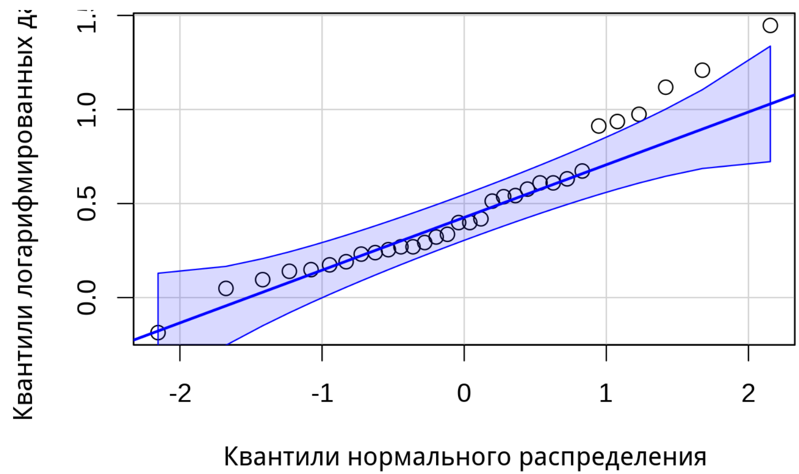
Проверяем условия для одновыборочного t-теста





Асимметрия вправо. Попробуем логарифмировать данные.





Неидеально, но явно лучше.

Одновыборочный t-тест по трансформированным данным

- $H_0 : \mu = 0$ — средний логарифм отношения биомасс не зависит от статуса
- $H_1 : \mu \neq 0$ — средний логарифм отношения биомасс различается в защищенных и не защищенных акваториях

Одновыборочный t-тест по трансформированным данным

- $H_0 : \mu = 0$ — средний логарифм отношения биомасс не зависит от статуса
- $H_1 : \mu \neq 0$ — средний логарифм отношения биомасс различается в защищенных и не защищенных акваториях

Формулировка изменилась, т.к. $log(1) = 0$

Одновыборочный t-тест по трансформированным данным

- $H_0 : \mu = 0$ — средний логарифм отношения биомасс не зависит от статуса
- $H_1 : \mu \neq 0$ — средний логарифм отношения биомасс различается в защищенных и не защищенных акваториях

Формулировка изменилась, т.к. $log(1) = 0$

В нашем примере средний логарифм отношения биомасс $\overline{log(I)} = \overline{log\Big(\frac{B_{\text{заповедник}}}{B_{\text{контроль}}}\Big)} = 0.479$.

Одновыборочный t-тест по трансформированным данным

- $H_0 : \mu = 0$ — средний логарифм отношения биомасс не зависит от статуса
- $H_1 : \mu \neq 0$ — средний логарифм отношения биомасс различается в защищенных и не защищенных акваториях

Формулировка изменилась, т.к. $log(1) = 0$

В нашем примере средний логарифм отношения биомасс $\overline{log(I)} = \overline{log\left(\frac{B_{\text{заповедник}}}{B_{\text{контроль}}}\right)} = 0.479$.

$s = 0.366, n = 32$

$t = \frac{0.479 - 0}{0.366/\sqrt{32}} = 7.40, df = 32 - 1 = 31$

$p = 2.49e - 08$

Т.е. средний логарифм биомассы статистически значимо выше в охраняемых акваториях.

Доверительный интервал по трансформированным данным

$$\overline{log(I)} - t_{\alpha,df} \cdot SE_{\overline{log(I)}} \leq \overline{log(I)} \leq \overline{log(I)} + t_{\alpha,df} \cdot SE_{\overline{log(I)}}$$

Доверительный интервал по трансформированным данным

$$\overline{log(I)} - t_{\alpha,df} \cdot SE_{\overline{log(I)}} \le \overline{log(I)} \le \overline{log(I)} + t_{\alpha,df} \cdot SE_{\overline{log(I)}}$$

В этом интервале лежит средний логарифм (в 95% провторных выборок):

$$0.479 - 2.04 \cdot \frac{0.366}{\sqrt{32}} \le \overline{log(I)} \le 0.479 + 2.04 \cdot \frac{0.366}{\sqrt{32}}$$

Доверительный интервал по трансформированным данным

$$\overline{log(I)} - t_{\alpha,df} \cdot SE_{\overline{log(I)}} \leq \overline{log(I)} \leq \overline{log(I)} + t_{\alpha,df} \cdot SE_{\overline{log(I)}}$$

В этом интервале лежит средний логарифм (в 95% провторных выборок):

$$0.479 - 2.04 \cdot \frac{0.366}{\sqrt{32}} \leq \overline{log(I)} \leq 0.479 + 2.04 \cdot \frac{0.366}{\sqrt{32}}$$

$$0.347 \leq \overline{log(I)} \leq 0.611$$

Доверительный интервал по трансформированным данным

$$\overline{log(I)} - t_{\alpha,df} \cdot SE_{\overline{log(I)}} \leq \overline{log(I)} \leq \overline{log(I)} + t_{\alpha,df} \cdot SE_{\overline{log(I)}}$$

В этом интервале лежит средний логарифм (в 95% провторных выборок):

$$0.479 - 2.04 \cdot \frac{0.366}{\sqrt{32}} \leq \overline{log(I)} \leq 0.479 + 2.04 \cdot \frac{0.366}{\sqrt{32}}$$

$$0.347 \leq \overline{log(I)} \leq 0.611$$

Само среднее (геометрическое - из-за логарифмирования)
лежит в другом интервале:

$$e^{0.347} \leq \text{среднее геометрическое } I \leq e^{0.611}$$

Доверительный интервал по трансформированным данным

$$\overline{log(I)} - t_{\alpha,df} \cdot SE_{\overline{log(I)}} \leq \overline{log(I)} \leq \overline{log(I)} + t_{\alpha,df} \cdot SE_{\overline{log(I)}}$$

В этом интервале лежит средний логарифм (в 95% провторных выборок):

$$0.479 - 2.04 \cdot \frac{0.366}{\sqrt{32}} \leq \overline{log(I)} \leq 0.479 + 2.04 \cdot \frac{0.366}{\sqrt{32}}$$

$$0.347 \leq \overline{log(I)} \leq 0.611$$

Само среднее (геометрическое - из-за логарифмирования)
лежит в другом интервале:

$$e^{0.347} \leq \text{среднее геометрическое } I \leq e^{0.611}$$

$$1.415 \leq \text{среднее геометрическое } I \leq 1.843$$

Доверительный интервал по трансформированным данным

$$\overline{log(I)} - t_{\alpha,df} \cdot SE_{\overline{log(I)}} \leq \overline{log(I)} \leq \overline{log(I)} + t_{\alpha,df} \cdot SE_{\overline{log(I)}}$$

В этом интервале лежит средний логарифм (в 95% провторных выборок):

$$0.479 - 2.04 \cdot \frac{0.366}{\sqrt{32}} \leq \overline{log(I)} \leq 0.479 + 2.04 \cdot \frac{0.366}{\sqrt{32}}$$

$$0.347 \leq \overline{log(I)} \leq 0.611$$

Само среднее (геометрическое - из-за логарифмирования)
лежит в другом интервале:

$$e^{0.347} \leq \text{среднее геометрическое } I \leq e^{0.611}$$

$$1.415 \leq \text{среднее геометрическое } I \leq 1.843$$

Т.е. в заповедниках биомасса в среднем выше, чем в неохраняемых акваториях в 1.415 — 1.843 раз.

Непараметрические методы

Непараметрические методы

- не делают никаких предположений о значениях **параметров** статистических распределений
- имеют менее жесткие условия применимости (например, не требуют нормальности)
- часто основаны на использовании рангов

Ранги

Допустим, вы поймали несколько ящериц и измерили их длину (см).

10, 4, 7, 8, 7, 6, 11, 6, 13, 6, 10

Ранги

Допустим, вы поймали несколько ящериц и измерили их длину (см).

10, 4, 7, 8, 7, 6, 11, 6, 13, 6, 10

Данные можно рассортировать.

4, 6, 6, 6, 7, 7, 8, 10, 10, 11, 13

Ранги

Допустим, вы поймали несколько ящериц и измерили их длину (см).

10, 4, 7, 8, 7, 6, 11, 6, 13, 6, 10

Данные можно рассортировать.

4, 6, 6, 6, 7, 7, 8, 10, 10, 11, 13

Ранг наблюдения — это его порядковый номер в отсортированном ряду. Одинаковые наблюдения ("связанные ранги", tied ranks) получат один и тот же ранг — чаще это среднее их номеров.

Ранги

Допустим, вы поймали несколько ящериц и измерили их длину (см).

10, 4, 7, 8, 7, 6, 11, 6, 13, 6, 10

Ранг наблюдения — это его порядковый номер в отсортированном ряду. Одинаковые наблюдения (“связанные ранги”, tied ranks) получат один и тот же ранг — чаще это среднее их номеров.

Данные можно рассортировать.

4, 6, 6, 6, 7, 7, 8, 10, 10, 11, 13

1, 3, 3, 3, 5.5, 5.5, 7, 8.5, 8.5, 10, 11

Ранги

Допустим, вы поймали несколько ящериц и измерили их длину (см).

10, 4, 7, 8, 7, 6, 11, 6, 13, 6, 10

Данные можно рассортировать.

4, 6, 6, 6, 7, 7, 8, 10, 10, 11, 13

Ранг наблюдения — это его порядковый номер в отсортированном ряду. Одинаковые наблюдения (“связанные ранги”, tied ranks) получат один и тот же ранг — чаще это среднее их номеров.

1, 3, 3, 3, 5.5, 5.5, 7, 8.5, 8.5, 10, 11

Исходные значения можно заменить на их ранги и работать уже с ними.

8.5, 1, 5.5, 7, 5.5, 3, 10, 3, 11, 3, 8.5

Ошибки I и II рода

Вероятность ошибки I рода

У параметрических тестов соответствует α только если выполнены условия применимости. Если нет, то увеличивается.

У непараметрических тестов — соответствует α, если выполнены более мягкие условия применимости.

Вероятность ошибки II рода

У непараметрических тестов увеличивается вероятность ошибки II рода β, из-за потери информации при переходе к рангам.

Проверка гипотез для одной или парных выборок

при помощи непараметрических тестов

Тест знаков

Sign test

Альтернатива одновыборочному или парному t-тестам, когда данные не подчиняются нормальному распределению.

Проверяет равенство медианы переменной x конкретному значению m_0.

- $H_0 : m = m_0$
- $H_A : m \neq m_0$

Тест знаков

Sign test

Альтернатива одновыборочному или парному t-тестам, когда данные не подчиняются нормальному распределению.

Проверяет равенство медианы переменной x конкретному значению m_0.

- $H_0 : m = m_0$
- $H_A : m \neq m_0$

Обозначим $x > m_0$ как “+”, а $x < m_0$ как “—“, остальные не будем учитывать вообще.

Тест знаков

Sign test

Альтернатива одновыборочному или парному t-тестам, когда данные не подчиняются нормальному распределению.

Проверяет равенство медианы переменной x конкретному значению m_0.

- $H_0 : m = m_0$
- $H_A : m \neq m_0$

Обозначим $x > m_0$ как “+”, а $x < m_0$ как “—“, остальные не будем учитывать вообще.

При H_0 можно ожидать что, доля “+” и “—“ равна 0.5.

Тест знаков

Sign test

Альтернатива одновыборочному или парному t-тестам, когда данные не подчиняются нормальному распределению.

Проверяет равенство медианы переменной x конкретному значению m_0.

- $H_0 : m = m_0$
- $H_A : m \neq m_0$

Обозначим $x > m_0$ как “+”, а $x < m_0$ как “—“, остальные не будем учитывать вообще.

При H_0 можно ожидать что, доля “+” и “—“ равна 0.5.

Для вычисления p используется биномиальное распределение.

Пример: половой конфликт и происхождение видов

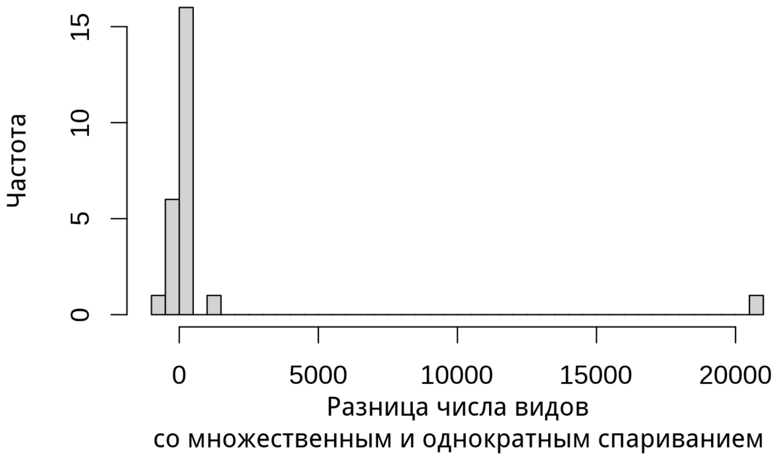
Половой конфликт более выражен у видов, чьи самки спариваются больше одного раза. Значит ли это, что у них быстрее видообразование (Arnqvist et al., 2000)?

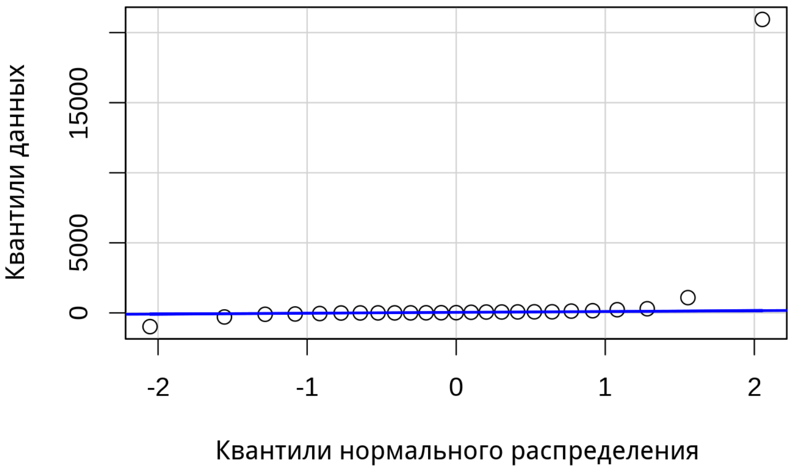
25 пар групп насекомых со множественным и однократным спариванием.

В каких из этих групп таксонов больше родственных видов?

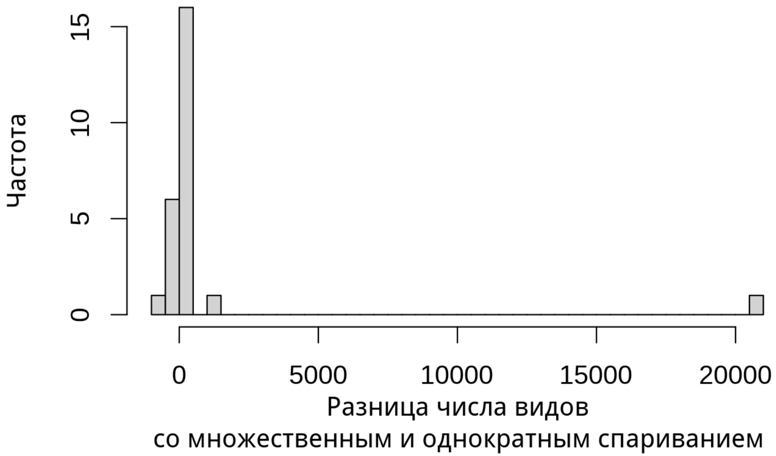
	Спаривание (число таксонов)		
Пара таксонов	Множественное	Однократное	Разница
A	53	10	43
B	73	120	-47
C	228	74	154
D	353	289	64
E	157	30	127
F	300	4	296
G	34	18	16
H	3400	3500	-100

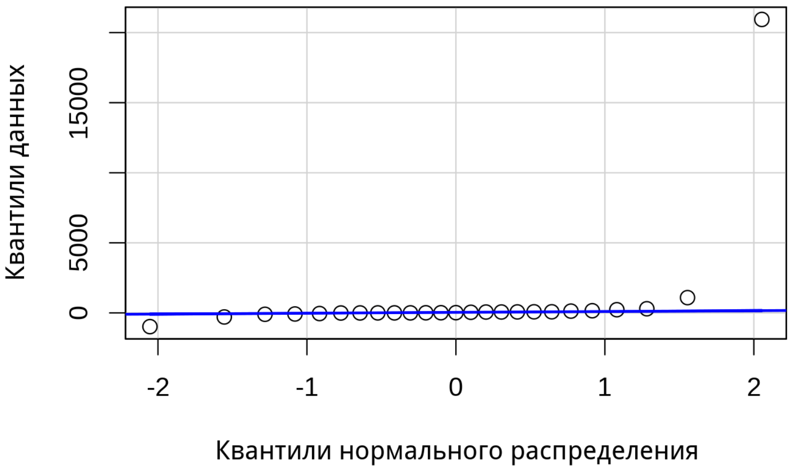
Проверяем условия для одновыборочного t-теста





Проверяем условия для одновыборочного t-теста





Есть наблюдение-выброс — таксон, в котором видов со множественным спариванием на 20400 больше, чем с однократным.

Никакая трансформация не поможет.

Гипотезы для теста знаков

- $H_0 : m = m_0$ — медианная разница числа видов между группами со множественным и однократным спариванием равна 0.
- $H_A : m \neq m_0$ — медианная разница числа видов между этими группами не равна 0.

Гипотезы для теста знаков

- $H_0 : m = m_0$ — медианная разница числа видов между группами со множественным и однократным спариванием равна 0.
- $H_A : m \neq m_0$ — медианная разница числа видов между этими группами не равна 0.

При H_0 разница числа видов > 0 у половины наблюдений, а у другой половины < 0.

Гипотезы для теста знаков

- $H_0 : m = m_0$ — медианная разница числа видов между группами со множественным и однократным спариванием равна 0.
- $H_A : m \neq m_0$ — медианная разница числа видов между этими группами не равна 0.

При H_0 разница числа видов > 0 у половины наблюдений, а у другой половины < 0.

Тест знаков — это биномиальный тест.

Расставляем знаки

	Спаривание (число таксонов)			
Пара таксонов	Множественное	Однократное	Разница	Знак
A	53	10	43	+
B	73	120	-47	-
C	228	74	154	+
D	353	289	64	+
E	157	30	127	+
F	300	4	296	+
G	34	18	16	+
H	3400	3500	-100	-
I	20	1000	-980	-
J	196	486	-290	-
K	1750	660	1090	+
L	55	63	-8	-

Расставляем знаки

	Спаривание (число таксонов)			
Пара таксонов	Множественное	Однократное	Разница	Знак
A	53	10	43	+
B	73	120	-47	-
C	228	74	154	+
D	353	289	64	+
E	157	30	127	+
F	300	4	296	+
G	34	18	16	+
H	3400	3500	-100	-
I	20	1000	-980	-
J	196	486	-290	-
K	1750	660	1090	+
L	55	63	-8	-

Подсчитаем количество

-	+
7	18

Тест знаков

Гипотезы можно переформулировать как для биномиального теста

$H_0 : \pi = 0.5$ — доля "-" в генеральной совокупности равна 0.5

$H_A : \pi \neq 0.5$ — доля "-" не равна 0.5

-	+
7	18

Тест знаков

Гипотезы можно переформулировать как для биномиального теста

$H_0 : \pi = 0.5$ — доля “-“ в генеральной совокупности равна 0.5

$H_A : \pi \neq 0.5$ — доля “-“ не равна 0.5

-	+
7	18

Биномиальное распределение:
$P(X\text{'успех'}) = \frac{n!}{X!(n-X)!}\pi^X(1-\pi)^{n-X}$

Тест знаков

Гипотезы можно переформулировать как для биномиального теста

-	+
7	18

$H_0 : \pi = 0.5$ — доля “-“ в генеральной совокупности равна 0.5

$H_A : \pi \neq 0.5$ — доля “-“ не равна 0.5

Биномиальное распределение:
$P(X\text{'успех'}) = \frac{n!}{X!(n-X)!}\pi^X(1-\pi)^{n-X}$

Какова вероятность получить ≤ 7 “-“ из 25 когда $p_{\text{"-"}} = 0.5$?

Тест знаков

Гипотезы можно переформулировать как для биномиального теста

$H_0 : \pi = 0.5$ — доля "-" в генеральной совокупности равна 0.5

$H_A : \pi \neq 0.5$ — доля "-" не равна 0.5

-	+
7	18

Биномиальное распределение:
$P(X\text{'успех'}) = \frac{n!}{X!(n-X)!}\pi^X(1-\pi)^{n-X}$

Какова вероятность получить ≤ 7 "-" из 25 когда $p_{\text{"-"}} = 0.5$?

Находим и суммируем вероятности для 0-7:

$$P(X \leq 7) = \sum_{i=0}^{7} \frac{25!}{i!(25-i)!} 0.5^i (1-0.5)^{25-i} = 0.02164$$

Тест знаков

Гипотезы можно переформулировать как для биномиального теста

$H_0 : \pi = 0.5$ — доля “-“ в генеральной совокупности равна 0.5

$H_A : \pi \neq 0.5$ — доля “-“ не равна 0.5

-	+
7	18

Биномиальное распределение:
$P(X\text{'успех'}) = \frac{n!}{X!(n-X)!}\pi^X(1-\pi)^{n-X}$

Какова вероятность получить ≤ 7 “-“ из 25 когда $p_{\text{"-"}} = 0.5$?

Находим и суммируем вероятности для 0-7:

$$P(X \leq 7) = \sum_{i=0}^{7} \frac{25!}{i!(25-i)!} 0.5^i (1-0.5)^{25-i} = 0.02164$$

Двусторонняя альтернативная гипотеза, поэтому $p = 2 \cdot 0.02164 = 0.043$

Тест знаковых рангов Вилкоксона

Wilcoxon signed-rank test

“Усовершенствованная” (плохо) версия теста знаков.

Сохраняет информацию о величине отличия от m_0.

Тест знаковых рангов Вилкоксона

Wilcoxon signed-rank test

“Усовершенствованная” (плохо) версия теста знаков.

Сохраняет информацию о величине отличия от m_0.

Осторожно!

Условие применимости: распределение измерений **симметрично** относительно медианы.

Почти невыполнимо, т.к. в большинстве случаев проблемы с нормальностью распределения именно из-за асимметрии.

Тест знаковых рангов Вилкоксона

Wilcoxon signed-rank test

“Усовершенствованная” (плохо) версия теста знаков.

Сохраняет информацию о величине отличия от m_0.

Осторожно!

Условие применимости: распределение измерений **симметрично** относительно медианы.

Почти невыполнимо, т.к. в большинстве случаев проблемы с нормальностью распределения именно из-за асимметрии.

Лучше использовать тест знаков.

Сравнение двух групп

при помощи непараметрических тестов

Тест U Манна-Уитни

The Mann-Whitney U-test

Альтернатива двухвыборочному t-тесту.

По ранжированным данным сравнивает среднее значение рангов в двух выборках.

p-значение зависит от степени несоответствия средних рангов.

Тест U Манна-Уитни

The Mann-Whitney U-test

Альтернатива двухвыборочному t-тесту.

По ранжированным данным сравнивает среднее значение рангов в двух выборках.

p-значение зависит от степени несоответствия средних рангов.

Особенности:

- Регистрирует в основном различия медиан
- Есть адаптации для связанных рангов

Тест U Манна-Уитни

The Mann-Whitney U-test

Альтернатива двухвыборочному t-тесту.

По ранжированным данным сравнивает среднее значение рангов в двух выборках.

p-значение зависит от степени несоответствия средних рангов.

Особенности:

- Регистрирует в основном различия медиан
- Есть адаптации для связанных рангов

= тест суммы рангов Вилкоксона (Wilcoxon rank sum test).

Не путайте с тестом знаковых рангов Вилкоксона.

Пример: канибализм у сверчков

У сверчков Cyphoderris strepitans во время спаривания самец предлагает самке съесть свои мясистые крылья.

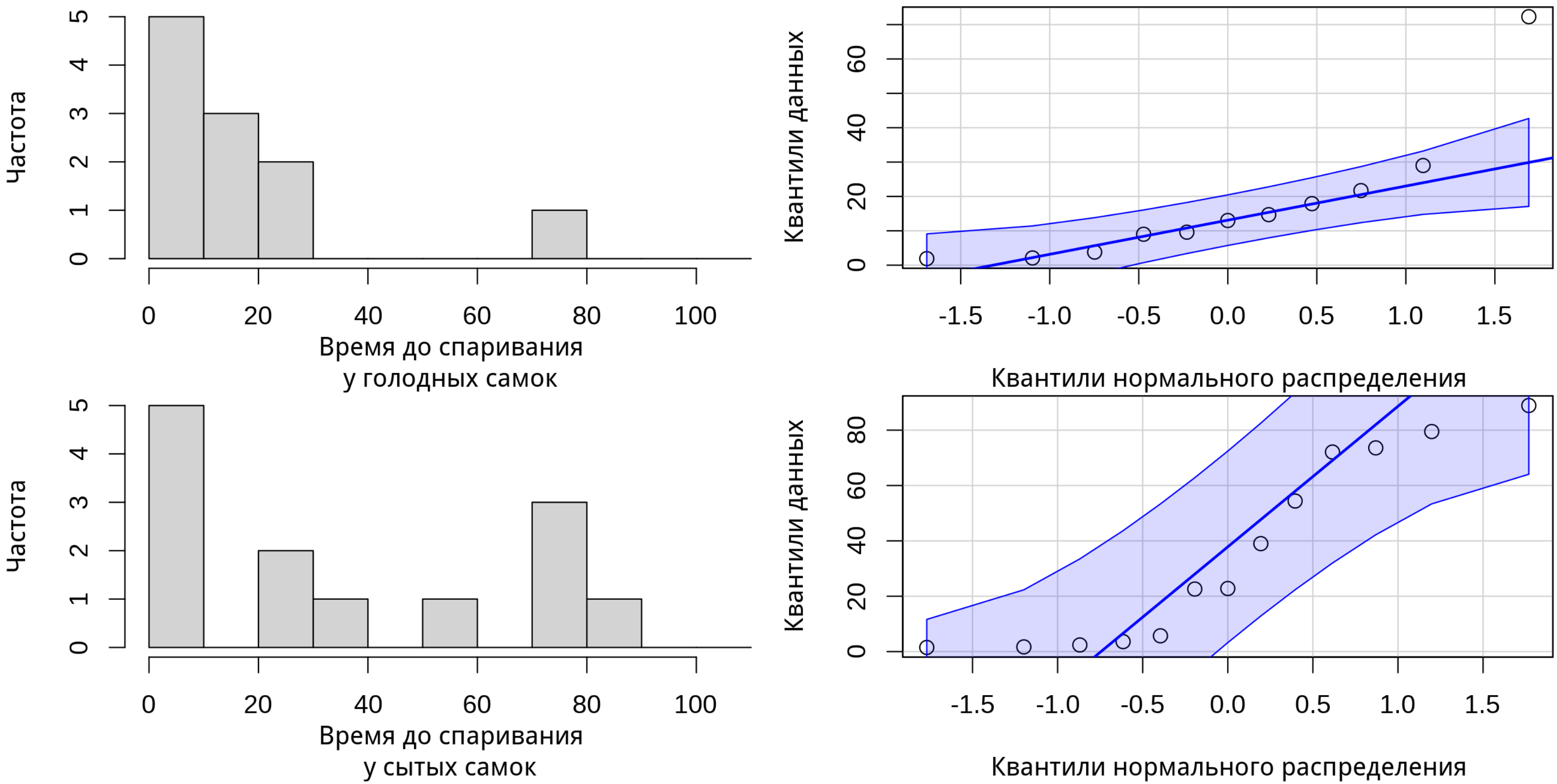
Зависит ли вероятность спаривания от того, голодные ли самки (Johnson, 1999)?

Свежих самцов предлагали самкам из двух групп:

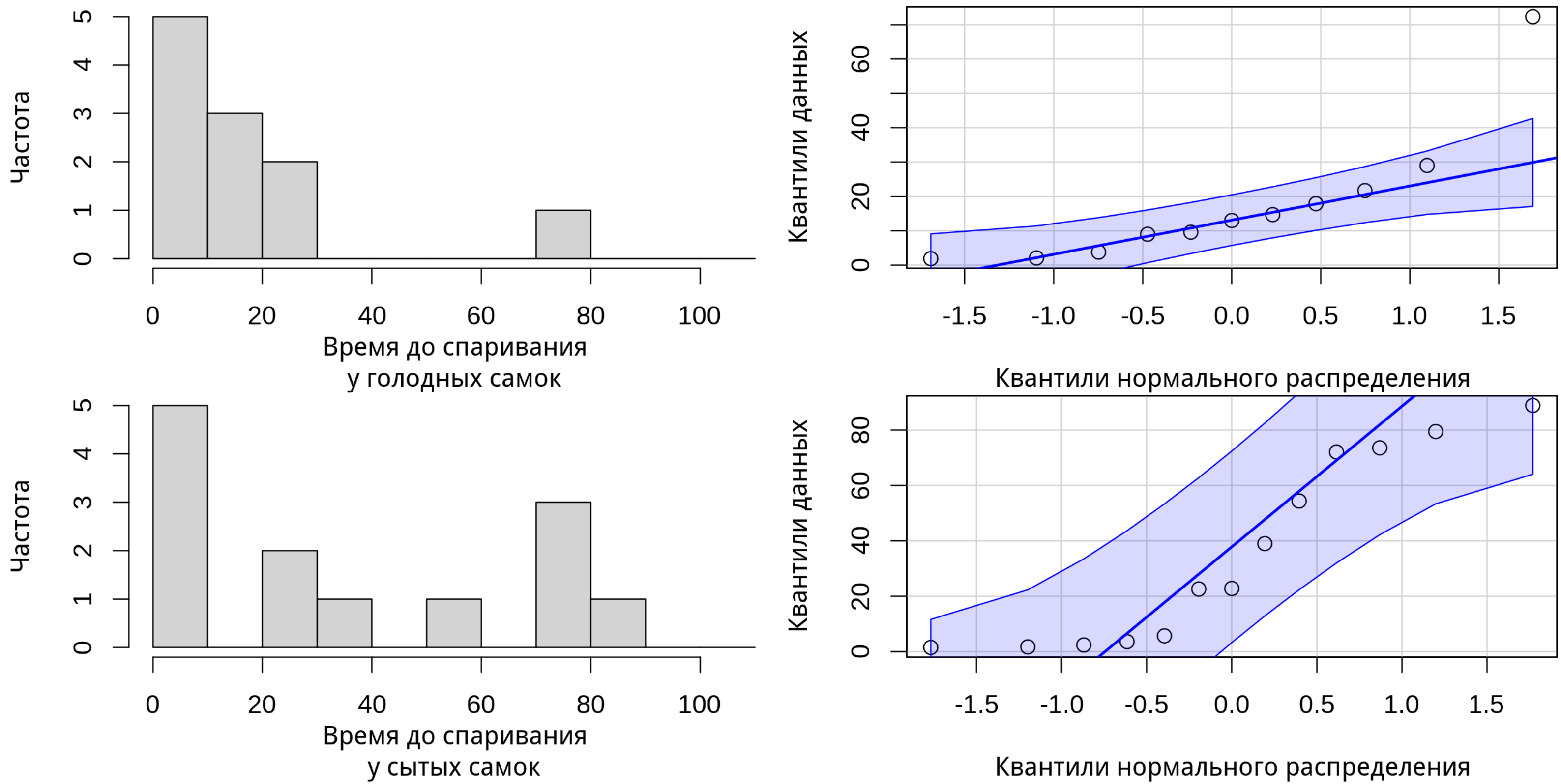
- 11 голодных
- 13 сытых

Регистрировали время до начала спаривания.

Проверяем условия для двухвыборочного t-теста



Проверяем условия для двухвыборочного t-теста

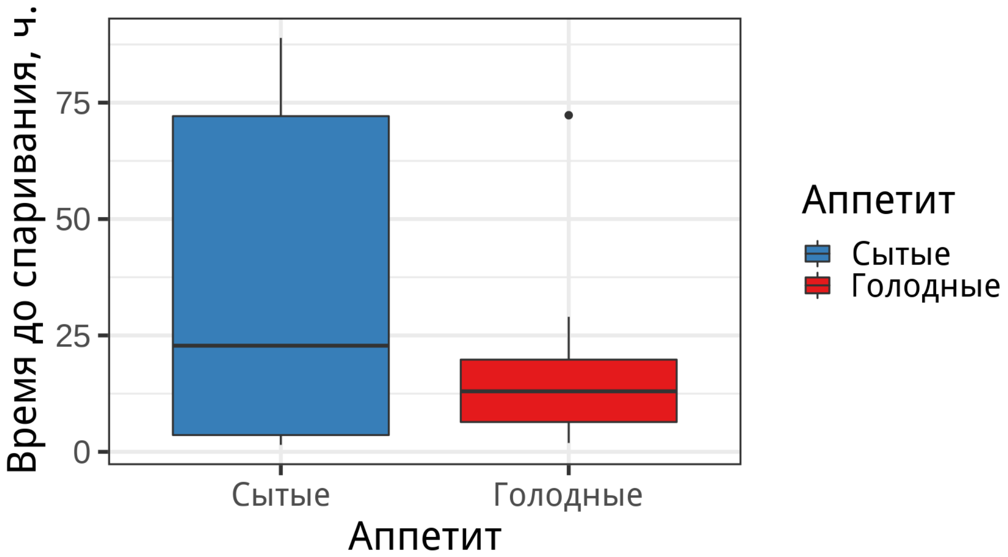


Асимметрия вправо. В одной группе есть выброс. В другой полимодальное распределение. Малый объем выборки.

Логарифмирование не поможет.

Медианное время до спаривания

- у голодных самок 13.0 ч.
- у сытых самок 22.8 ч.



Гипотезы для теста Манна-Уитни

- $H_0 : m_1 - m_2 = 0$ — медианное время до спаривания одинаковое у сытых и голодных самок.
- $H_A : m_1 - m_2 \neq 0$ — медианное время до спаривания разное у сытых и голодных самок.

Ранжируем данные

Аппетит	Время до спаривания	Ранг
Сытые	1.5	1
Сытые	1.7	2
Голодные	1.9	3
Голодные	2.1	4
Сытые	2.4	5
Сытые	3.6	6
Голодные	3.8	7
Сытые	5.7	8
Голодные	9.0	9
Голодные	9.6	10
Голодные	13.0	11
Голодные	14.7	12
Голодные	17.9	13

Сумма рангов в группе 1 (голодные): $R_1 = 121$

Ранжируем данные

Аппетит	Время до спаривания	Ранг
Сытые	1.5	1
Сытые	1.7	2
Голодные	1.9	3
Голодные	2.1	4
Сытые	2.4	5
Сытые	3.6	6
Голодные	3.8	7
Сытые	5.7	8
Голодные	9.0	9
Голодные	9.6	10
Голодные	13.0	11
Голодные	14.7	12
Голодные	17.9	13

Сумма рангов в группе 1 (голодные): $R_1 = 121$

Сколько раз у наблюдений из группы 1 меньший ранг, чем в группе 2 (во всех возможных парах):

$$U_1 = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - R_1 =$$

$$= 11 \cdot 13 + \frac{11(11+1)}{2} - 121 = 88$$

Ранжируем данные

Аппетит	Время до спаривания	Ранг
Сытые	1.5	1
Сытые	1.7	2
Голодные	1.9	3
Голодные	2.1	4
Сытые	2.4	5
Сытые	3.6	6
Голодные	3.8	7
Сытые	5.7	8
Голодные	9.0	9
Голодные	9.6	10
Голодные	13.0	11
Голодные	14.7	12
Голодные	17.9	13

Сумма рангов в группе 1 (голодные): $R_1 = 121$

Сколько раз у наблюдений из группы 1 меньший ранг, чем в группе 2 (во всех возможных парах):

$$U_1 = n_1 \cdot n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 =$$

$$= 11 \cdot 13 + \frac{11(11 + 1)}{2} - 121 = 88$$

Аналогично для группы 2:
$U_2 = n_1 \cdot n_2 - U_1 = 11 \cdot 13 - 88 = 55$

Ранжируем данные

Аппетит	Время до спаривания	Ранг
Сытые	1.5	1
Сытые	1.7	2
Голодные	1.9	3
Голодные	2.1	4
Сытые	2.4	5
Сытые	3.6	6
Голодные	3.8	7
Сытые	5.7	8
Голодные	9.0	9
Голодные	9.6	10
Голодные	13.0	11
Голодные	14.7	12
Голодные	17.9	13

Сумма рангов в группе 1 (голодные): $R_1 = 121$

Сколько раз у наблюдений из группы 1 меньший ранг, чем в группе 2 (во всех возможных парах):

$$U_1 = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - R_1 =$$

$$= 11 \cdot 13 + \frac{11(11+1)}{2} - 121 = 88$$

Аналогично для группы 2:
$U_2 = n_1 \cdot n_2 - U_1 = 11 \cdot 13 - 88 = 55$

Максимальное из них — это U-статистика:
$U = U_1 = 88$

Ранжируем данные

Аппетит	Время до спаривания	Ранг
Сытые	1.5	1
Сытые	1.7	2
Голодные	1.9	3
Голодные	2.1	4
Сытые	2.4	5
Сытые	3.6	6
Голодные	3.8	7
Сытые	5.7	8
Голодные	9.0	9
Голодные	9.6	10
Голодные	13.0	11
Голодные	14.7	12
Голодные	17.9	13

Сумма рангов в группе 1 (голодные): $R_1 = 121$

Сколько раз у наблюдений из группы 1 меньший ранг, чем в группе 2 (во всех возможных парах):

$$U_1 = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - R_1 =$$

$$= 11 \cdot 13 + \frac{11(11+1)}{2} - 121 = 88$$

Аналогично для группы 2:
$U_2 = n_1 \cdot n_2 - U_1 = 11 \cdot 13 - 88 = 55$

Максимальное из них — это U-статистика:
$U = U_1 = 88$

P-значение определяем по U-распределению.

Критическое значение $U_{0.05, df=2, n_1=11, n_2=13} = 106$, поэтому $p > 0.05$, сохраняем H_0.

Тест Колмогорова-Смирнова

The Kolmogorov-Smirnov test

Сравнивает кумулятивное распределение в двух выборках (goodness-of-fit test).

p-значение зависит от степени несоответствия кумулят.

Тест Колмогорова-Смирнова

The Kolmogorov-Smirnov test

Сравнивает кумулятивное распределение в двух выборках (goodness-of-fit test).

p-значение зависит от степени несоответствия кумулят.

Недостатки:

- Регистрирует любые различия формы распределений. Т.е. не умеет отличать положение медианы от различий размаха и асимметрии.
- Плохо работает со связанными рангами.

Тест Колмогорова-Смирнова

The Kolmogorov-Smirnov test

Сравнивает кумулятивное распределение в двух выборках (goodness-of-fit test).

p-значение зависит от степени несоответствия кумулят.

Недостатки:

- Регистрирует любые различия формы распределений. Т.е. не умеет отличать положение медианы от различий размаха и асимметрии.
- Плохо работает со связанными рангами.

Лучше использовать тест U Манна-Уитни

Пермутационные методы

Пермутационные методы

- используют перестановки (пермутации, permutations) исходных данных для тестирования гипотез

Пермутационные методы

- используют перестановки (пермутации, permutations) исходных данных для тестирования гипотез
    - о равенстве средних (аналог t-тестов)

Пермутационные методы

- используют перестановки (пермутации, permutations) исходных данных для тестирования гипотез
    - о равенстве средних (аналог t-тестов)
    - о связи между категориальными переменными (таблицы сопряженности)

Пермутационные методы

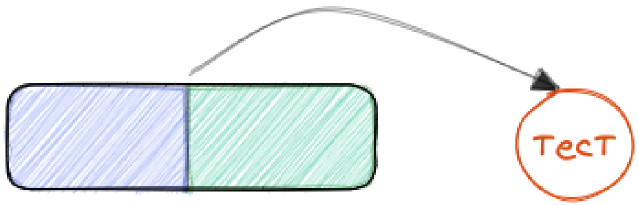
- используют перестановки (пермутации, permutations) исходных данных для тестирования гипотез
  - о равенстве средних (аналог t-тестов)
  - о связи между категориальными переменными (таблицы сопряженности)
  - о связи между непрерывными переменными (коэффициент корреляции)

Пермутационные методы

- используют перестановки (пермутации, permutations) исходных данных для тестирования гипотез
  - о равенстве средних (аналог t-тестов)
  - о связи между категориальными переменными (таблицы сопряженности)
  - о связи между непрерывными переменными (коэффициент корреляции)
- делают меньше предположений о распределении данных

Пермутационный тест в общем виде

0. Вычисляем тестовую статистику на исходных данных.

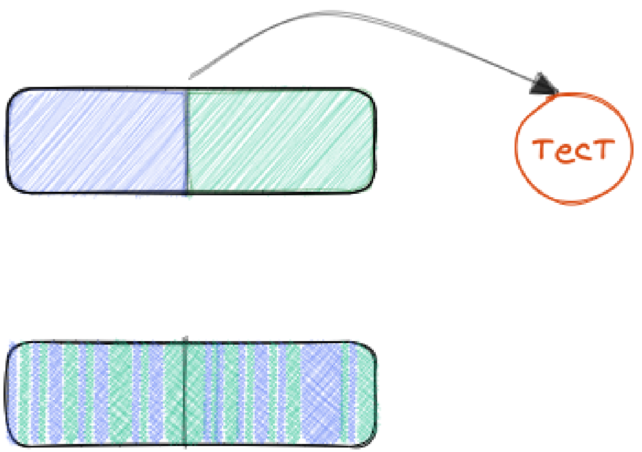


Пермутационный тест в общем виде

0. Вычисляем тестовую статистику на исходных данных.

1. Делаем пермутации

Случайно переставляем данные, смешивая сравниваемые группы.
Это моделирует ситуацию для H_0.



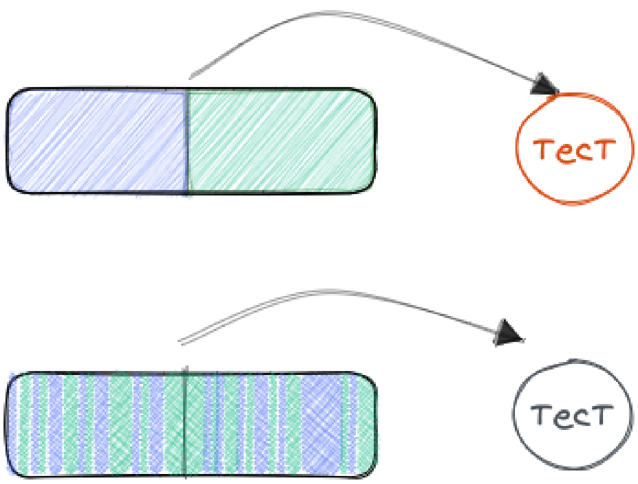
Пермутационный тест в общем виде

0. Вычисляем тестовую статистику на исходных данных.

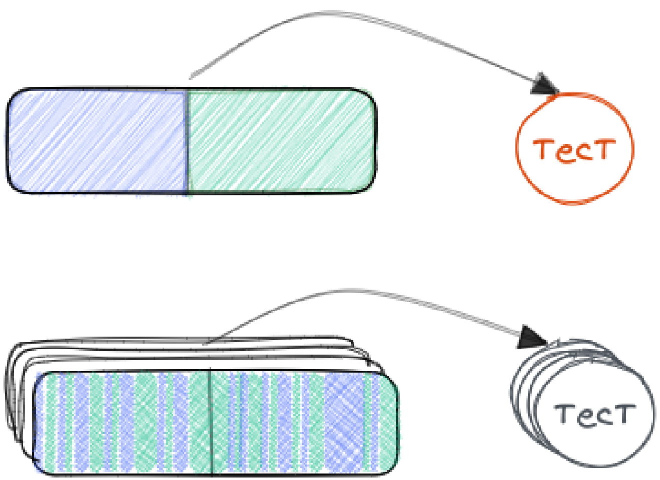
1. Делаем пермутации

Случайно переставляем данные, смешивая сравниваемые группы.
Это моделирует ситуацию для H_0.

2. Вычисляем тестовую статистику на пермутированных данных.



Пермутационный тест в общем виде



0. Вычисляем тестовую статистику на исходных данных.

1. Делаем пермутации

Случайно переставляем данные, смешивая сравниваемые группы.
Это моделирует ситуацию для H_0.

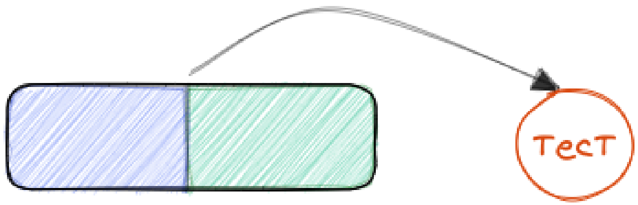
2. Вычисляем тестовую статистику на пермутированных данных.

3. Повторяем 1 и 2 много раз (≥ 1000)

Получаем распределение тестовой статистики при H_0 (с учетом вычисленного про исходным данным значения).

Пермутационный тест в общем виде

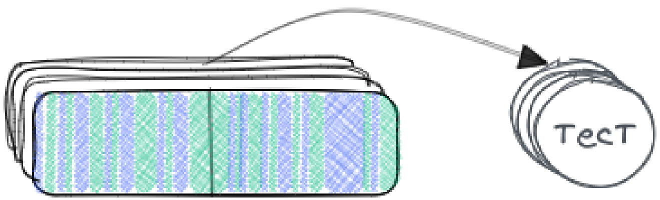
0. Вычисляем тестовую статистику на исходных данных.



1. Делаем пермутации

Случайно переставляем данные, смешивая сравниваемые группы.
Это моделирует ситуацию для H_0.

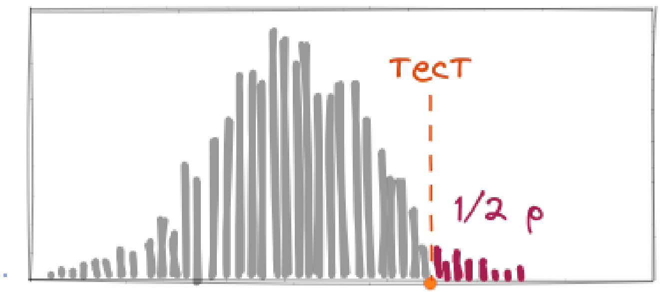
2. Вычисляем тестовую статистику на пермутированных данных.



3. Повторяем 1 и 2 много раз (≥ 1000)

Получаем распределение тестовой статистики при H_0 (с учетом вычисленного про исходным данным значения).

4. Вычисляем p



Например, в двустороннем тесте разницы средних, если статистика >0, *p* — удвоенная доля пермутаций, где значение статистики больше исходного (и наоборот, если < 0).

Условия применимости пермутационного теста о сравнении средних

- данные — это случайная выборка
- распределение переменной одинаковое в сравниваемых группах (устойчив при больших выборках)

Условия применимости пермутационного теста о сравнении средних

- данные — это случайная выборка
- распределение переменной одинаковое в сравниваемых группах (устойчив при больших выборках)
- p-значение зависит от количества пермутаций

Условия применимости пермутационного теста о сравнении средних

- данные — это случайная выборка
- распределение переменной одинаковое в сравниваемых группах (устойчив при больших выборках)
- p-значение зависит от количества пермутаций

Число возможных пермутаций $n!$,
т.е. годится для не слишком малых выборок ($n > 5$ или 7).

Мощность

- При малых выборках мощность меньше, чем у параметрического t-теста, но больше, чем у теста U Манна-Уитни.
- При больших выборках сходная мощность у всех трех типов.

Пермутационный тест для сверчков-каннибалов

Считаем статистику на исходных данных

Аппетит	Время до спаривания
Голодные	1.9
Голодные	2.1
Голодные	3.8
Голодные	9
Голодные	9.6
Голодные	13
Голодные	14.7
Голодные	17.9
Голодные	21.7
Голодные	29
Голодные	72.3
Сытые	1.5
Сытые	1.7
Сытые	2.4
Сытые	3.6
Сытые	5.7

Среднее время до спаривания

у голодных самок $\bar{y}_{\text{голодные}} = 17.73$

у сытых самок $\bar{y}_{\text{сытые}} = 35.98$

Разница $y_{\text{голодные}} - y_{\text{сытые}} = -18.26$

Считаем статистику на пермутированных данных

Аппетит	Время до спаривания
Сытые	1.9
Голодные	2.1
Сытые	3.8
Сытые	9
Голодные	9.6
Сытые	13
Сытые	14.7
Голодные	17.9
Голодные	21.7
Голодные	29
Голодные	72.3
Сытые	1.5
Голодные	1.7
Сытые	2.4
Сытые	3.6
[illegible]	[illegible]

Среднее время до спаривания в этой пермутации получилось

у голодных самок $\bar{y}_{\text{голодные}_1} = 35.66$

у сытых самок $\bar{y}_{\text{сытые}_1} = 20.81$

Разница $y_{\text{голодные}_1} - y_{\text{сытые}_1} = 14.86$

Распределение статистики по пермутированным данным

Если процесс повторить 10000 раз, получится такое распределение

В нашем двустороннем тесте p-значение — это удвоенная доля пермутаций, в которых значение меньше исходного.

$p = 2 \cdot 0.0672 = 0.134$

Распределение статистики по пермутированным данным

Если процесс повторить 10000 раз, получится такое распределение

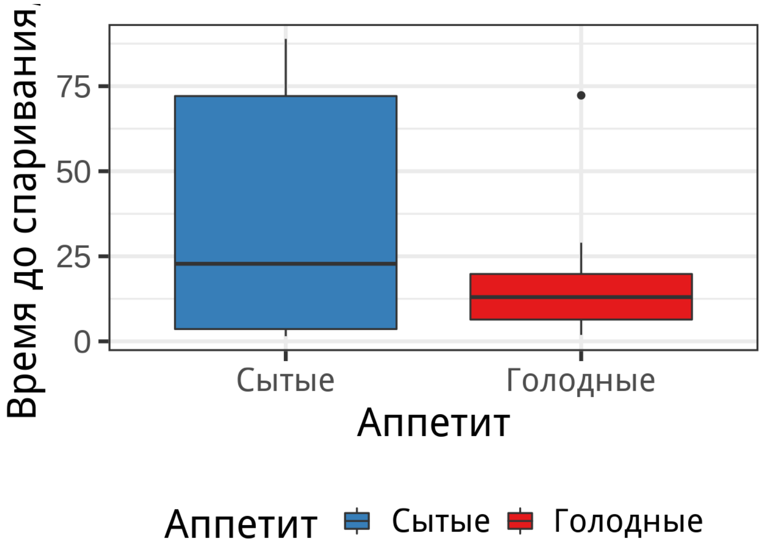
В нашем двустороннем тесте p-значение — это удвоенная доля пермутаций, в которых значение меньше исходного.

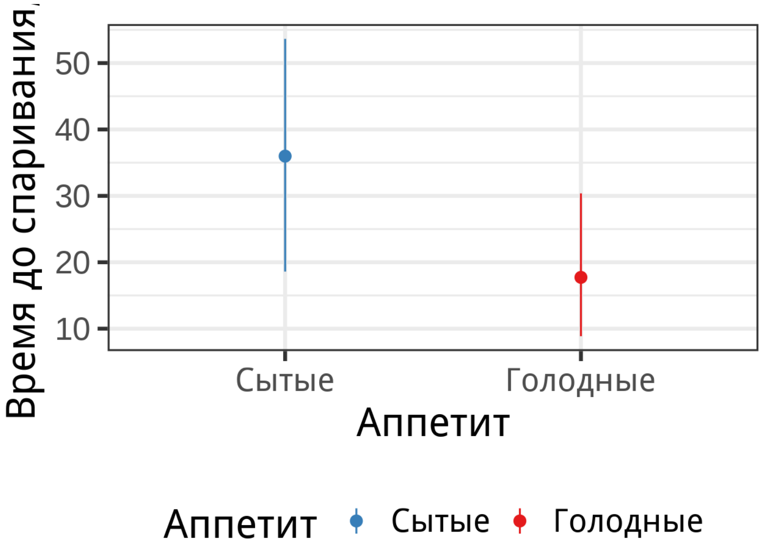
$p = 2 \cdot 0.0672 = 0.134$

Сохраняем H_0

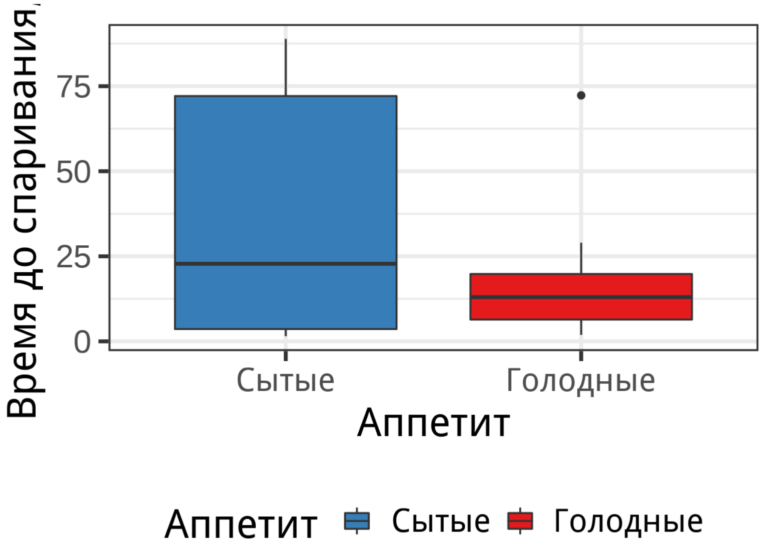
Т.е. не обнаружено статистически-значимых различий среднего времени до спаривания у голодных и сытых самок.

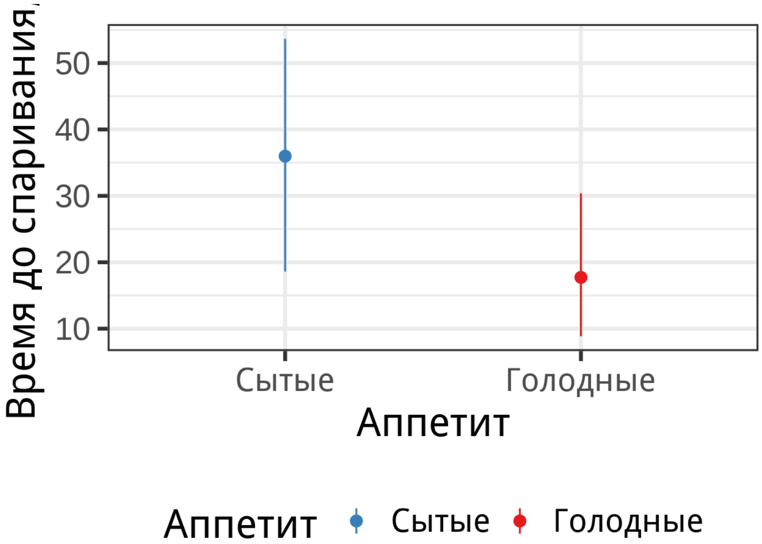
Обращайте внимание, что вы сравниваете





Обращайте внимание, что вы сравниваете





Тест U Манна-Уитни сравнивает медианы, а в пермутационном тесте мы решили сравнивать средние значения (хотя могли бы медианы). Это разные гипотезы.

Summary

Summary

- Условия применимости статистических методов не всегда выполняются.
- Если нарушаются условия применимости метода, можно использовать
    - другой параметрический метод
    - трансформировать данные
    - непараметрические методы
    - пермутационные методы

Что почитать

Whitlock, M., & Schluter, D. (2015). The analysis of biological data (Second edition). Roberts and Company Publishers.