

Прикладной статистический анализ данных. 3. Проверка непараметрических гипотез.

Ольга Добролюбова
Юлиан Сердюк
cs.msu.psad@gmail.com

25.02.2022

Виды задач

Одновыборочные:

X^n

среднее выборки равно заданному числу 1 3 8

Двухвыборочные:

$X_1^{n_1}, X_2^{n_2}$

средние выборок равны

X_1, X_2 связанные 2 4 9

X_1, X_2 независимые 5 10

дисперсии выборок равны 6 11

Варианты двухвыборочных гипотез

О положении:

$$H_0: \mathbb{E}X_1 = \mathbb{E}X_2,$$

$$H_1: \mathbb{E}X_1 <\neq> \mathbb{E}X_2;$$

$$H_0: \text{med } X_1 = \text{med } X_2,$$

$$H_1: \text{med } X_1 <\neq> \text{med } X_2;$$

$$H_0: \mathbf{P}(X_1 > X_2) = 0.5,$$

$$H_1: \mathbf{P}(X_1 > X_2) <\neq> 0.5;$$

$$H_0: F_{X_1}(x) = F_{X_2}(x),$$

$$H_1: F_{X_1}(x) = F_{X_2}(x + \Delta), \Delta <\neq> 0;$$

$$H_0: F_{X_1}(x) = F_{X_2}(x),$$

$$H_1: F_{X_1}(x) <\neq> F_{X_2}(x).$$

О рассеянии:

$$H_0: \mathbb{D}X_1 = \mathbb{D}X_2,$$

$$H_1: \mathbb{D}X_1 <\neq> \mathbb{D}X_2;$$

$$H_0: F_{X_1}(x) = F_{X_2}(x + \Delta),$$

$$H_1: F_{X_1}(x) = F_{X_2}(\sigma x + \Delta), \sigma <\neq> 1.$$

(1) Одновыборочный критерий знаков

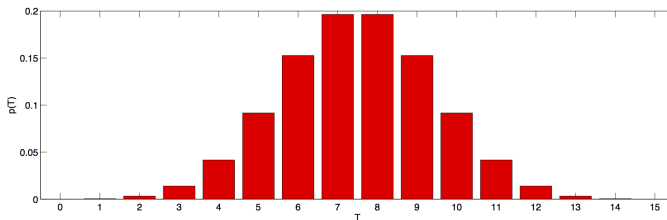
выборка: $X^n = (X_1, \dots, X_n), X_i \neq m_0$

нулевая гипотеза: $H_0: \text{med } X = m_0$

альтернатива: $H_1: \text{med } X < \neq > m_0$

статистика: $T(X^n) = \sum_{i=1}^n [X_i > m_0]$

нулевое распределение: $\text{Bin}(n, \frac{1}{2})$



(1) Одновыборочный критерий знаков

Пример 1 (Dinse, 1982): выживаемость пациентов с лимфоцитарной лимфомой (в неделях):

49, 58, 75, 110, 112, 132, 151, 276, 281, 362*

Исследование длилось 7 лет, поэтому для пациентов, проживших дольше, выживаемость неизвестна (выборка цензурирована сверху).

Превышает ли среднее время дожития 200 недель?

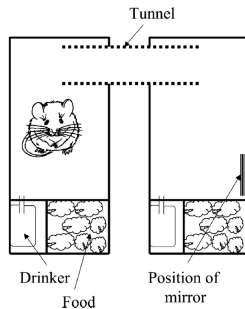
H_0 : медиана времени дожития не больше 200 недель.

H_1 : медиана времени дожития больше 200 недель.

Критерий знаков: $p = 0.9453$.

(1) Одновыборочный критерий знаков

Пример 2: (Shervin, 2004): 16 лабораторных мышей были помещены в двухкомнатные клетки, в одной из комнат висело зеркало. Измерялась доля времени, которое каждая мышь проводила в каждой из своих двух клеток.

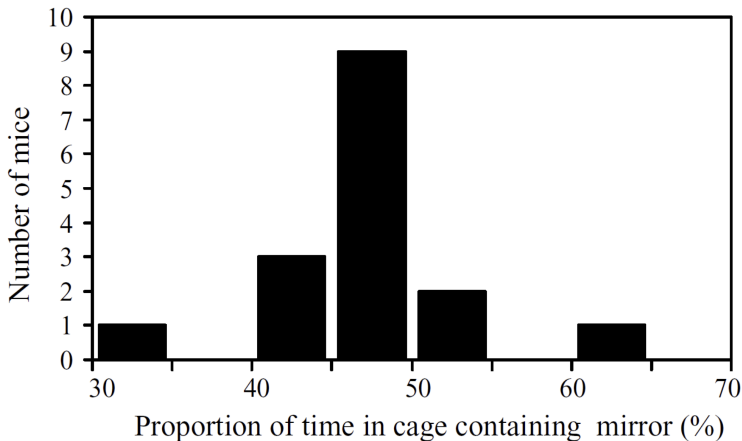


Общая постановка:

H_0 : мышам всё равно, висит в клетке зеркало или нет.

H_1 : у мышей есть какие-то предпочтения насчёт зеркала.

(1) Одновыборочный критерий знаков



Средняя доля времени, проводимого в клетке с зеркалом — $47.6 \pm 4.7\%$.

(1) Одновыборочный критерий знаков

H_0 : медиана доли времени, проводимого в клетке с зеркалом, равна $\frac{1}{2}$.

H_1 : медиана доли времени, проводимого в клетке с зеркалом, не равна $\frac{1}{2}$.

Редуцированные данные: 0 — мышь провела больше времени в комнате с зеркалом, 1 — в комнате без зеркала.

Статистика: T — число единиц в выборке.

13 из 16 мышей провели больше времени в комнате без зеркала.

Критерий знаков: $p = 0.0213$; доверительный интервал для медианы доли времени, проведённого в комнате с зеркалом:

- $[0.4507, 0.4887]$ — с уровнем доверия 92.32%
- $[0.4263, 0.4894]$ — с уровнем доверия 97.87%
- $[0.4389, 0.4890]$ — приближённый 95% (линейная интерполяция)

(2) Двухвыборочный критерий знаков

выборки: $X_1^n = (X_{11}, \dots, X_{1n})$
 $X_2^n = (X_{21}, \dots, X_{2n}), X_{1i} \neq X_{2i}$

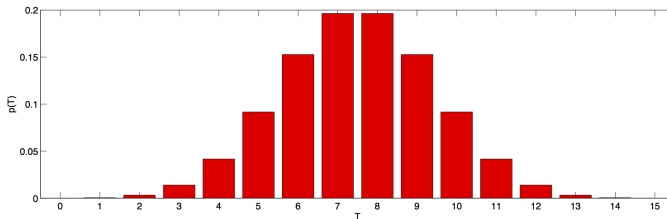
выборки связанные

нулевая гипотеза: $H_0: \mathbf{P}(X_1 > X_2) = \frac{1}{2}$

альтернатива: $H_1: \mathbf{P}(X_1 > X_2) < \neq > \frac{1}{2}$

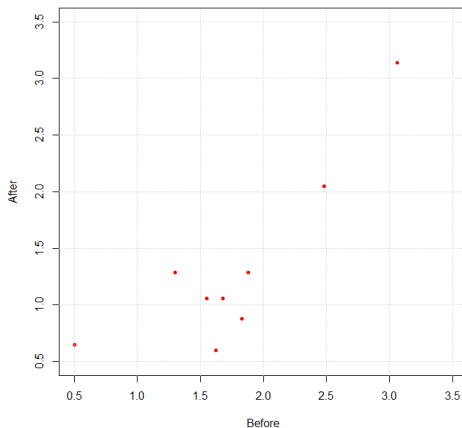
статистика: $T(X_1^n, X_2^n) = \sum_{i=1}^n [X_{1i} > X_{2i}]$

нулевое распределение: $\text{Bin}(n, \frac{1}{2})$



(2) Двухвыборочный критерий знаков

Пример 1 (Hollander & Wolfie, 29f): депрессивность 9 пациентов была измерена по шкале Гамильтона до и после первого приёма транквилизатора. Подействовал ли транквилизатор?



(2) Двухвыборочный критерий знаков

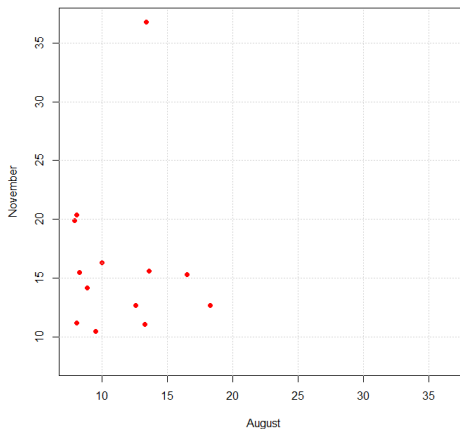
H_0 : уровень депрессивности не изменился.

H_1 : уровень депрессивности снизился.

Критерий знаков: $p = 0.09$, 95% нижний доверительный предел для медианы изменения — -0.041 .

(2) Двухвыборочный критерий знаков

Пример 2: (Laureysens et al., 2004): для 13 разновидностей тополей, растущих в зоне интенсивного загрязнения, в августе и ноябре измерялась средняя концентрация алюминия в микрограммах на грамм древесины.



(2) Двухвыборочный критерий знаков

H_0 : концентрация алюминия не менялась.

H_1 : концентрация алюминия изменилась.

Для тополей 10 из 13 разновидностей концентрация алюминия увеличилась.

Критерий знаков: $p = 0.0923$, 95% доверительный интервал для медианы изменения — $[-0.687, 10.107]$.

Причины использовать критерий знаков

- Точные разности Δx_i неизвестны, известны только их знаки (сравнение агрессивности комаров).
- Разности Δx_i при H_1 могут быть небольшими по модулю, но иметь систематический характер по знаку (пример с мышами).
- Разности Δx_i при H_0 могут быть большими по модулю, но случайными по знаку (влияние меди на число личинок комаров).

Вариационный ряд, ранги, связи

$$X_1, \dots, X_n \Rightarrow X_{(1)} \leq \dots < \underbrace{X_{(k_1)} = \dots = X_{(k_2)}}_{\text{связка размера } k_2 - k_1 + 1} < \dots \leq X_{(n)}$$

Ранг наблюдения X_i :

если X_i не в связке, то $\text{rank}(X_i) = r: X_i = X_{(r)}$,

если X_i в связке $X_{(k_1)}, \dots, X_{(k_2)}$, то $\text{rank}(X_i) = \frac{k_1 + k_2}{2}$.

(3) Одновыборочный критерий знаковых рангов Уилкоксона

выборка: $X^n = (X_1, \dots, X_n), X_i \neq m_0$

$F(X)$ симметрично относительно медианы

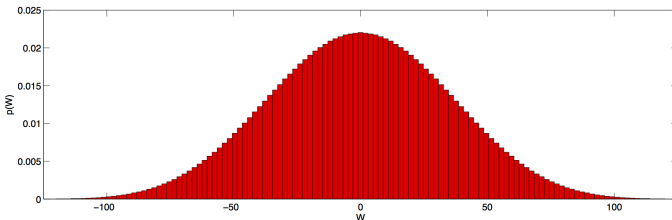
$F(x) = 1 - F(2m - x)$, где m - медиана $F(X)$

нулевая гипотеза: $H_0: \text{med } X = m_0$

альтернатива: $H_1: \text{med } X < \neq > m_0$

статистика: $W(X^n) = \sum_{i=1}^n \text{rank}(|X_i - m_0|) \cdot \text{sign}(X_i - m_0)$

нулевое распределение: табличное



(3) Одновыборочный критерий знаковых рангов Уилкоксона

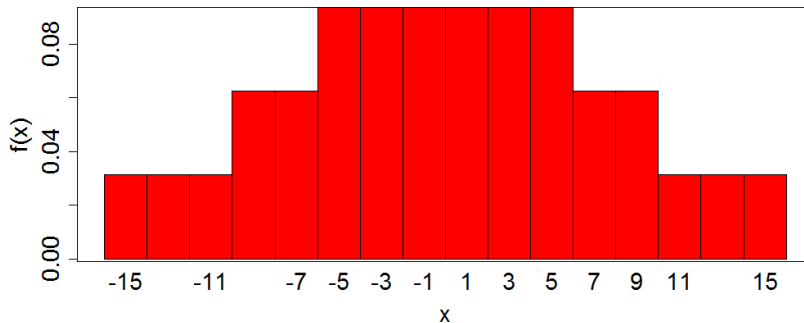
Откуда берётся табличное распределение?

1	2	3	4	5	W
—	—	—	—	—	—15
+	—	—	—	—	—13
—	+	—	—	—	—11
+	+	—	—	—	—9
—	—	+	—	—	—9
...
+	+	—	+	+	9
—	—	+	+	+	9
+	—	+	+	+	11
—	+	+	+	+	13
+	+	+	+	+	15

Всего 2^n вариантов.

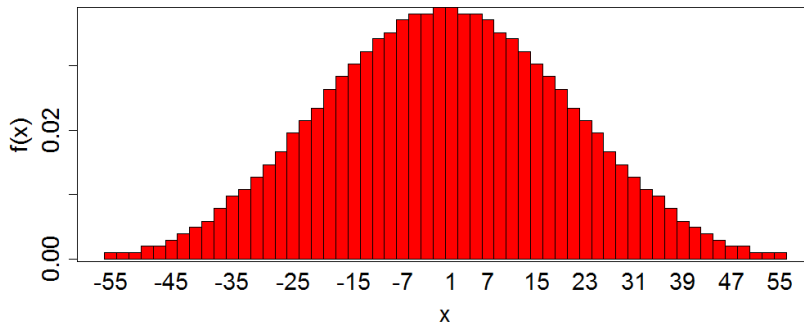
(3) Одновыборочный критерий знаковых рангов Уилкоксона

$n = 5$:



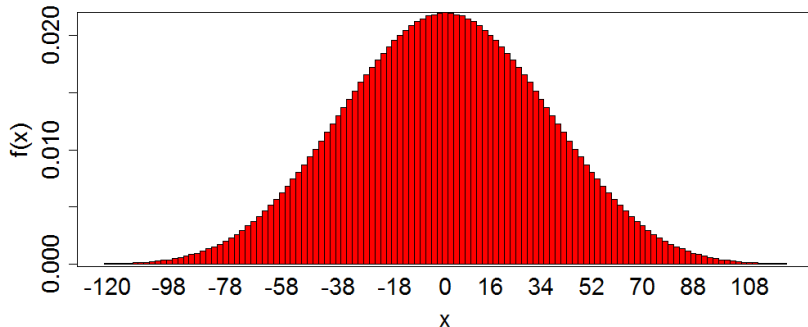
(3) Одновыборочный критерий знаковых рангов Уилкоксона

$n = 10$:



(3) Одновыборочный критерий знаковых рангов Уилкоксона

$n = 15$:

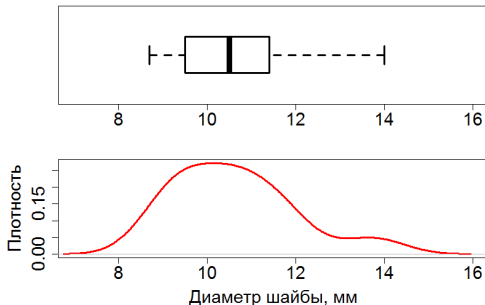


Аппроксимация для $n > 20$:

$$W \approx N \left(0, \frac{n(n+1)(2n+1)}{6} \right).$$

(3) Одновыборочный критерий знаковых рангов Уилкоксона

Пример 1 (Bonpini, табл. 1.4): диаметры шайб на производстве ($n = 24$):



Соответствуют ли шайбы стандартному размеру 10 мм?

H_0 : средний диаметр шайбы — 10 мм, $\text{med } X = 10$.

H_1 : средний диаметр шайбы не соответствует стандарту, $\text{med } X \neq 10$.

Критерий знаковых рангов: $p = 0.0673$, выборочная медиана диаметра — 10.5 мм (95% доверительный интервал — $[9.95, 11.15]$ мм).

(3) Одновыборочный критерий знаковых рангов Уилкоксона

Пример 2 (зеркала в клетках мышей):

H_0 : медиана доли времени, проводимого в клетке с зеркалом, равна $\frac{1}{2}$.

H_1 : медиана доли времени, проводимого в клетке с зеркалом, не равна $\frac{1}{2}$.

Критерий знаковых рангов: $p = 0.0934$.

(4) Критерий знаковых рангов Уилкоксона для связанных выборок

выборки: $X_1^n = (X_{11}, \dots, X_{1n})$
 $X_2^n = (X_{21}, \dots, X_{2n}), X_{1i} \neq X_{2i}$

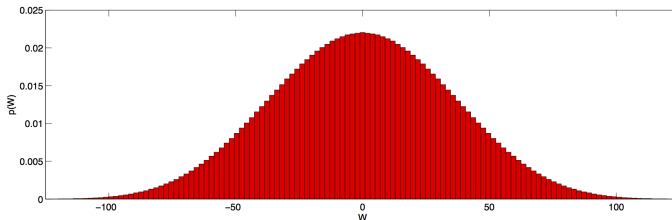
выборки связанные

нулевая гипотеза: $H_0: \text{med}(X_1 - X_2) = 0$

альтернатива: $H_1: \text{med}(X_1 - X_2) < \neq > 0$

статистика: $W(X_1^n, X_2^n) = \sum_{i=1}^n \text{rank}(|X_{1i} - X_{2i}|) \cdot \text{sign}(X_{1i} - X_{2i})$

нулевое распределение: табличное



(4) Критерий знаковых рангов Уилкоксона для связанных выборок

Пример 1 (Kanji, критерий 48): управляемый вручную станок на каждом шаге процесса производит пару пружин. Для 14 пар измерена прочность:

$X_1: \{1.38, 0.39, 1.42, 0.54, 5.94, 0.59, 2.67, 2.44, 0.56, 0.69, 0.71, 0.95, 0.50, 9.69\},$

$X_2: \{1.42, 0.39, 1.46, 0.55, 6.15, 0.61, 2.69, 2.68, 0.53, 0.72, 0.72, 0.93, 0.53, 10.37\}.$

Одинакова ли прочность пружин в паре?

H_0 : средние значение прочности пружин в паре равны.

H_1 : средние значение прочности пружин в паре не равны $\Rightarrow p = 0.0142$,
95% доверительный интервал для медианной разности — $[0.005, 0.14]$.

(4) Критерий знаковых рангов Уилкоксона для связанных выборок

Пример 2 (алюминий в тополях):

H_0 : медиана изменения концентрации алюминия равна нулю.

H_1 : медиана изменения концентрации алюминия не равна нулю $\Rightarrow 0.0398$,
95% доверительный интервал для медианы изменения — $[0.35, 9.3]$.

(5) Критерий Манна-Уитни-Уилкоксона

выборки: $X_1^{n_1} = (X_{11}, \dots, X_{1n_1})$

$X_2^{n_2} = (X_{21}, \dots, X_{2n_2})$

выборки независимые

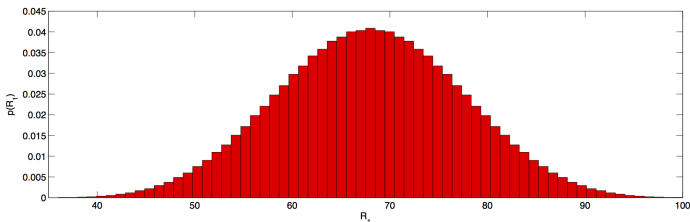
нулевая гипотеза: $H_0: F_{X_1}(x) = F_{X_2}(x)$

альтернатива: $H_1: F_{X_1}(x) = F_{X_2}(x + \Delta), \Delta < \neq > 0$

статистика: $X_{(1)} \leq \dots \leq X_{(n_1+n_2)}$ — вариационный ряд
объединённой выборки $X = X_1^{n_1} \cup X_2^{n_2}$

$$R_1(X_1^{n_1}, X_2^{n_2}) = \sum_{i=1}^{n_1} \text{rank}(X_{1i})$$

нулевое распределение: табличное



(5) Критерий Манна-Уитни-Уилкоксона

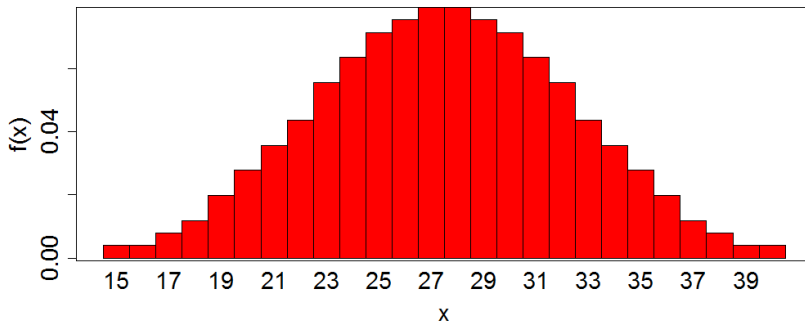
Откуда берётся табличное распределение?

X_1	X_2	R_1
{1,2,3}	{4,5,6,7}	6
{1,2,4}	{3,5,6,7}	7
{1,2,5}	{3,4,6,7}	8
{1,2,6}	{3,4,5,7}	9
{1,2,7}	{3,4,5,6}	10
{1,3,4}	{2,5,6,7}	8
...
{3,5,7}	{1,2,4,6}	15
{3,6,7}	{1,2,4,5}	16
{4,5,6}	{1,2,3,7}	15
{4,5,7}	{1,2,3,6}	16
{4,6,7}	{1,2,3,5}	17
{5,6,7}	{1,2,3,4}	18

Всего $C_{n_1+n_2}^{m_1}$ вариантов.

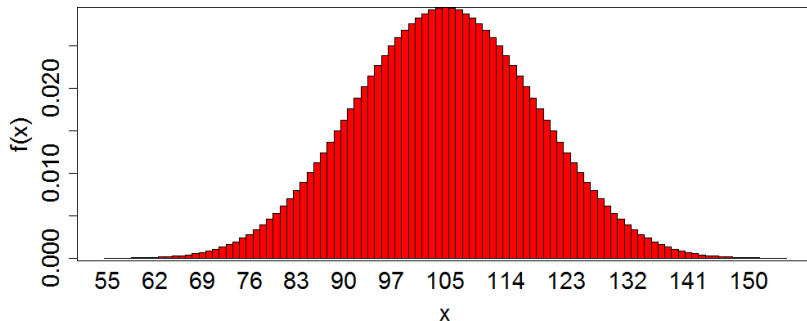
(5) Критерий Манна-Уитни-Уилкоксона

$n_1 = n_2 = 5$:



(5) Критерий Манна-Уитни-Уилкоксона

$n_1 = n_2 = 10$:



Аппроксимация для $n_1, n_2 > 10$:

$$R_1 \sim N \left(\frac{n_1(n_1 + n_2 + 1)}{2}, \frac{n_1 n_2 (n_1 + n_2 + 1)}{12} \right).$$

(5) Критерий Манна-Уитни-Уилкоксона

Пример 1 (Капji, критерий 52): сотрудник налоговой службы хочет сравнить средние значения в двух выборках заявленных трат на компенсацию командировочных расходов в одной и той же компании в двух разных периодах (расходы скорректированы на инфляцию).

$$X_1: \{50.5, 37.5, 49.8, 56.0, 42.0, 56.0, 50.0, 54.0, 48.0\},$$

$$X_2: \{57.0, 52.0, 51.0, 44.2, 55.0, 62.0, 59.0, 45.2, 53.5, 44.4\}.$$

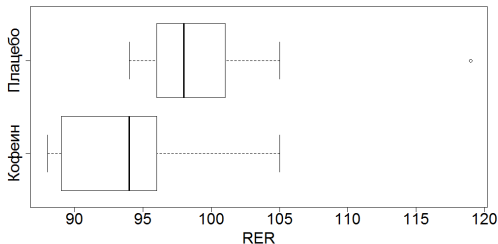
Равны ли средние расходы?

H_0 : средние расходы равны.

H_1 : средние расходы не равны $\Rightarrow p = 0.3072$, 95% доверительный интервал для медианной разности — $[-9, 4]$.

(5) Критерий Манна-Уитни-Уилкоксона

RER — соотношение числа молекул CO_2 и O_2 в выдыхаемом воздухе. В эксперименте измерялся респираторный обмен 18 испытуемых в процессе физических упражнений. За час до этого 9 из них получили таблетку кофеина, 9 — плацебо.



Повлиял ли кофеин на значение RER?

H_0 : среднее значение показателя респираторного обмена не отличается в двух группах.

H_1 : среднее значение показателя респираторного обмена отличается в двух группах.

(5) Критерий Манна-Уитни-Уилкоксона

Ранг	Наблюдение	Номер наблюдения	Наблюдение	Ранг
16.5	105	1	96	9
18	119	2	99	13
14	100	3	94	5.5
11	97	4	89	3
9	96	5	96	9
15	101	6	93	4
5.5	94	7	88	1.5
7	95	8	105	16.5
12	98	9	88	1.5

Статистика R_1 — сумма рангов в одной из групп.

$p = 0.0521$, сдвиг между средними — 6 пунктов, (95% доверительный интервал — $[-0.00005, 12]$ пт).

(6) Критерий Ансари-Брэдли

выборки: $X_1^{n_1} = (X_{11}, \dots, X_{1n_1})$
 $X_2^{n_2} = (X_{21}, \dots, X_{2n_2})$

выборки независимые, $\text{med}(X_1) = \text{med}(X_2)$

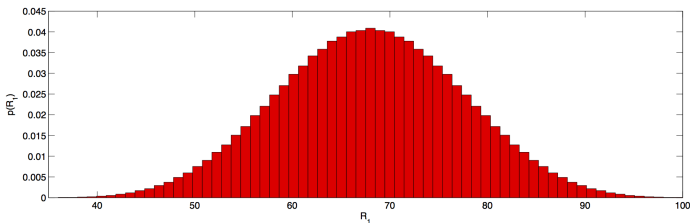
нулевая гипотеза: $H_0: \mathbb{D}X_1 = \mathbb{D}X_2$

альтернатива: $H_1: \mathbb{D}X_1 < \neq \mathbb{D}X_2$

статистика: $X_{(1)} \leq \dots \leq X_{(N)}$ — вариационный ряд
 объединённой выборки $X^N = X_1^{n_1} \cup X_2^{n_2}, N = n_1 + n_2$

$$R_1(X_1^{n_1}, X_2^{n_2}) = \sum_{i=1}^{n_1} \widetilde{\text{rank}}(X_{1i})$$

нулевое распределение: табличное

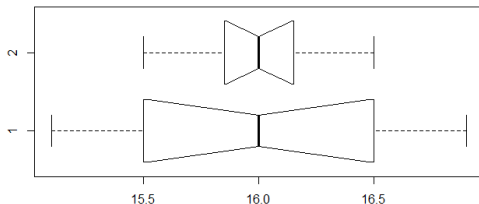


Ранги присваиваются от краёв к центру:

$$\begin{array}{ccccccccccc} X_{(i)} & X_{(1)} & X_{(2)} & X_{(3)} & \dots & X_{(N-2)} & X_{(N-1)} & X_{(N)} \\ \widetilde{\text{rank}}(X_{(i)}) & 1 & 2 & 3 & & 3 & 2 & 1 \end{array}$$

(6) Критерий Ансари-Брэдли

Пример (Bonini, табл. 2.1): два поставщика шестнадцатикилограммовых свинцовых слитков выслали по выборке образцов. Средний вес образцов в обеих выборках соответствует норме; различаются ли дисперсии?



H_0 : дисперсия веса слитков не отличается для двух поставщиков.

H_1 : дисперсия веса слитков для двух поставщиков отличается

$\Rightarrow p = 0.014$.

Построение доверительных интервалов

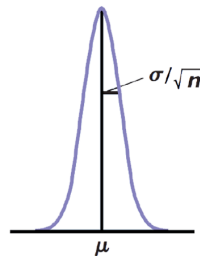
Как можно оценить $F_{\hat{\theta}_n}(x)$ — выборочное распределение статистики $\hat{\theta}_n$? (Hesterberg, 2005):

- параметрический метод:



НОРМАЛЬНАЯ ПОПУЛЯЦИЯ
неизвестное среднее μ

Теория
→



Выборочное распределение

Сделать предположение, что X распределена по закону $F_X(x)$, при выполнении которого закон распределения $\hat{\theta}_n$ известен.

Построение доверительных интервалов

- наивный метод:



ПОПУЛЯЦИЯ

неизвестное среднее μ

SRS объёма n

SRS объёма n

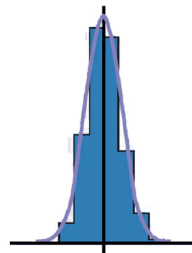
SRS объёма n

\bar{x}

\bar{x}

\bar{x}

·
·
·

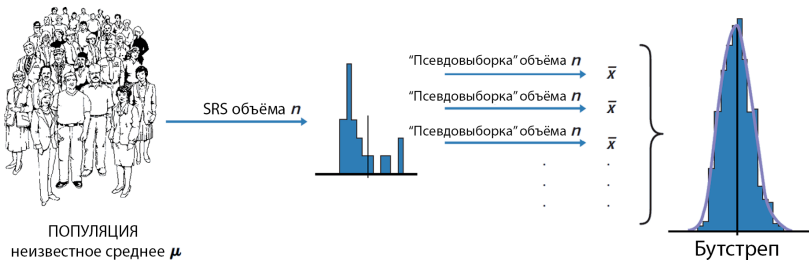


Выборочное распределение

Извлечь из генеральной совокупности N выборки объёма n и оценить выборочное распределение $\hat{\theta}_n$ эмпирическим.

Построение доверительных интервалов

• бутстреп:

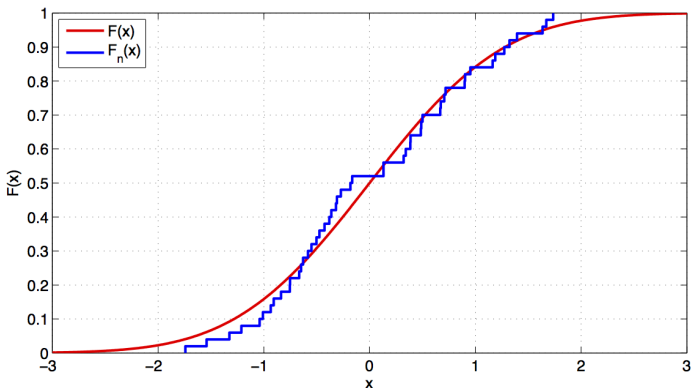


Сгенерировать N «псевдовыборок» объёма n и оценить выборочное распределение $\hat{\theta}_n$ «псевдоэмпирическим».

Бутстреп

Извлечение выборок из генеральной совокупности — сэмплирование из неизвестного распределения $F_X(x)$.

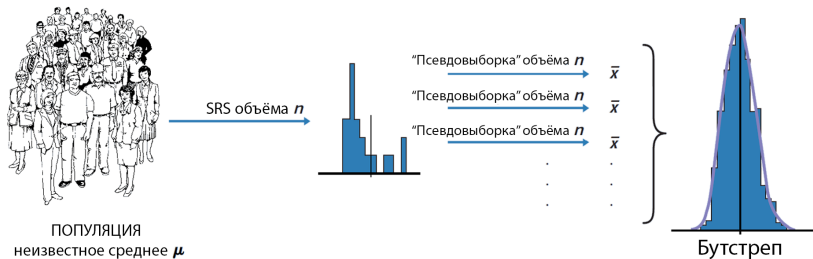
Лучшая оценка $F_X(x)$, которая у нас есть — $F_{X^n}(x)$:



Сэмплировать из неё — это то же самое, что делать из X^n выборки с возвращением объёма n .

Бутстреп-распределение

X^{1*}, \dots, X^{N*} — бутстреп-псевдовыборки из X^n объёма n ,
 $\hat{\theta}_n^{1*}, \dots, \hat{\theta}_n^{N*}$ — значения статистики на них,
 $F_{\hat{\theta}_n}^{boot}(x)$ — бутстреп-распределение $\hat{\theta}_n$ — эмпирическая функция
распределения, построенная по значениям статистики на псевдовыборках.



По $F_{\hat{\theta}_n}^{boot}(x)$ можно строить доверительные интервалы для θ !

Доверительные интервалы

- Посчитаем S_n^{boot} — выборочное стандартное отклонение $\hat{\theta}_n$ на псевдовыборках;

$$\mathbf{P}\left(\hat{\theta}_n - t_{n-1, 1-\frac{\alpha}{2}} S_n^{boot} \leq \theta \leq \hat{\theta}_n + t_{n-1, 1-\frac{\alpha}{2}} S_n^{boot}\right) \approx 1 - \alpha.$$

Это стьюдентизированный бутстреп.

- Возьмём выборочные квантили бутстреп-распределения:

$$\mathbf{P}\left(\left(F_{\hat{\theta}_n}^{boot}\right)^{-1}\left(\frac{\alpha}{2}\right) \leq \theta \leq \left(F_{\hat{\theta}_n}^{boot}\right)^{-1}\left(1 - \frac{\alpha}{2}\right)\right) \approx 1 - \alpha.$$

Это базовый бутстреп.

Доверительные интервалы

- Слегка изменим наивный бутстреп:

$$\mathbf{P}\left(\left(F_{\hat{\theta}_n}^{boot}\right)^{-1}(\alpha_1) \leq \theta \leq \left(F_{\hat{\theta}_n}^{boot}\right)^{-1}(\alpha_2)\right) \approx 1 - \alpha,$$

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\frac{\alpha}{2}}}{1 - \hat{a}\left(\hat{z}_0 + z_{\frac{\alpha}{2}}\right)}\right),$$

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\frac{\alpha}{2}}}{1 - \hat{a}\left(\hat{z}_0 + z_{1-\frac{\alpha}{2}}\right)}\right),$$

$$\hat{z}_0 = \Phi^{-1}\left(\frac{1}{N} \sum_{i=1}^N \left[\hat{\theta}_n^{i*} < \hat{\theta}_n\right]\right),$$

\hat{a} не поместится на этом слайде.

Это несмещённый ускоренный бутстреп.

Свойства бутстрепа

- асимптотическая состоятельность
- простота использования даже для самых сложных статистик
- плохо работает для статистик, значение которых зависит от небольшого числа элементов выборки

Перестановочные критерии

Ранговые критерии:

- 1 выборки \Rightarrow ранги
- 2 дополнительное предположение (о равенстве распределений / медиан и пр.)
- 3 перестановки \Rightarrow нулевое распределение статистики

Что если пропустить пункт 1?

(8) Одновыборочный перестановочный критерий, гипотеза о среднем

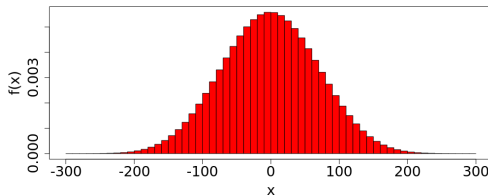
выборка: $X_1^n = (X_1, \dots, X_n)$ $F(X)$ симметрично относительно матожиданиянулевая гипотеза: $H_0: \mathbb{E}X = m_0$ альтернатива: $H_1: \mathbb{E}X < \neq > m_0$ статистика: $T(X^n) = \sum_{i=1}^n (X_i - m_0)$ нулевое распределение: порождается перебором 2^n знаков
перед слагаемыми $X_i - m_0$

Достижимый уровень значимости — доля перестановок знаков, на которых получилось такое же или ещё более экстремальное значение статистики.

(8) Одновыборочный перестановочный критерий, гипотеза о среднем

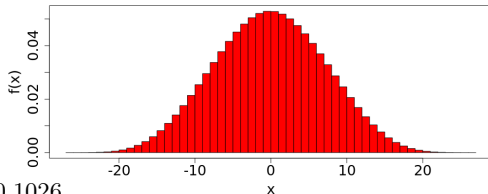
Пример (диаметры шайб):

Критерий знаковых рангов:



$$p = 0.0673$$

Перестановочный критерий:



$$T = 14.6, p = 0.1026$$

95% доверительный интервал для среднего диаметра (ВСа бутстреп) — [10.11, 11.20].

(9) Двухвыборочный перестановочный критерий, гипотеза о средних, связанные выборки

выборки: $X_1^n = (X_{11}, \dots, X_{1n})$

$X_2^n = (X_{21}, \dots, X_{2n})$

выборки связанные

распределение попарных разностей симметрично

нулевая гипотеза: $H_0: \mathbb{E}(X_1 - X_2) = 0$

альтернатива: $H_1: \mathbb{E}(X_1 - X_2) < \neq > 0$

статистика: $D_i = X_{1i} - X_{2i}$

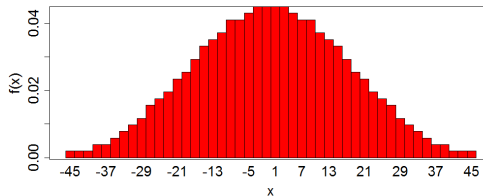
$$T(X_1^n, X_2^n) = \sum_{i=1}^n D_i$$

нулевое распределение: порождается перебором 2^n знаков перед слагаемыми D_i

(9) Двухвыборочный перестановочный критерий, гипотеза о средних, связанные выборки

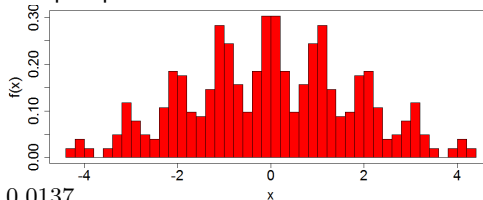
Пример (лечение депрессии):

Критерий знаковых рангов:



$$p = 0.019$$

Перестановочный критерий:



$$T = 3.887, p = 0.0137$$

95% доверительный интервал для среднего уменьшения депрессивности (ВСа бутстреп) — $[0.1658, 0.6834]$.

(10) Двухвыборочный перестановочный критерий, гипотеза о средних, независимые выборки

выборки: $X_1^{n_1} = (X_{11}, \dots, X_{1n_1})$

$X_2^{n_2} = (X_{21}, \dots, X_{2n_2})$

нулевая гипотеза: $H_0: F_{X_1}(x) = F_{X_2}(x)$

альтернатива: $H_1: F_{X_1}(x) = F_{X_2}(x + \Delta), \Delta \neq 0$

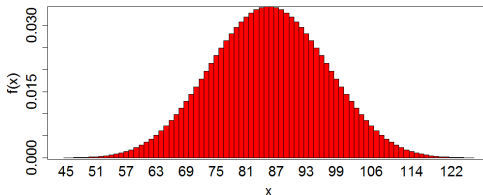
статистика: $T(X_1^{n_1}, X_2^{n_2}) = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{1i} - \frac{1}{n_2} \sum_{i=1}^{n_2} X_{2i}$

нулевое распределение: порождается перебором $C_{n_1+n_2}^{n_1}$
размещений объединённой выборки

(10) Двухвыборочный перестановочный критерий, гипотеза о средних, независимые выборки

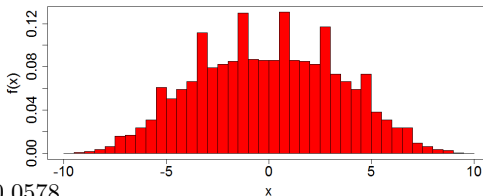
Пример (кофеин и респираторный обмен):

Критерий Манна-Уитни:



$$p = 0.0521$$

Перестановочный критерий:



$$T = 6.33, p = 0.0578$$

95% доверительный интервал для разности средних (BCa бутстреп) —
[1.556, 13.667].

(11) Двухвыборочный перестановочный критерий, гипотеза о дисперсиях, статистика Али

выборки: $X_1^n = (X_{11}, \dots, X_{1n})$

$X_2^n = (X_{21}, \dots, X_{2n})$

выборки независимые

нулевая гипотеза: $H_0: \mathbb{D}X_1 = \mathbb{D}X_2$

альтернатива: $H_1: \mathbb{D}X_1 < \neq > \mathbb{D}X_2$

статистика: $\delta(D_1^{n-1}) = \sum_{i=1}^{n-1} i(n-i)D_{1i},$

$D_{1i} = X_{1(i+1)} - X_{1(i)}$

нулевое распределение: порождается перебором 2^{n-1}

попарных перестановок D_{1i} и D_{2i}

Особенности перестановочных критериев

- Статистику критерия можно выбрать разными способами.
В некоторых случаях разные статистики приведут к одному и тому же достигаемому уровню значимости:

$$X^n, \quad H_0: \mathbb{E}X = 0, \quad H_1: \mathbb{E}X \neq 0,$$

$$T_1(X^n) = \sum_{i=1}^n X_i \sim T_2(X^n) = \bar{X}.$$

В других случаях достигаемый уровень значимости будет зависеть от выбора статистики:

$$T_2(X^n) = \bar{X} \approx T_3(X^n) = \frac{\bar{X}}{S/\sqrt{n}}.$$

- Если множество перестановок G слишком велико, для оценки нулевого распределения T достаточно взять случайное подмножество $G' \in G$. При этом стандартное отклонение достигаемого уровня значимости будет равно примерно $\sqrt{\frac{p(1-p)}{|G'|}}$.

Перестановки и бутстреп

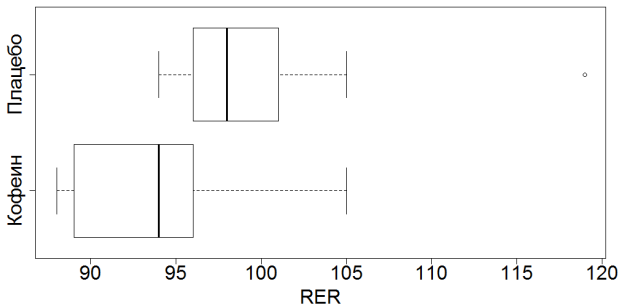
Перестановочные критерии:

- 1 выборки, статистика
- 2 дополнительное предположение
- 3 перестановки \Rightarrow нулевое распределение статистики

Бутстреповые доверительные интервалы:

- 1 выборки, статистика, оценивающая параметр
- 2 бутстреп-псевдовыборки \Rightarrow приближённое распределение статистики

Кофеин и респираторный обмен



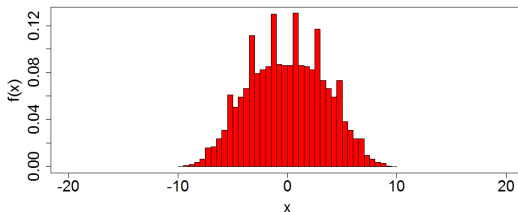
H_0 : среднее значение показателя респираторного обмена не отличается в двух группах.

H_1 : под воздействием кофеина среднее значение показателя респираторного обмена снижается.

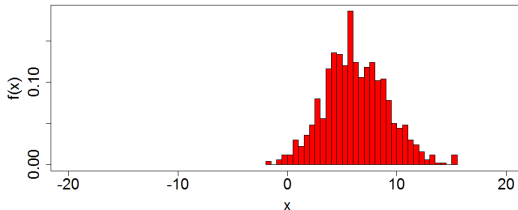
$$\bar{X}_{1n} - \bar{X}_{2n} = 6.33$$

Кофеин и респираторный обмен

Нулевое распределение перестановочного критерия со статистикой $\bar{X}_{1n} - \bar{X}_{2n}$:

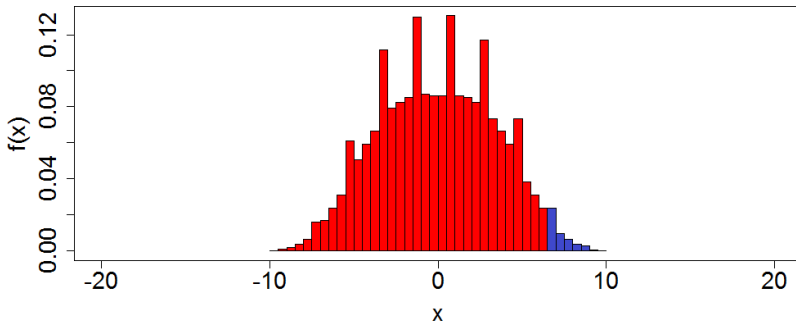


Бутстреп-распределение статистики $\bar{X}_{1n} - \bar{X}_{2n}$:



Кофеин и респираторный обмен

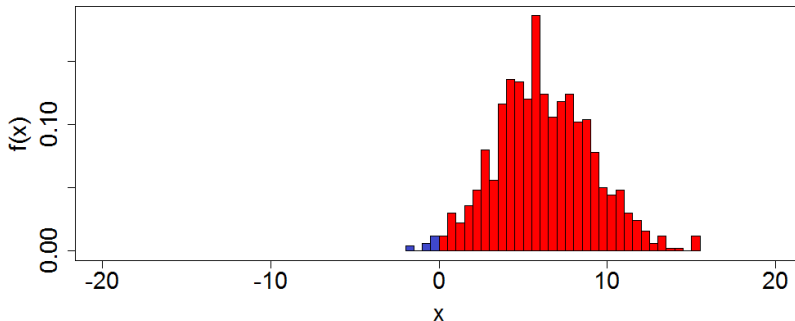
Нулевое распределение перестановочного критерия со статистикой $\bar{X}_{1n} - \bar{X}_{2n}$:



Доля перестановок, на которых среднее больше либо равно 6.33 — 0.0289. Это точный достигаемый уровень значимости перестановочного критерия.

Кофеин и респираторный обмен

Бутстреп-распределение статистики $\bar{X}_{1n} - \bar{X}_{2n}$:



Доля псевдовыборок, на которых среднее меньше либо равно нулю — 0.011.

Это приближённый достигаемый уровень значимости бутстреп-критерия.

Перестановки vs. бутстреп

- ①
 - Перестановочный критерий измеряет расстояние от 0 до \bar{D}_n
 - Бутстреп-критерий измеряет расстояние от \bar{D}_n до 0
- ②
 - Перестановочный критерий точный
 - Бутстреп-критерий приближённый

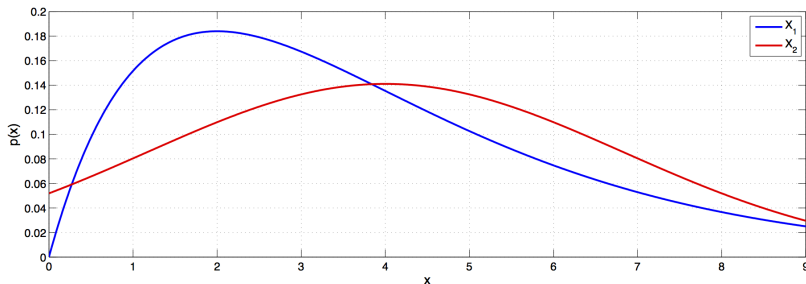
- ③
 - Перестановочный критерий проверяет

$$H_0: F_{X_1}(x) = F_{X_2}(x) \quad \text{против} \quad H_1: F_{X_1}(x) = F_{X_2}(x + \Delta), \Delta > 0$$

- Бутстреп-критерий проверяет

$$H_0: \mathbb{E}X_1 = \mathbb{E}X_2 \quad \text{против} \quad H_1: \mathbb{E}X_1 > \mathbb{E}X_2$$

Различия между моментами высокого порядка



$$X_1 \sim \chi_4^2, \quad X_2 \sim N(4, \sqrt{8});$$

$$\mathbb{E}X_1 = \mathbb{E}X_2, \quad \mathbb{D}X_1 = \mathbb{D}X_2.$$

Двухвыборочные критерии согласия

выборки: $X_1^{n_1} = (X_{11}, \dots, X_{1n_1})$

$X_2^{n_2} = (X_{21}, \dots, X_{2n_2})$

выборки независимые

нулевая гипотеза: $H_0: F_{X_1}(x) = F_{X_2}(x)$

альтернатива: $H_1: H_0$ неверна

Критерий Смирнова

статистика: $D(X_1^{n_1}, X_2^{n_2}) = \sup_{-\infty < x < \infty} |F_{n_1 X_1}(x) - F_{n_2 X_2}(x)|$

Критерий Андерсона (модификация критерия Смирнова-Крамера-фон Мизеса)

статистика:
$$T(X_1^{n_1}, X_2^{n_2}) = \frac{1}{n_1 n_2 (n_1 + n_2)} \left(n_1 \sum_{i=1}^{n_1} (\text{rank}(X_{1i}) - i)^2 + \right. \\ \left. + n_2 \sum_{j=1}^{n_2} (\text{rank}(X_{2j}) - j)^2 \right) - \frac{4n_1 n_2 - 1}{6(n_1 + n_2)}$$

Статистики имеют табличные распределения при H_0 .

Литература

- критерии знаков (sign tests) — Kanji, №№ 45, 46;
- критерии знаковых рангов (signed-rank tests) — Kanji, №№ 47, 48;
- критерий Манна-Уитни-Уилкоксона (Mann-Whitney-Wilcoxon test) — Kanji, № 52;
- перестановочные критерии (permutation tests) — Good, 3.2.1, 3.6.4, 3.7.2 (с ошибкой, исправлено в Ramsey);
- двухвыборочные критерии согласия (two-sample goodness-of-fit tests) — Кобзарь, 3.1.2.8.

Кобзарь А.И. *Прикладная математическая статистика*, 2006.

Bonnini S., Corain L., Marozzi M., Salmaso S. *Nonparametric Hypothesis Testing - Rank and Permutation Methods with Applications in R*, 2014.

Dinse G.E. (1982). *Nonparametric estimation for partially-complete time and type of failure data*. Biometrics, 38, 417–431.

Good P. *Permutation, Parametric and Bootstrap Tests of Hypotheses: A Practical Guide to Resampling Methods for Testing Hypotheses*, 2005.

Литература

Hollander M., Wolfe D.A. *Nonparametric statistical methods*, 1973.

Kanji G.K. *100 statistical tests*, 2006.

Laureysens I., Blust R., De Temmerman L., Lemmens C., Ceulemans R. (2004). *Clonal variation in heavy metal accumulation and biomass production in a poplar coppice culture. I. Seasonal variation in leaf, wood and bark concentrations*. Environmental Pollution, 131, 485-494.

Ramsey P.H., Ramsey P.P. (2008). *Brief investigation of tests of variability in the two-sample case*. Journal of Statistical Computation and Simulation, 78(12), 1125-1131.

Shervin C.M. (2004) *Mirrors as potential environmental enrichment for individually housed laboratory mice*. Applied Animal Behaviour Science, 87(1-2), 95-103.