Прикладная статистика 3. Непараметрическая проверка гипотез.

Рябенко Евгений
riabenko.e@gmail.com

24 февраля 2014 г.

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 7 11

Варианты двухвыборочных гипотез

О положении:

$H_0\colon \operatorname{med} X_1 = \operatorname{med} X_2,$ $H_1\colon \operatorname{med} X_1 <\neq> \operatorname{med} X_2;$

$H_0\colon \mathbb{E}X_1 = \mathbb{E}X_2,$ $H_1\colon \mathbb{E}X_1 <\neq> \mathbb{E}X_2;$

$H_0\colon p\left(X_1 > X_2\right) = \frac{1}{2},$ $H_1\colon p\left(X_1 > X_2\right) <\neq> \frac{1}{2};$

$H_0\colon F_{X_1}(x) = F_{X_2}(x),$ $H_1\colon F_{X_1}(x) <\neq> F_{X_2}(x);$

$H_0\colon F_{X_1}(x) = F_{X_2}(x),$ $H_1\colon F_{X_1}(x) = F_{X_2}(x+\Delta), \Delta <\neq> 0.$

О рассеянии:

$H_0\colon \mathbb{D}X_1 = \mathbb{D}X_2,$ $H_1\colon \mathbb{D}X_1 <\neq> \mathbb{D}X_2;$

$H_0\colon \mathbb{D}X_1 = \mathbb{D}X_2, \operatorname{med} X_1 = \operatorname{med} X_2,$ $H_1\colon \mathbb{D}X_1 <\neq> \mathbb{D}X_2;$

$H_0\colon F_{X_1}(x) = F_{X_2}(x+\Delta),$ $H_1\colon F_{X_1}(x) = F_{X_2}(\sigma x+\Delta), \sigma <\neq> 1.$

(1) Одновыборочный критерий знаков

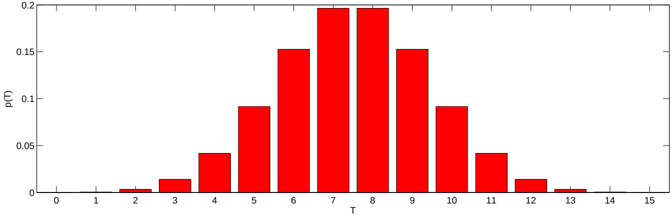
выборка: $X^n = (X_1, \dots, X_n), X_i \neq m_0;$

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

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

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

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



достигаемый уровень значимости:

$$p(t) = \begin{cases} binocdf(t, n, 0.5), & H_0\colon \text{ med } X_i <> m_0, \\ \min(1, 2 * binocdf(t, n, 0.5)), & H_0\colon \text{ med } X_i \neq m_0. \end{cases}$$

(1) Одновыборочный критерий знаков

Пример: 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 $\Rightarrow p \approx 1.$

(2) Двухвыборочный критерий знаков

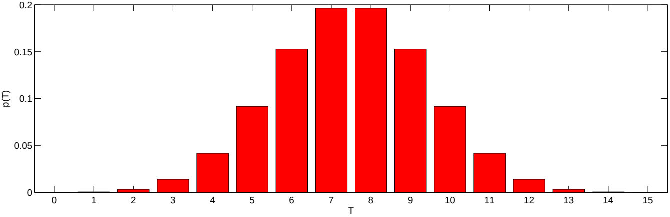
выборки: $X_1^n = (X_{11}, \dots, X_{1n})$,
$X_2^n = (X_{21}, \dots, X_{2n})$, $X_{1i} \neq X_{2i}$, выборки связанные;

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

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

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

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



достигаемый уровень значимости:

$$p(t) = \begin{cases} binocdf(t, n, 0.5), & H_0\colon p(X_{1i} > X_{2i}) <> \frac{1}{2}, \\ \min(1, 2 * binocdf(t, n, 0.5)), & H_0\colon p(X_{1i} > X_{2i}) \neq \frac{1}{2}. \end{cases}$$

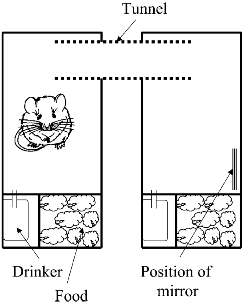
(2) Двухвыборочный критерий знаков

Пример: показатель качества работы 10 машин производственного цикла был измерен до и после модификации производства. Для трёх машин значение показателя упало, для семи — повысилось. В среднем влияет ли модификация на значение показателя качества?

H_0: модификация в среднем не влияет на значение показателя качества работы машин.
H_1: модификация в среднем влияет на значение показателя качества работы машин $\Rightarrow p = 0.3438$.

Зеркала в клетках мышей

Shervin, Mirrors as potential environmental enrichment for individually housed laboratory mice, 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.

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 $\frac{1}{2}$.

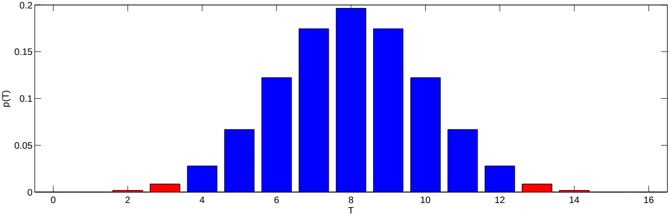
Редуцированные данные: 0 — мышь провела больше времени в комнате с зеркалом, 1 — в комнате без зеркала.

0000000000000000	0100000000000000	1000000000000000	1100000000000000
0000000000000001	0100000000000001	1000000000000001	1100000000000001
0000000000000010	0100000000000010	1000000000000010	1100000000000010
0000000000000011	0100000000000011	1000000000000011	1100000000000011
0000000000000100	0100000000000100	1000000000000100	1100000000000100
0000000000000101	0100000000000101	1000000000000101	1100000000000101
0000000000000110	0100000000000110	1000000000000110	1100000000000110
...	...	...	...

Используем критерий знаков.

Зеркала в клетках мышей

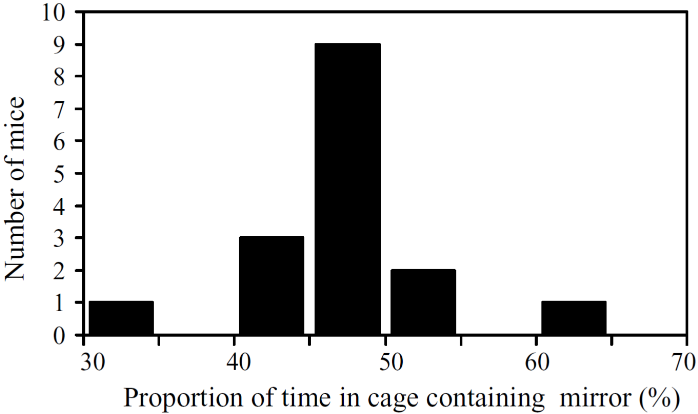
Статистика: T — число единиц в выборке.



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

Критерий знаков: $p = 0.0213$.

Зеркала в клетках мышей



Гистограмма распределения доли времени, проводимого в клетке с зеркалом.

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

Причины использовать критерий знаков

- 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).
- Точные разности Δx_i неизвестны, известны только их знаки (сравнение агрессивности 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 измерения X_i:

$$r_i = r\left(X_i\right) = mean\left\{r \,\middle|\, X_i = X_{(r)}\right\}$$

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 ранг $r_i = \frac{k_1 + k_2}{2}$.

(3) Одновыборочный критерий знаковых рангов Уилкоксона

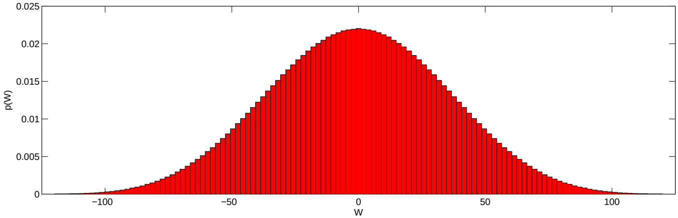
выборка: $X^n = (X_1, \dots, X_n), X_i \neq m_0;$

нулевая гипотеза: $H_0\colon \operatorname{med} X_i = m_0;$

альтернатива: $H_1\colon \operatorname{med} X_i <\neq> m_0;$

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

$W(X^n)$ 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: значение депозитной ставки измеряется на выборке из 10 инвесторов после рекламной кампании; среднее значение ставки до начала кампании — 0.28. Значения депозитной ставки после кампании:

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

Изменилось ли среднее значение депозитной ставки?

$H_0:$ среднее значение депозитной ставки после кампании осталось прежним.
$H_1:$ среднее значение депозитной ставки после кампании изменилось
$\Rightarrow p = 0.5536.$

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: $p = 0.0703$.

(4) Критерий знаковых рангов Уилкоксона для связанных выборок

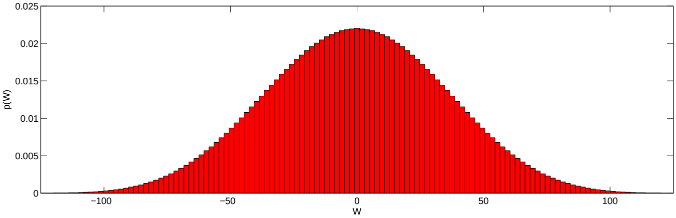
выборки: $X_1^n = (X_{11}, \dots, X_{1n})$,
$X_2^n = (X_{21}, \dots, X_{2n})$, $X_{1i} \neq X_{2i}$, выборки связанные;

нулевая гипотеза: $H_0\colon \operatorname{med}(X_{1i} - X_{2i}) = 0$;

альтернатива: $H_1\colon \operatorname{med}(X_{1i} - X_{2i}) <\neq> 0$;

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

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



(4) Критерий знаковых рангов Уилкоксона для связанных выборок

Пример: 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 $\Rightarrow p = 0.0142$.

(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;

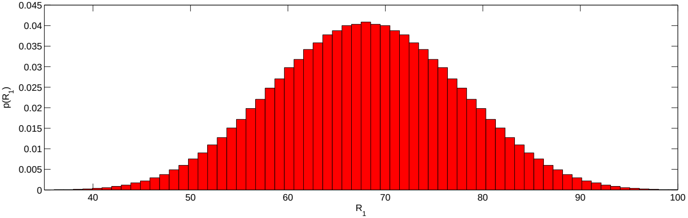
нулевая гипотеза: $H_0 \colon p(X_{1i} > X_{2j}) = \frac{1}{2}$;

альтернатива: $H_1 \colon p(X_{1i} > X_{2j}) <\neq> \frac{1}{2}$;

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

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

$R_1\left(X_1^{n_1}, X_2^{n_2}\right)$ 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}	8
{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.

$$p\left(X_{1i} > X_{2j}\right) = \frac{1}{2} \Rightarrow \operatorname{med}\left(X_{1i} - X_{2j}\right) = 0 \not\Rightarrow \operatorname{med} X_{1i} = \operatorname{med} X_{2j}.$$

(5) Критерий Манна-Уитни-Уилкоксона

Пример: 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?

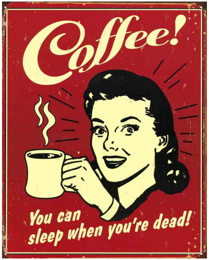
$H_0\colon$ средние расходы равны.
$H_1\colon$ средние расходы не равны $\Rightarrow p = 0.3072.$

Кофеин и респираторный обмен

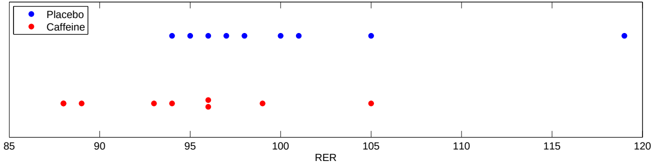
RER (респираторный обмен)— соотношение числа молекул CO_2 и O_2 в выдыхаемом воздухе. Является косвенным признаком того, из жиров или углеводов вырабатывается энергия в момент измерения.

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?



Кофеин и респираторный обмен



Значение показателя респираторного обмена в двух группах.

H_0: среднее значение показателя респираторного обмена не отличается в двух группах.
H_1: среднее значение показателя респираторного обмена отличается в двух группах.

Кофеин и респираторный обмен

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.

R: $p = 0.0521$.
Matlab: $p = 0.0468$.
Matlab при исключении наблюдения RER=119: $p = 0.0865$.

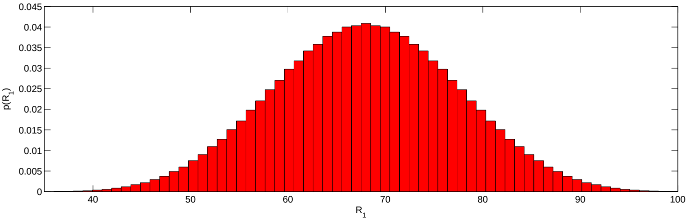
(6) Критерий Зигеля-Тьюки

выборки: $X_1^{n_1} = (X_{11}, \dots, X_{1n_1})$,
$X_2^{n_2} = (X_{21}, \dots, X_{2n_2})$, выборки независимые;
$\operatorname{med} X_{1i} = \operatorname{med} X_{2j}$;

нулевая гипотеза: $H_0\colon \mathbb{D}X_{1i} = \mathbb{D}X_{2j}$;

альтернатива: $H_1\colon \mathbb{D}X_{1i} <\neq> \mathbb{D}X_{2j}$;

статистика: $X_{(1)} \leq \ldots \leq X_{(N)}$ — вариационный ряд объединённой выборки $X^N = X_1^{n_1} \bigcup X_2^{n_2}, N = n_1 + n_2$,
$$R_1\left(X_1^{n_1}, X_2^{n_2}\right) = \sum_{i=1}^{n_1} \tilde{r}\left(X_{1i}\right);$$
$R_1\left(X_1^{n_1}, X_2^{n_2}\right)$ имеет табличное распределение при H_0.



Ранги присваиваются необычным образом:

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

(6) Критерий Зигеля-Тьюки

Пример: 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 $\Rightarrow p = 0.0288.$

(7) WM-критерий

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

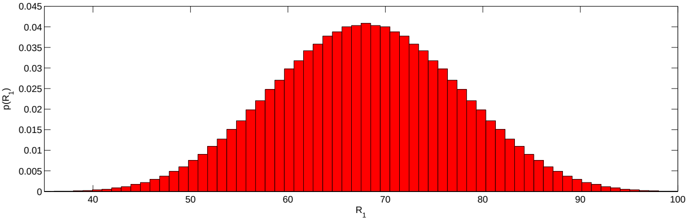
нулевая гипотеза: $H_0\colon \mathbb{D}X_{1i} = \mathbb{D}X_{2j}$;

альтернатива: $H_1\colon \mathbb{D}X_{1i} <\neq> \mathbb{D}X_{2j}$;

статистика: $D_1^{N_1} = (|X_{1i} - X_{1j}|)$, $N_1 = \lfloor \frac{n_1}{2} \rfloor$,
$D_2^{N_2} = (|X_{2i} - X_{2j}|)$, $N_2 = \lfloor \frac{n_2}{2} \rfloor$,
$D_{(1)} \leq \ldots \leq D_{(N_1+N_2)}$ — вариационный ряд объединённой выборки $D = D_1^{N_1} \bigcup D_2^{N_2}$,

$$R_1\left(X_1^{n_1}, X_2^{n_2}\right) = \sum_{i=1}^{N_1} r\left(D_{1i}\right);$$

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



(7) WM-критерий

Пример: 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. Возможно, диспенсеры откалиброваны по-разному.

$H_0:$ дисперсия количества соуса в упаковке не отличается для двух диспенсеров.
$H_1:$ дисперсия количества соуса в упаковке для двух диспенсеров отличается $\Rightarrow p = 0.2222$.

Перестановочные критерии

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

Например, если в одновыборочной задаче справедлива гипотеза $H_0\colon \mathbb{E}X_i = 0$, то с той же вероятностью, что и X^n, могла реализоваться выборка

$$gX^n = X^n \cdot (s_1, \dots, s_n), s_i \in \{-1, 1\}.$$

Если нулевая гипотеза заключается в том, что выборки в паре $(X_1^{n_1}, X_2^{n_2}) \equiv (X_1, \dots, X_N)$ одинаково распределены, то с той же вероятностью могла реализоваться пара

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

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

(8) Одновыборочный перестановочный критерий, гипотеза о среднем

выборка: $X_1^n = (X_1, \ldots, X_n)\,, \quad F(\mathbb{E}X - X) = 1 - F(\mathbb{E}X + X)\,;$
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.

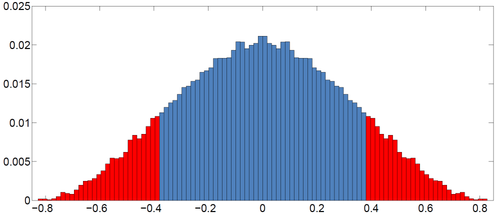
Достигаемый уровень значимости:

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

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



$p = \frac{\#[|T| \geq |t|]}{2^n}.$
$p = 0.2292.$

(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;

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

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

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

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

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

Достигаемый уровень значимости:

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

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

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

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

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

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

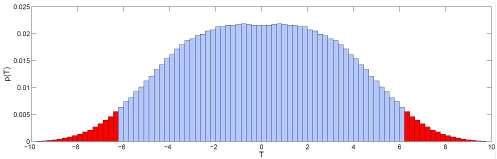
Достигаемый уровень значимости:

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

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



$$p = \frac{\#\left[\left|T-\overline{T}\right| \geq \left|t-\overline{T}\right|\right]}{C_{n_1+n_2}^{n_1}}.$$
$$p = 0.0578.$$

(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: $H_0 : \mathbb{D}X_{1i} = \mathbb{D}X_{2j}$;

альтернатива: $H_1 : \mathbb{D}X_{1i} <\neq> \mathbb{D}X_{2j}$;

статистика: $X_{1(1)} \leq \ldots \leq X_{1(n)}, \quad X_{2(1)} \leq \ldots \leq X_{2(n)}$ — вариационные ряды выборок,

$D_{1i} = X_{1(i+1)} - X_{1(i)}, \; D_{2i} = X_{2(i+1)} - X_{2(i)}$,

$$\delta\left(D_1^{n-1}\right) = \sum_{i=1}^{n-1} i(n-i)D_{1i}.$$

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:

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

Особенности перестановочных критериев

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

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

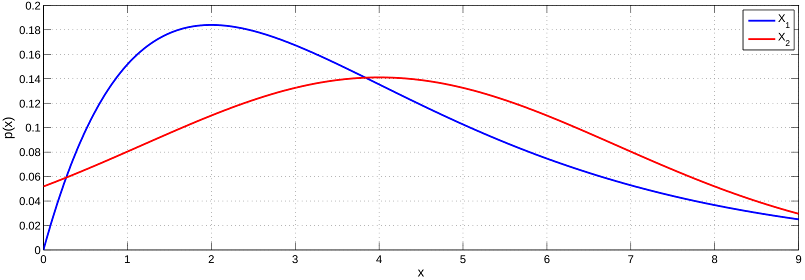
$$T_1(X^n) = \sum_{i=1}^{n} X_i \quad \sim \quad T_2(X^n) = \bar{X}.$$

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

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

- Если $|G|$ слишком велико, для оценки нулевого распределения T достаточно взять случайное подмножество $G' \in G$.

Различия между моментами высокого порядка



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

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

Критерий Колмогорова-Смирнова:
статистика: $$D\left(X_1^{n_1}, X_2^{n_2}\right) = \sup_{-\infty<x<\infty} \left|F_{n_1 X_1}(x) - F_{n_2 X_2}(x)\right|.$$

Критерий Андерсона (модификация критерия Смирнова-Крамера-фон Мизеса):
статистика: $$T\left(X_1^{n_1}, X_2^{n_2}\right) = \frac{1}{n_1 n_2 (n_1+n_2)} \left(n_1 \sum_{i=1}^{n_1} \left(r\left(X_{1i}\right) - i\right)^2 + \right.$$
$$\left. + n_2 \sum_{j=1}^{n_1} \left(r\left(X_{2j}\right) - j\right)^2 \right) - \frac{4n_1n_2-1}{6(n_1+n_2)}.$$

Статистики имеют табличные распределения при H_0.

Функция сдвига

Вместо равномерного сдвига

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

можно предположить, что сдвиг зависит от x:

$$F_{X_1}(x) = F_{X_2}(x + \Delta(x)).$$

Такая модель позволяет оценить эффект отдельно для разных групп популяции. Примеры:

- удобрение увеличивает крупные экземпляры растения и не влияет на рост мелких;
- гамма-облучение семян увеличивает вариабельность сельскохозяйственных растений.

Функция сдвига

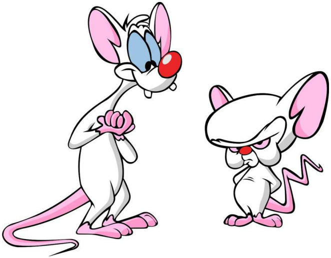
На основе инвертированного распределения двухвыборочного критерия Колмогорова-Смирнова можно построить доверительную ленту.

Реализация в R:

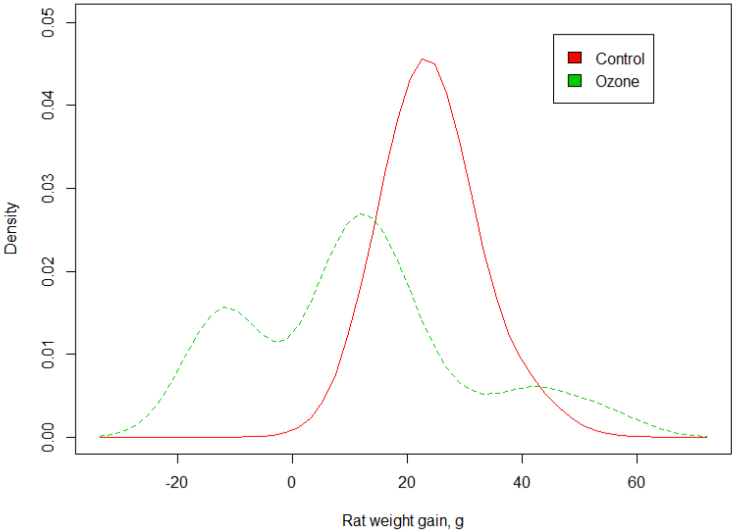
```
> install.packages("akima")
> install.packages("WRS repos="http://R-Forge.R-project.org")
> library("WRS")
> sband(x1,x2)
```

Функция сдвига

Пример: (Doksum, Sievers, Plotting with confidence: graphical comparisons of two populations, 1976) из двух групп лабораторных мышей одна живёт в обычных условиях, а вторая — в среде, обогащённой озоном. Известен прирост веса для каждой особи за фиксированный контрольный период. Есть ли различия между группами?



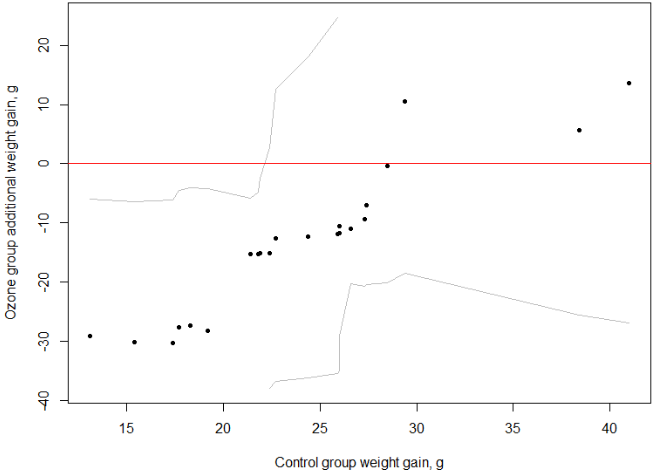
Функция сдвига



Функция сдвига

Критерий Уэлша: $p = 0.005$, критерий Манна-Уитни: $p = 0.001$.

Функция сдвига:



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;
- 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;
- функция сдвига (shift function) — Wilcox, 5.1.

Кобзарь А.И. *Прикладная математическая статистика*. — М.: Физматлит, 2006.
Kanji G.K. *100 statistical tests*. — London: SAGE Publications, 2006.
Good P. *Permutation, Parametric and Bootstrap Tests of Hypotheses: A Practical Guide to Resampling Methods for Testing Hypotheses*. — New York: Springer, 2005.
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.
Wilcox R.R. *Introduction to Robust Estimation and Hypothesis Testing*. — Academic Press, 2012.