Прикладной статистический анализ данных.
3. Непараметрическая проверка гипотез.

Рябенко Евгений
riabenko.e@gmail.com

I/2016

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:

$$H_0\colon \mathbb{E}X_1 = \mathbb{E}X_2, \qquad H_1\colon \mathbb{E}X_1 <\neq> \mathbb{E}X_2;$$
$$H_0\colon \operatorname{med} X_1 = \operatorname{med} X_2, \qquad H_1\colon \operatorname{med} X_1 <\neq> \operatorname{med} X_2;$$
$$H_0\colon p\left(X_1 > X_2\right) = 0.5, \qquad H_1\colon p\left(X_1 > X_2\right) <\neq> 0.5;$$
$$H_0\colon F_{X_1}\left(x\right) = F_{X_2}\left(x\right), \qquad H_1\colon F_{X_1}\left(x\right) = F_{X_2}\left(x+\Delta\right), \Delta <\neq> 0;$$
$$H_0\colon F_{X_1}\left(x\right) = F_{X_2}\left(x\right), \qquad H_1\colon F_{X_1}\left(x\right) <\neq> F_{X_2}\left(x\right).$$

О рассеянии:

$$H_0\colon \mathbb{D}X_1 = \mathbb{D}X_2, \qquad H_1\colon \mathbb{D}X_1 <\neq> \mathbb{D}X_2;$$
$$H_0\colon F_{X_1}\left(x\right) = F_{X_2}\left(x+\Delta\right), \qquad H_1\colon F_{X_1}\left(x\right) = F_{X_2}\left(\sigma x+\Delta\right), \sigma <\neq> 1.$$

(1) Одновыборочный критерий знаков

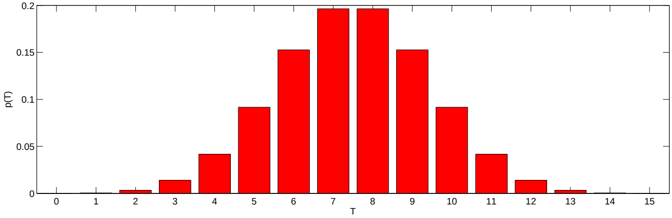
выборка: $X^n = (X_1, \dots, X_n)\,, X_i \neq m_0;$

нулевая гипотеза: $H_0\colon\ \mathrm{med}\,X = m_0;$

альтернатива: $H_1\colon\ \mathrm{med}\,X \ {<}{\neq}{>}\ m_0;$

статистика: $$T(X^n) = \begin{cases} n_1 = \sum\limits_{i=1}^{n} [X_i > m_0]\,, & H_1\colon\ \mathrm{med}\,X_i > m_0, \\ n_2 = \sum\limits_{i=1}^{n} [X_i < m_0]\,, & H_1\colon\ \mathrm{med}\,X_i < m_0, \\ \min(n_1, n_2)\,, & H_1\colon\ \mathrm{med}\,X_i \neq m_0; \end{cases}$$

$T(X^n) \sim Bin(n, \frac{1}{2})$ при H_0.



(1) Одновыборочный критерий знаков

Пример 1 (Kanji, критерий 45): предполагается, что стоимость материала, получаемого при переработке строительной конструкции, составляет в среднем 0.28 долларов. Взята случайная выборка из 10 конструкций, все они переработаны; стоимость в долларах полученного из каждой конструкции материала составила:

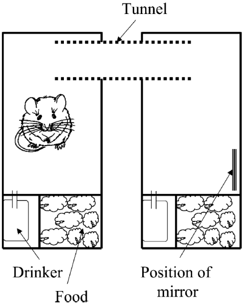
$$\{0.28, 0.18, 0.24, 0.30, 0.40, 0.36, 0.15, 0.42, 0.23, 0.48\}.$$

Правомерно ли использовать гипотезу о том, что она взята из популяции с медианой стоимости переработанного материала 0.28 долларов?

H_0: медиана стоимости переработанного материала составляет 0.28 долларов.
H_1: медиана стоимости переработанного материала отличается от 0.28 долларов $\Rightarrow p = 1$, 95% доверительный интервал для медианы — $[0.196, 0.414]$.

(1) Одновыборочный критерий знаков

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



Общая постановка:

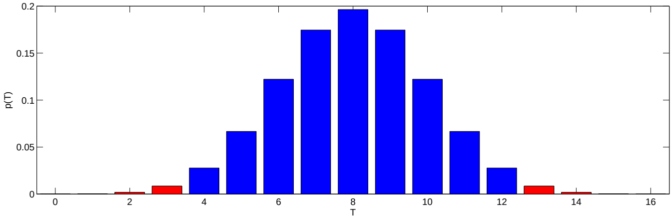
H_0: мышам всё равно, висит в клетке зеркало или нет.

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

(1) Одновыборочный критерий знаков

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 $\frac{1}{2}$.

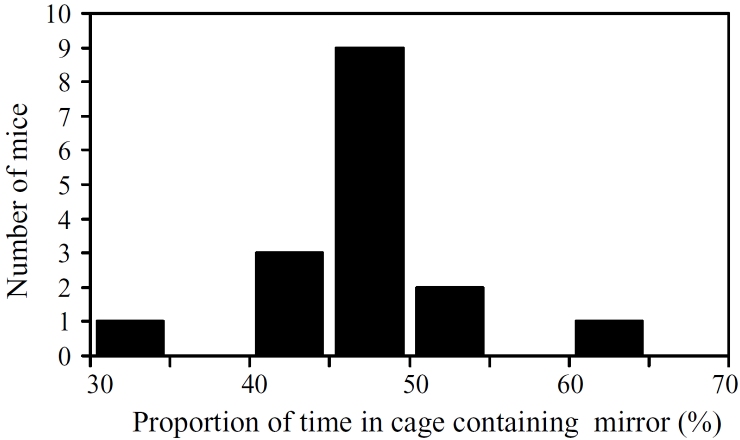
Редуцированные данные: 0 — мышь провела больше времени в комнате с зеркалом, 1 — в комнате без зеркала.
Статистика: T — число единиц в выборке.



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

Критерий знаков: $p = 0.0213$, 95% доверительный интервал для вероятности (что мышь проведёт больше времени в комнате без зеркала) — $[0.54, 0.96]$.

(1) Одновыборочный критерий знаков



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

(2) Двухвыборочный критерий знаков

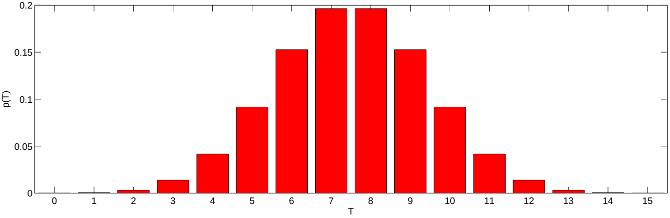
выборки: $X_1^n = (X_{11}, \ldots, X_{1n})$,
$X_2^n = (X_{21}, \ldots, X_{2n})$, $X_{1i} \neq X_{2i}$,
выборки связанные;

нулевая гипотеза: $H_0\colon p(X_1 > X_2) = \frac{1}{2}$;

альтернатива: $H_1\colon p(X_1 > X_2) <\neq> \frac{1}{2}$;

статистика:
$$T(X_1^n, X_2^n) = \begin{cases} n_1 = \sum\limits_{i=1}^n [X_{1i} > X_{2i}], & H_1\colon >, \\ n_2 = \sum\limits_{i=1}^n [X_{1i} < X_{2i}], & H_1\colon <, \\ \min(n_1, n_2), & H_1\colon \neq; \end{cases}$$

$T(X_1^n, X_2^n) \sim Bin(n, \frac{1}{2})$ при H_0.



(2) Двухвыборочный критерий знаков

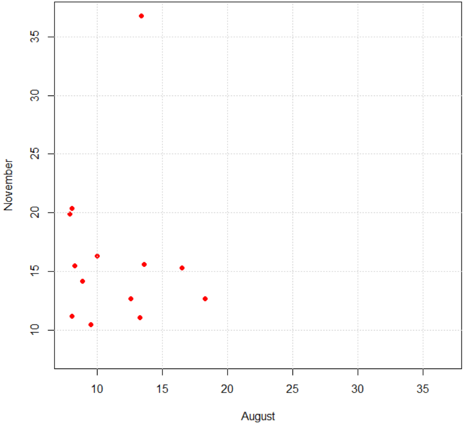
Пример 1 (Hollander & Wolfie, 29f): депрессивность 9 пациентов была измерена по шкале Гамильтона до и после первого приёма транквилизатора. Подействовал ли транквилизатор?

H_0: уровень депрессивности не изменился.
H_1: уровень депрессивности снизился $\Rightarrow p = 0.09$, 95% односторонний доверительный интервал для медианы изменения:

- $[-0.041, \infty]$ — сглаженный,
- $[-0.08, \infty]$ — консервативный (с уровнем доверия 98.05%).

(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 но знаку (влияние меди на число личинок комаров).

Вариационный ряд, ранги, связки

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

Ранг наблюдения X_i:

$$\operatorname{rank}(X_i) = \mathbb{E}\left\{r \,\middle|\, X_i = X_{(r)}\right\}$$

т. е. если X_i не в связке, то ранг — номер X_i в вариационном ряду,
если X_i в связке $X_{(k_1)}, \ldots, X_{(k_2)}$, то $\operatorname{rank}(X_i) = \frac{k_1+k_2}{2}$.

(3) Одновыборочный критерий знаковых рангов Уилкоксона

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

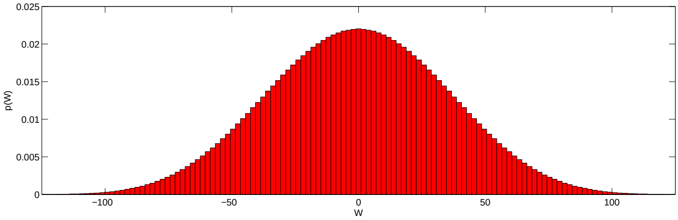
$F(X)$ симметрично относительно медианы;

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

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

статистика: $W(X^n) = \sum\limits_{i=1}^{n} \operatorname{rank}(|X_i - m_0|) \cdot \operatorname{sign}(X_i - m_0)\,;$

$W(X^n)$ имеет табличное распределение при H_0.



(3) Одновыборочный критерий знаковых рангов Уилкоксона

Откуда берётся табличное распределение?

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	W
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-120
+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-118
-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-116
+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-114
-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-114
+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-112
-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-110
+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-108
-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-112
+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-110
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	110
+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	112
-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	108
+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	110
-	+	-	+	+	+	+	+	+	+	+	+	+	+	+	112
+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	114
-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	114
+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	116
-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	118
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	120

(3) Одновыборочный критерий знаковых рангов Уилкоксона

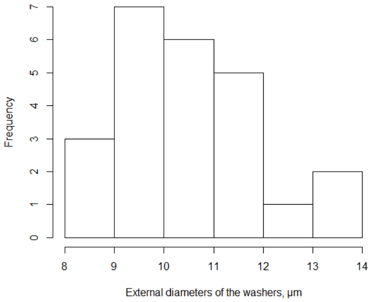
Аппроксимация для $n > 20$:

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

Обработка связок: зависит от реализации.

(3) Одновыборочный критерий знаковых рангов Уилкоксона

Пример 1 (Bonnini, табл. 1.4): целевой внешний диаметр втулки — 10 мкм. Служба контроля качества производства измерила его для 24 втулок и получила следующее распределение:



$H_0:$ средний внешний диаметр втулок соответствует норме.
$H_1:$ средний внешний диаметр втулок не соответствует норме
$\Rightarrow p = 0.0673$, 95% доверительный интервал для медианы — $[9.95, 11.15]$.

(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 $\frac{1}{2}$ $\Rightarrow p = 0.0703$.

(4) Критерий знаковых рангов Уилкоксона для связанных выборок

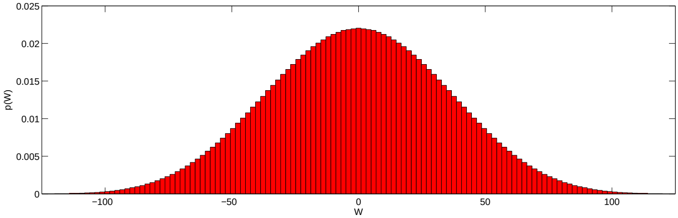
выборки: $X_1^n = (X_{11}, \ldots, X_{1n})$,
$X_2^n = (X_{21}, \ldots, X_{2n})$, $X_{1i} \neq X_{2i}$,
выборки связанные;

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

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

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

$W(X_1^n, X_2^n)$ имеет табличное распределение при H_0.



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

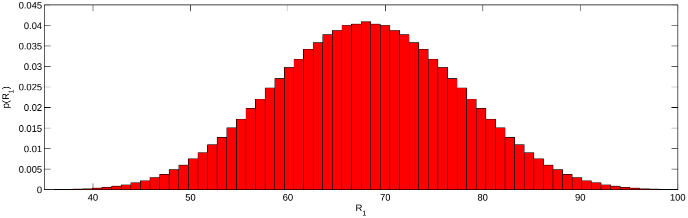
нулевая гипотеза: $H_0\colon F_{X_1}(x) = F_{X_2}(x)$;

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

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

$$R_1\left(X_1^{n_1}, X_2^{n_2}\right) = \sum_{i=1}^{n_1} \operatorname{rank}(X_{1i});$$

$R_1\left(X_1^{n_1}, X_2^{n_2}\right)$ имеет табличное распределение при H_0.



(5) Критерий Манна-Уитни-Уилкоксона

Откуда берётся табличное распределение?

Первая выборка	Вторая выборка	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
{1,3,5}	{2,4,6,7}	9
{1,3,6}	{2,4,5,7}	10
{1,3,7}	{2,4,5,6}	11
{1,4,5}	{2,3,6,7}	6
...	...	...
{3,4,5}	{1,2,6,7}	12
{3,4,6}	{1,2,5,7}	13
{3,4,7}	{1,2,5,6}	14
{3,5,6}	{1,2,4,7}	14
{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

(5) Критерий Манна-Уитни-Уилкоксона

Аппроксимация для $n_1, n_2 > 10$:

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

Обработка связок: зависит от реализации.

(5) Критерий Манна-Уитни-Уилкоксона

Пример 1 (Kanji, 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).

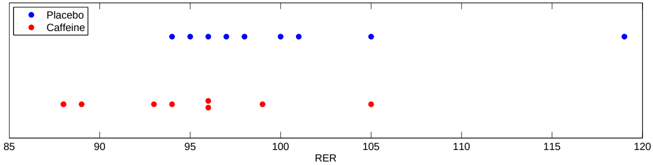
$$X_1\colon \{50.5, 37.5, 49.8, 56.0, 42.0, 56.0, 50.0, 54.0, 48.0\},$$
$$X_2\colon \{57.0, 52.0, 51.0, 44.2, 55.0, 62.0, 59.0, 45.2, 53.5, 44.4\}.$$

Равны ли средние расходы?

$H_0\colon$ средние расходы равны.
$H_1\colon$ средние расходы не равны $\Rightarrow p = 0.3072$, 95% доверительный интервал для медианной разности — $[-9, 4]$.

(5) Критерий Манна-Уитни-Уилкоксона

Пример 2: RER (респираторный обмен)— соотношение числа молекул CO_2 и O_2 в выдыхаемом воздухе. Является косвенным признаком того, из жиров или углеводов вырабатывается энергия в момент измерения. Изучалось влияние кофеина на мышечный метаболизм. В эксперименте принимало участие 18 испытуемых, респираторный обмен которых измерялся в процессе физических упражнений. За час до этого 9 из них получили таблетку кофеина, оставшиеся 9 — плацебо.
Повлиял ли кофеин на значение показателя респираторного обмена?



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.

$p = 0.0521$, 95% доверительный интервал для медианной разности — $[-0.00005, 1.2]$.

(6) Критерий Зигеля-Тьюки

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

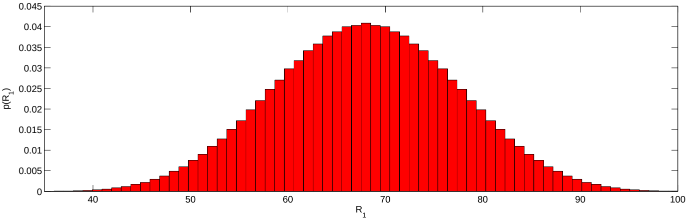
нулевая гипотеза: $H_0 \colon \mathbb{D}X_1 = \mathbb{D}X_2$;

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

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

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

$R_1\left(X_1^{n_1}, X_2^{n_2}\right)$ имеет табличное распределение при H_0.



Ранги присваиваются необычным образом:

$X_{(i)}$	$X_{(1)}$	$\leq$	$X_{(2)}$	$\leq$	$X_{(3)}$	$\leq \ldots \leq$	$X_{(N-2)}$	$\leq$	$X_{(N-1)}$	$\leq$	$X_{(N)}$
$\widetilde{\operatorname{rank}}\left(X_{(i)}\right)$	1		4		5		6		3		2

(6) Критерий Зигеля-Тьюки

Пример (Kanji, критерий 53): менеджер по кейтерингу хочет проверить, одинакова ли дисперсия количества соуса в упаковке при расфасовке с помощью двух диспенсеров. Каждым из диспенсеров он наполнил 10 упаковок.

$$X_1: \{2.4, 6.1, 7.3, 8.5, 8.8, 9.4, 9.8, 10.1, 10.1, 12.6\},$$
$$X_2: \{2.9, 3.3, 3.6, 4.2, 4.9, 7.3, 11.7, 13.1, 15.3, 16.5\}.$$

Предполагается, что оба диспенсера хорошо откалиброваны, то есть, дают одинаковое среднее количество соуса в упаковке.

H_0: дисперсия количества соуса в упаковке не отличается для двух диспенсеров.
H_1: дисперсия количества соуса в упаковке для двух диспенсеров отличается $\Rightarrow p = 0.0288$, 95% доверительный интервал для отношения дисперсий — $[-11, -1.00003]$.

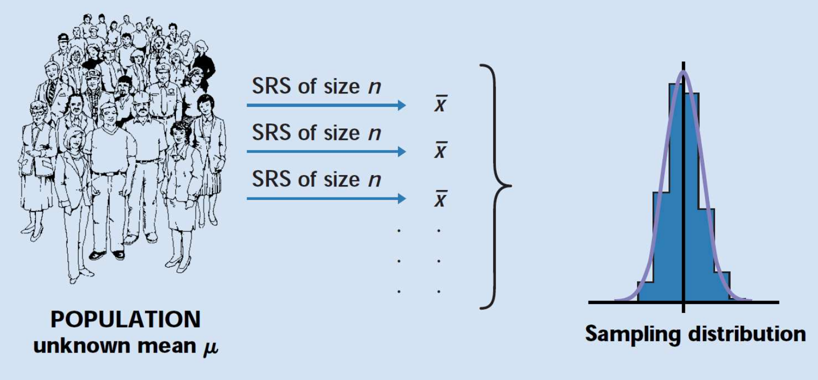
(6) Критерий Зигеля-Тьюки

В стандартных реализациях критерия Зигеля-Тьюки в R часто бывает ошибка.
Верная реализация: `https://yadi.sk/d/iilLwa1NerKeU`

Построение доверительных интервалов

Чтобы построить доверительный интервал для статистики $T_n = T(X^n)$, нужно знать её выборочное распределение $F_{T_n}(x)$. Как его оценить? (Hesterberg, 2005):

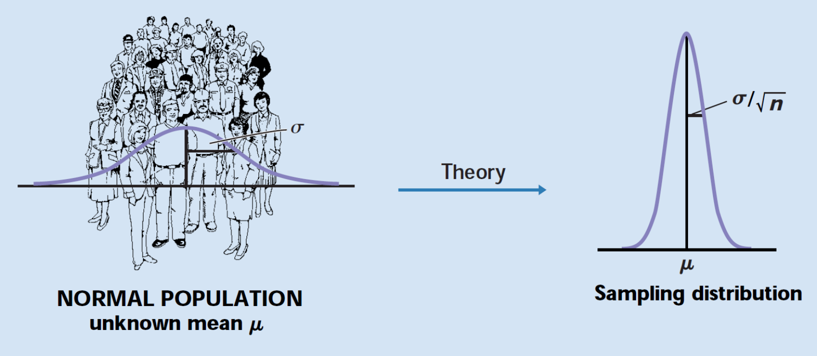
- наивный метод



Извлечь из генеральной совокупности N выборок объёма n и оценить выборочное распределение T_n эмпирическим.

Построение доверительных интервалов

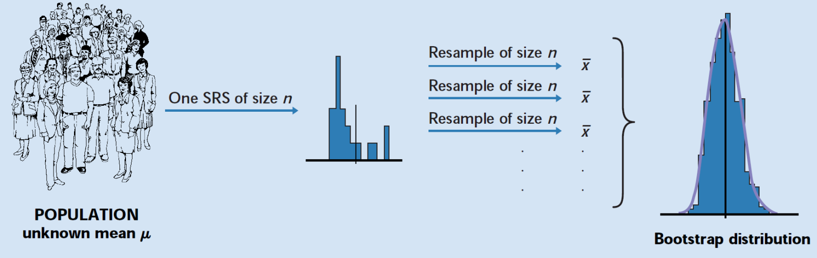
- параметрический метод



Сделать предположение, что X распределена по закону $F_X(x)$, при выполнении которого закон распределения T_n известен.

Построение доверительных интервалов

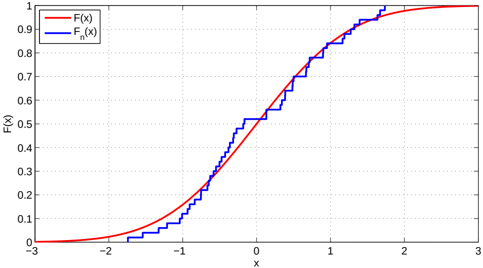
- бутстреп



Сгенерировать N «псевдовыборок» объёма n и оценить выборочное распределение T_n «псевдоэмпирическим».

Бутстреп

Извлечение выборок из генеральной совокупности — сэмплирование из неизвестного распределения $F_X(x)$. Лучшая оценка $F_X(x)$, которая у нас есть — $F_{X_n}(x)$; будем сэмплировать из неё. Это то же самое, что делать из X^n выборки с возвращением объёма n.



Способы построения доверительных интервалов с помощью бутстрепа:

- нормальный: $t \pm t_{1-\frac{\alpha}{2}} S_b$;
- базовый: границы — $\frac{\alpha}{2}$ и $1-\frac{\alpha}{2}$-квантили распределения значений T_n на псевдовыборках;
- ...

Перестановочные критерии

С помощью перестановок можно не только строить доверительные интервалы, но и проверять гипотезы.
Но для проверки гипотез нужно оценить не выборочное распределение статистики, а нулевое, то есть распределение статистики при справедливости нулевой гипотезы.

Идея: найти такую группу перестановок G исходной выборки X^n, что распределение X^n при справедливости нулевой гипотезы не отличается от распределения $gX^n, g \in G$.

(8) Одновыборочный перестановочный критерий, гипотеза о среднем

выборка: $X_1^n = (X_1, \ldots, X_n)$,
$F(X)$ симметрично относительно матожидания;
нулевая гипотеза: $H_0\colon \mathbb{E}X = 0$;
альтернатива: $H_1\colon \mathbb{E}X <\neq> 0$;
статистика: $T(X^n) = \sum\limits_{i=1}^{n} X_i$.

Распределение $T(X^n)$ при H_0 порождается группой перестановок

$$G = \{g = (s_1, \ldots, s_n)\}, s_i \in \{-1, 1\},$$

$$|G| = 2^n.$$

Для проверки гипотезы $H_0\colon \mathbb{E}X = \mu_0$ группа строится по аналогии.

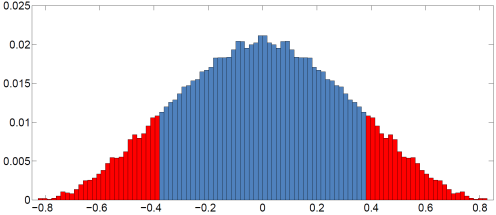
Достигаемый уровень значимости:

$$p(t) = \begin{cases} \frac{\sum\limits_{g\in G}[t(gx^n) \leq \geq t(x^n)]}{2^n}, & H_1\colon \mathbb{E}X <> 0, \\ \frac{\sum\limits_{g\in G}[|t(gx^n)| \geq |t(x^n)|]}{2^n}, & H_1\colon \mathbb{E}X \neq 0. \end{cases}$$

(8) Одновыборочный перестановочный критерий, гипотеза о среднем

Пример (зеркала в клетках мышей):
H_0: в клетке с зеркалом мыши проводят в среднем половину времени.
H_1: в клетке с зеркалом мыши проводят в среднем не половину времени.

Статистика: $T = \sum\limits_{i=1}^{n} (X_i - 0.5)\,; \quad t = -0.3784.$



$p = \frac{\#[|T| \geq |t|]}{2^n} = 0.2292.$
95% доверительный интервал для доли времени в клетке с зеркалом (базовый бутстреп) — $[0.444, 0.506]$.

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

выборки: $X_1^n = \left(X_{11}, \dots, X_{1n}\right)$,
$X_2^n = \left(X_{21}, \dots, X_{2n}\right)$,
выборки связанные;

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

альтернатива: $H_1\colon F_{X_1}(x) <\neq> F_{X_2}(x)$;

статистика: $D^n = \left(X_{1i} - X_{2i}\right)$,

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

Распределение $T\left(D^n\right)$ при H_0 порождается группой перестановок

$$G = \left\{g = \left(s_1, \dots, s_n\right)\right\}, s_i \in \{-1, 1\},$$

$$|G| = 2^n.$$

Достигаемый уровень значимости:

$$p\left(t\right) = \begin{cases} \dfrac{\sum\limits_{g \in G} \left[t(gd^n) \leq\geq t(d^n)\right]}{2^n}, & H_1\colon F_{X_1}(x) <> F_{X_2}(x), \\ \dfrac{\sum\limits_{g \in G} \left[|t(gd^n)| \geq |t(d^n)|\right]}{2^n}, & H_1\colon F_{X_1}(x) \neq F_{X_2}(x). \end{cases}$$

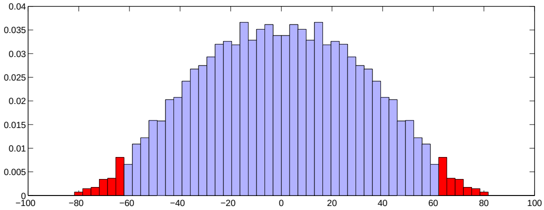
(9) Двухвыборочный перестановочный критерий, гипотеза о средних, связанные выборки

Пример (алюминий в тополях):

H_0: среднее изменения концентрации алюминия равна нулю.

H_1: среднее изменения концентрации алюминия не равна нулю.

Статистика: $T = \sum\limits_{i=1}^{n} (X_{1i} - X_{2i})\,; \quad t = -63.7.$



$p = \frac{\#[|T| \geq |t|]}{2^n} = 0.0054.$

95% доверительный интервал для средней разности (базовый бутстреп) — $[0.692, 8.577]$.

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

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

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

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

статистика: $T\left(X_1^{n_1}, X_2^{n_2}\right) = \frac{1}{n_1}\sum\limits_{i=1}^{n_1} X_{1i} - \frac{1}{n_2}\sum\limits_{i=1}^{n_2} X_{2i}$.

Распределение $T\left(X_1^{n_1}, X_2^{n_2}\right)$ при H_0 порождается группой перестановок G :

$$g\left(X_1^{n_1}, X_2^{n_2}\right) = \left(X_{\pi_{11}}, \ldots, X_{\pi_{1n_1}}, X_{\pi_{21}}, \ldots, X_{\pi_{2n_2}}\right) = \left(gX_1^{n_1}, gX_2^{n_2}\right),$$

где $\pi_{11}, \ldots, \pi_{1n_1}$ — сочетание из $N = n_1 + n_2$ по n_1, $\pi_{21}, \ldots, \pi_{2n_2}$ — его дополнение до множества $\{1, \ldots, N\}$.

$$|G| = C_N^{n_1} = C_N^{n_2}.$$

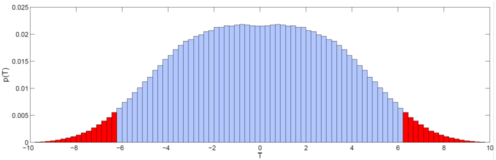
Достигаемый уровень значимости:

$$p(t) = \begin{cases} \dfrac{\sum\limits_{g\in G}\left[t\left(gx_1^{n_1}, gx_2^{n_2}\right) \leq\geq t\left(x_1^{n_1}, x_2^{n_2}\right)\right]}{C_N^{n_1}}, & H_1\colon \Delta <> 0, \\ \dfrac{\sum\limits_{g\in G}\left[\left|t\left(gx_1^{n_1}, gx_2^{n_2}\right)\right| \geq \left|t\left(x_1^{n_1}, x_2^{n_2}\right)\right|\right]}{C_N^{n_1}}, & H_1\colon \Delta \neq 0. \end{cases}$$

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

Пример (кофеин и респираторный обмен):
H_0: среднее значение показателя респираторного обмена не отличается в двух группах.
H_1: среднее значение показателя респираторного обмена отличается в двух группах.

Статистика: $T = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i} - \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2i}; \quad t = 6.33.$



$p = \frac{\#\left[\left|T-\overline{T}\right|\geq\left|t-\overline{T}\right|\right]}{C_{n_1+n_2}^{n_1}} = 0.0578.$
95% доверительный интервал для сдвига (базовый бутстреп) — $[0.111, 11.889]$.

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

выборки: $X_1^n = (X_{11}, \ldots, X_{1n})$,
$X_2^n = (X_{21}, \ldots, X_{2n})$,
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: $\delta\left(D_1^{n-1}\right) = \sum\limits_{i=1}^{n-1} i(n-i)D_{1i}$,
$D_{1i} = X_{1(i+1)} - X_{1(i)}$.

Распределение $\delta\left(D_1^{n-1}\right)$ при H_0 порождается группой G попарных перестановок D_{1i} и D_{2i}:

$$g\left(D_1^{n-1}, D_2^{n-1}\right) = \left(D_{\pi_1 1}, \ldots, D_{\pi_{n-1}(n-1)}, D_{\pi'_1 1}, \ldots, D_{\pi'_{n-1}(n-1)}\right) = \left(gD_1^{n-1}, gD_2^{n-1}\right),$$

где $\forall i = 1, \ldots, n-1$ либо $\pi_i = 1, \pi'_i = 2$, либо $\pi_i = 2, \pi'_i = 1$.

$$|G| = 2^{n-1}.$$

Достигаемый уровень значимости:

$$p(\delta) = \begin{cases} p_1 = \frac{\sum\limits_{g\in G}\left[\delta\left(gd_1^{n-1}\right) \geq \delta\left(d_1^{n-1}\right)\right]}{2^{n-1}}, & H_1 : \mathbb{D}X_1 > \mathbb{D}X_2, \\ p_2 = \frac{\sum\limits_{g\in G}\left[\delta\left(gd_1^{n-1}\right) \leq \delta\left(d_1^{n-1}\right)\right]}{2^{n-1}}, & H_1 : \mathbb{D}X_1 < \mathbb{D}X_2, \\ 2\cdot\min(p_1, p_2), & H_1 : \mathbb{D}X_1 \neq \mathbb{D}X_2. \end{cases}$$

Особенности перестановочных критериев

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

$$X^n, \;\; H_0\colon \mathbb{E}X = 0, \; H_1\colon \mathbb{E}X \neq 0,$$

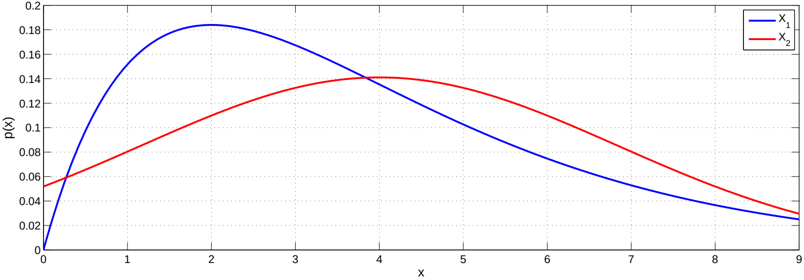
$$T_1\left(X^n\right) = \sum_{i=1}^{n} X_i \;\; \sim \;\; T_2\left(X^n\right) = \bar{X}.$$

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

$$T_2\left(X^n\right) = \bar{X} \;\; \nsim \;\; T_3\left(X^n\right) = \frac{\bar{X}}{S/\sqrt{n}}.$$

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

Различия между моментами высокого порядка



$X_1 \sim \chi^2_4, \;\; X_2 \sim N\left(4, \sqrt{8}\right);$
$\mathbb{E}X_1 = \mathbb{E}X_2, \;\; \mathbb{D}X_1 = \mathbb{D}X_2.$

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: $H_0\colon F_{X_1}(x) = F_{X_2}(x)$;
альтернатива: $H_1\colon H_0$ неверна.

Критерий Смирнова:
статистика: $D(X_1^{n_1}, X_2^{n_2}) = \sup\limits_{-\infty<x<\infty} |F_{n_1X_1}(x) - F_{n_2X_2}(x)|$.

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

$$T(X_1^{n_1}, X_2^{n_2}) = \frac{1}{n_1n_2(n_1+n_2)}\left(n_1\sum_{i=1}^{n_1}(\mathrm{rank}(X_{1i}) - i)^2 + \right.$$
$$\left. + n_2\sum_{j=1}^{n_1}(\mathrm{rank}(X_{2j}) - j)^2\right) - \frac{4n_1n_2-1}{6(n_1+n_2)}.$$

Статистики имеют табличные распределения при H_0.

Примеры

Продажная стоимость недвижимости в Сиэтле (независимые выборки):
`https://yadi.sk/d/rGiG2AkHerKbd`

Поведенческая терапия при анорексии (связные выборки):
`https://yadi.sk/d/CLSiAkGderUYR`

Качество регрессионных моделей (для самостоятельной работы):
`https://yadi.sk/i/JOHDkSSFexkkZ`

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) — Кобзарь, 4.2.1.1.2.2;
- бутстреп (bootstrap) — Hesterberg (попроще), Davidson (посложнее);
- 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*. — М.: Физматлит, 2006.

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

Davison A.C., Hinkley D.V. *Bootstrap Methods and their Application*. — Cambridge University Press, 1997.

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

Литература

Hesterberg T., Monaghan S., Moore D.S., Clipson A., Epstein R. (2005). *Bootstrap methods and permutation tests*. In Introduction to the Practice of Statistics. `http://statweb.stanford.edu/~tibs/stat315a/Supplements/bootstrap.pdf`

Hollander M., Wolfe D.A. *Nonparametric statistical methods*. — John Wiley & Sons, 1973.

Kanji G.K. *100 statistical tests*. — London: SAGE Publications, 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.