Лекция 9

Некоторые непараметрические критерии

Напоминалка: связанные пары и независимые выборки

Пример 1 *(Newbold)*. A random sample of six salespersons who had attended a motivational course on sales techniques was monitored in the three months before and the three months after the course. The table shows the values of sales, in thousands of dollars, generated by these six salespersons in the two periods.

Salesperson	*Before the course*	*After the course*
1	212	237
2	282	291
3	203	191
4	327	341
5	165	192
6	198	180

Assuming that the population distributions are normal find an 80% confidence interval for the difference between two population means.

Пример 2 *(Демешев)*. Вашему вниманию представлены результаты прыжков в длину Васи Сидорова на двух соревнованиях. На первых среди болельщиц присутствовала Аня Иванова:

1.83 1.64 2.27 1.78 1.89 2.33 1.61 2.31

На вторых Аня среди болельщиц не присутствовала:

1.26 1.41 2.05 1.07 1.59 1.96 1.29 1.52 1.18 1.47

Влияет ли присутствие Ани Ивановой на результаты Васи Сидорова?

Пример 1 — связанные пары *Пример 2 — независимые выборки*

Что будет сегодня

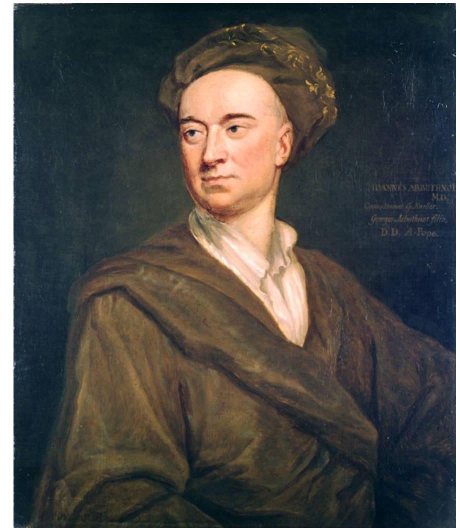
Посмотрим, как те же задачи можно решить без предпосылки о нормальности (или другом распределении).

Связанные пары:

- ► критерий знаков;
- ► критерий знаковых рангов.

Независимые выборки:

- ► критерий ранговых сумм.



Джон Арбетнот

Фрэнк Уилкоксон

Критерий знаков
(sign test)

Пример *(Newbold)*. A random sample of six salespersons who had attended a motivational course on sales techniques was monitored in the three months before and the three months after the course. The table shows the values of sales, in thousands of dollars, generated by these six salespersons in the two periods.

Salesperson	*Before the course*	*After the course*
1	212	237
2	282	291
3	203	191
4	327	341
5	165	192
6	198	180

Попробуем проверить гипотезу о неэффективности курсов на уровне 10%, опираясь не на объёмы продаж, а только на соотношения «больше-меньше».

Критерий знаков
(sign test)

Пример *(Newbold)*. A random sample of six salespersons who had attended a motivational course on sales techniques was monitored in the three months before and the three months after the course. The table shows the values of sales, in thousands of dollars, generated by these six salespersons in the two periods.

Salesperson	*Before the course(X)*	*After the course(Y)*	*Diff.*	*Sign*
1	212	237	-25	-
2	282	291	-9	-
3	203	191	12	+
4	327	341	-14	-
5	165	192	-27	-
6	198	180	18	+

Попробуем проверить гипотезу о неэффективности курсов на уровне 10%, опираясь не на объёмы продаж, а только на соотношения «больше-меньше».

В двух случаях объём продаж снизился («+»), в четырёх — повысился («-»).

Сформулируем гипотезы:

$$\begin{aligned} &\mathrm{H_0}:\ \mathrm{P}(X_i > Y_i) = \mathrm{P}(X_i < Y_i) \\ &\mathrm{H_A}:\ \mathrm{P}(X_i > Y_i) < \mathrm{P}(X_i < Y_i) \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} &\mathrm{H_0}:\ \mathrm{Med}(X_i - Y_i) = 0 \\ &\mathrm{H_A}:\ \mathrm{Med}(X_i - Y_i) < 0 \end{aligned}$$

Какой будет критерий?

Критерий знаков

$$H_0: \ P(X_i > Y_i) = P(X_i < Y_i)$$
$$H_A: \ P(X_i > Y_i) < P(X_i < Y_i)$$

Пусть пары $(X_1, Y_1), \dots, (X_6, Y_6)$ таковы, что

$$d_i = X_i - Y_i \ \sim \ \text{i.i.d.}$$

n^+ - число наблюдений, где $d_i > 0 \quad (X_i > Y_i)$.

n^- - число наблюдений, где $d_i < 0 \quad (X_i > Y_i)$.

Предположим, что $P(X_i = Y_i) = 0$ (на практике такие наблюдения просто выкидываются).

Тогда

$$n^+ \overset{H_0}{\sim} Bi(6,\ 0.5), \qquad n^- \overset{H_0}{\sim} Bi(6,\ 0.5).$$

Критерий знаков

Пусть пары $(X_1, Y_1), \dots, (X_6, Y_6)$ таковы, что

$$d_i = X_i - Y_i \ \sim \ \text{i.i.d.}$$

n^+ - число наблюдений, где $d_i > 0 \quad (X_i > Y_i)$.

n^- - число наблюдений, где $d_i < 0 \quad (X_i > Y_i)$.

Предположим, что $\mathrm{P}(X_i = Y_i) = 0$ (на практике такие наблюдения просто выкидываются).

Тогда

$$n^+ \overset{\mathrm{H}_0}{\sim} Bi(6,\, 0.5), \qquad n^- \overset{\mathrm{H}_0}{\sim} Bi(6,\, 0.5).$$

Возьмём в качестве статистики $T = n^+$

В пользу $\mathrm{H}_\mathrm{A}: \ \mathrm{P}(X_i > Y_i) < \mathrm{P}(X_i < Y_i)$ будут говорить значения T, близкие к 0.

Выберем критическую область возле 0 так, чтобы вероятность попасть туда при H_0 *не превосходила* уровень значимости 10%.

Критерий знаков

$$n^{+} \overset{H_0}{\sim} Bi(6,\ 0.5), \quad n^{-} \overset{H_0}{\sim} Bi(6,\ 0.5).$$

Возьмём в качестве статистики $T = n^{+}$

В пользу $H_A:\ P(X_i > Y_i) < P(X_i < Y_i)$ будут говорить значения n^{+}, близкие к 0.

Выберем критическую область возле 0 так, чтобы вероятность попасть туда при H_A *не превосходила* уровень значимости 10%.

$$P(n^{+} = 0) = C_6^0 0.5^0 (1-0.5)^6 = 0.5^6 \approx 0.0156$$

$$P(n^{+} = 1) = C_6^1 0.5^6 \approx 0.0938$$

Критическая область	Вероятность ошибки первого рода	
$T \le 0$	0.0156	<10%
$T \le 1$	0.0156+0.0938=0.1094	>10%

Нужно брать критическое значение 0. То есть отвергать H_0 при $T \le 0$.

Критерий знаков

$$n^{+} \overset{H_0}{\sim} Bi(6,\,0.5), \qquad n^{-} \overset{H_0}{\sim} Bi(6,\,0.5).$$

Возьмём в качестве статистики $T = n^{+}$

В пользу $H_A:\ P(X_i > Y_i) < P(X_i < Y_i)$ будут говорить значения n^{+}, близкие к 0.

Выберем критическую область возле 0 так, чтобы вероятность попасть туда при H_A *не превосходила* уровень значимости 10%.

$$P(n^{+} = 0) = C_6^0 0.5^0 (1-0.5)^6 = 0.5^6 \approx 0.0156$$

$$P(n^{+} = 1) = C_6^1 0.5^6 \approx 0.0938$$

Критическая область	Вероятность ошибки первого рода	
$T \le 0$	0.0156	<10%
$T \le 1$	0.0156+0.0938=0.1094	>10%

Нужно брать критическое значение 0. То есть отвергать H_0 при $T \le 0$.

В примере с продавцами рассчитанное значение статистики $T = 2 > 0$,
так что нет оснований отвергать основную гипотезу и считать, что курсы способствуют повышению продаж.

Критерий знаков в общем виде

Выборка из связанных пар:

$$\left(X_1, Y_1\right), \dots, \left(X_{\tilde{n}}, Y_{\tilde{n}}\right)$$

Гипотезы:

$$\mathrm{H}_0 : \mathrm{P}\left(X_i > Y_i\right) = \mathrm{P}\left(X_i < Y_i\right) \quad \Longleftrightarrow \quad \mathrm{H}_0 : \mathrm{Med}\left(X_i - Y_i\right) = 0$$

$$\left[\begin{array}{l} \mathrm{H}_\mathrm{A} : \mathrm{P}\left(X_i > Y_i\right) \neq \mathrm{P}\left(X_i < Y_i\right) \\ \mathrm{H}_\mathrm{A} : \mathrm{P}\left(X_i > Y_i\right) > \mathrm{P}\left(X_i < Y_i\right) \\ \mathrm{H}_\mathrm{A} : \mathrm{P}\left(X_i > Y_i\right) < \mathrm{P}\left(X_i < Y_i\right) \end{array}\right. \quad \Longleftrightarrow \quad \left[\begin{array}{l} \mathrm{H}_\mathrm{A} : \mathrm{Med}\left(X_i - Y_i\right) \neq 0 \\ \mathrm{H}_\mathrm{A} : \mathrm{Med}\left(X_i - Y_i\right) > 0 \\ \mathrm{H}_\mathrm{A} : \mathrm{Med}\left(X_i - Y_i\right) < 0 \end{array}\right.$$

Выборка разностей $d_i = X_i - Y_i$:

$$d_1, \dots, d_{\tilde{n}} \ \sim \ \text{i.i.d.}$$

n^+ - число наблюдений, где $d_i > 0 \ \left(X_i > Y_i\right);$

n^- - число наблюдений, где $d_i < 0 \ \left(X_i < Y_i\right).$

Удаляем те наблюдения, где $d_i = 0 \ \left(X_i = Y_i\right).$

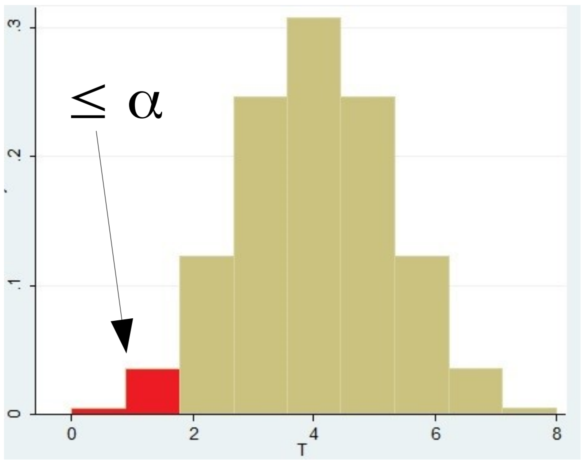
Остаётся $n = n^+ + n^-$ наблюдений.

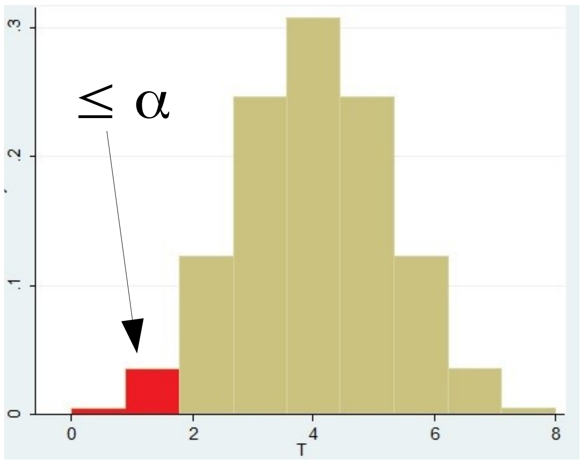
$$n^+ \overset{\mathrm{H}_0}{\sim} Bi\left(n, 0.5\right), \qquad n^- \overset{\mathrm{H}_0}{\sim} Bi\left(n, 0.5\right).$$

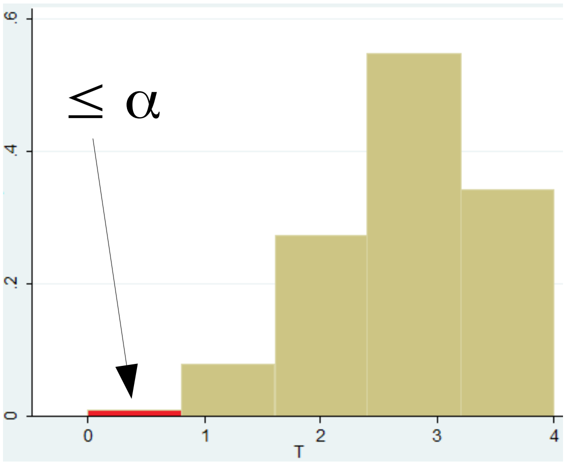
Критерий знаков в общем виде

$$H_0: \mathrm{Med}(X_i - Y_i) = 0$$

$H_A: \mathrm{Med}(X_i - Y_i) > 0$	$H_A: \mathrm{Med}(X_i - Y_i) < 0$	$H_A: \mathrm{Med}(X_i - Y_i) \neq 0$
Статистика: $T = n^- \overset{H_0}{\sim} Bi(n, 0.5)$	$T = n^+ \overset{H_0}{\sim} Bi(n, 0.5)$	$T = \min(n^-, n^+)$ $P(T=k) \overset{H_0}{=} 2C_n^k 0.5^n = P(n^+=k) + P(n^-=k)$, $k = 0,1,\dots,\lfloor n/2 \rfloor$
Критическая область: $T \le T_{n,\alpha}$	$T \le T_{n,\alpha}$	$T \le T_{n,\alpha/2}$







$$T_{n,\alpha} = \max\{k : P_{H_0}(n^+ \le k) \le \alpha\} = \max\{k : P_{H_0}(n^- \le k) \le \alpha\} = \max\{k : \sum_{i=0}^{k} C_n^i 0.5^n \le \alpha\}.$$

Table J Critical Values for the Sign Test

Reject the null hypothesis if the smaller number of positive or negative signs is less than or equal to the value in the table.

	One-tailed, $\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.05$
n	Two-tailed, $\alpha = 0.01$	$\alpha = 0.02$	$\alpha = 0.05$	$\alpha = 0.10$
8	0	0	0	1
9	0	0	1	1
10	0	0	1	1
11	0	1	1	2
12	1	1	2	2
13	1	1	2	3
14	1	2	3	3
15	2	2	3	3
16	2	2	3	4
17	2	3	4	4
18	3	3	4	5
19	3	4	4	5
20	3	4	5	5
21	4	4	5	6
22	4	5	5	6
23	4	5	6	7
24	5	5	6	7
25	5	6	6	7

Note: Table J is for one-tailed or two-tailed tests. The term n represents the total number of positive and negative signs. The test value is the number of less frequent signs.

Source: Table 1, p. 560, from "The Statistical Sign Test" by W. J. Dixon and A. M. Mood, vol. 41. no. 236 (Dec. 1946), pp. 557–566.

Критерий знаков: нормальное приближение

Рассмотрим величину n^+ (с тем же успехом можно взять n^-).

Её распределение при $\mathrm{H}_0 : \mathrm{P}(X_i > Y_i) = \mathrm{P}(X_i < Y_i)$:

$$n^+ \sim Bi(n,\ 0.5)$$

$$\mathrm{E}(n^+) = n/2;$$
$$\mathrm{D}(n^+) = n/4.$$

Центрируем и нормируем:

$$Z = \frac{n^+ - \frac{n}{2}}{\sqrt{\frac{n}{4}}} = \frac{2n^+ - n}{\sqrt{n}}.$$

По интегральной теореме Муавра-Лапласа:

$$\boxed{Z = \frac{2n^+ - n}{\sqrt{n}} \underset{asy}{\overset{\mathrm{H}_0}{\sim}} N(0,1).}$$

Критические области — как обычно.

В книжке P. Newbold «Statistics for Business and Economics» статистика чуть отличается.

Пример

Разработчик ПО собирается существенно поменять интерфейс поставляемой программы и желает оценить, каково отношение потенциальных пользователей к перемене.
Группе в 30 человек было предложено оценить два варианта (новый и старый) по пятибалльной шкале. Результаты обследования:

8 человек дали более высокую оценку старому варианту;
18 человек предпочли новый вариант;
4 человека дали одинаковые оценки обеим вариантам.

Есть ли основание считать, что большая часть пользователей предпочтёт новый интерфейс? Используйте уровень значимости 5%.

Пример

Разработчик ПО собирается существенно поменять интерфейс поставляемой программы и желает оценить, каково отношение потенциальных пользователей к перемене.
Группе в 30 человек было предложено оценить два варианта (новый и старый) по пятибалльной шкале. Результаты обследования:

8 человек дали более высокую оценку старому варианту;
18 человек предпочли новый вариант;
4 человека дали одинаковые оценки обеим вариантам.

Есть ли основание считать, что большая часть пользователей предпочтёт новый интерфейс? Используйте уровень значимости 5%.

Решение. Если обозначить за X оценку нового варианта, а за Y — старого, то

$$n^+ = 18, \quad n^- = 8, \quad n = n^+ + n^- = 26.$$

Статистика:

$$Z = \frac{2n^+ - n}{\sqrt{n}} = \frac{2\times 18 - 26}{\sqrt{26}} = 1.961$$

Пример

Разработчик ПО собирается существенно поменять интерфейс поставляемой программы и желает оценить, каково отношение потенциальных пользователей к перемене.
Группе в 30 человек было предложено оценить два варианта (новый и старый) по пятибалльной шкале. Результаты обследования:

8 человек дали более высокую оценку старому варианту;
18 человек предпочли новый вариант;
4 человека дали одинаковые оценки обеим вариантам.

Есть ли основание считать, что бóльшая часть пользователей предпочтёт новый интерфейс? Используйте уровень значимости 5%.

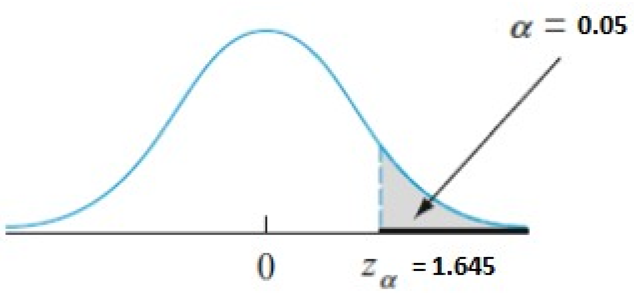
Решение. Если обозначить за X оценку нового варианта, а за Y — старого, то

$$n^+ = 18, \quad n^- = 8, \quad n = n^+ + n^- = 26.$$

Статистика:

$$Z = \frac{2n^+ - n}{\sqrt{n}} = \frac{2\times 18 - 26}{\sqrt{26}} = 1.961$$

Критическое значение: $z_\alpha = z_{0.05} = 1.645$



Вывод: бóльшая часть пользователей поддерживает новый вариант.

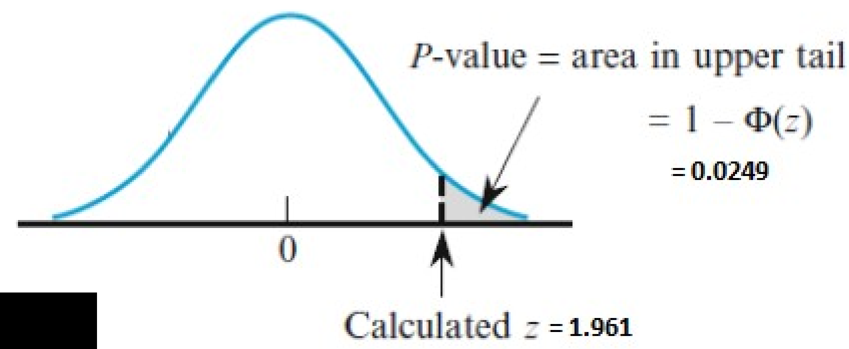
Сравним приближённый и точный критерий

Статистика: $n^{+} = 18, \quad n^{-} = 8, \quad n = n^{+} + n^{-} = 26.$

$$Z = \frac{2n^{+} - n}{\sqrt{n}} = \frac{2\times 18 - 26}{\sqrt{26}} = 1.961$$

p-значение по нормальному приближению:

$$\mathrm{P}(\xi > 1.961) = 0.0249, \quad \xi \sim N(0,1).$$



Точное *p*-значение:

```
Sign test

        sign |    observed    expected
-------------+------------------------
    positive |          18          13
    negative |           8          13
        zero |           4           4
-------------+------------------------
         all |          30          30

One-sided tests:
  Ho: median of x - y = 0 vs.
  Ha: median of x - y > 0
      Pr(#positive >= 18) =
         Binomial(n = 26, x >= 18, p = 0.5) =  0.0378

  Ho: median of x - y = 0 vs.
  Ha: median of x - y < 0
      Pr(#negative >= 8) =
         Binomial(n = 26, x >= 8, p = 0.5) =  0.9855

Two-sided test:
  Ho: median of x - y = 0 vs.
  Ha: median of x - y != 0
      Pr(#positive >= 18 or #negative >= 18) =
         min(1, 2*Binomial(n = 26, x >= 18, p = 0.5)) =  0.0755
```

0.0249 и 0.0378 — разница может показаться существенной...

Наблюдений не достаточно, чтобы нормальное приближение поразило своей точностью.

Что такое ранг?

Ранг наблюдения по признаку X это место в упорядоченном по возрастанию X ряду.

Пример:

i	1	2	3	4	5	6	7	8	9
X_i	5	6	3	5	10	8	8	1	8
$rank(X_i)$	3.5	5	2	3.5	9	7	7	1	7

Значения признака в наблюдениях 1 и 4 совпадают — эти наблюдения делят 3 и 4 места в упорядоченном ряду. Им в соответствие ставится средний ранг (3+4)/2=3.5.

Критерий знаковых рангов
(Signed rank test)

Выборка из независимых пар:

$$\left(X_1, Y_1\right), \ldots, \left(X_n, Y_n\right)$$

Гипотезы:

$$\mathrm{H}_0\colon \operatorname{Med}\left(X_i - Y_i\right) = 0$$

$$\begin{cases} \mathrm{H}_\mathrm{A}\colon \operatorname{Med}\left(X_i - Y_i\right) \neq 0 \\ \mathrm{H}_\mathrm{A}\colon \operatorname{Med}\left(X_i - Y_i\right) > 0 \\ \mathrm{H}_\mathrm{A}\colon \operatorname{Med}\left(X_i - Y_i\right) < 0 \end{cases}$$

Переходим к разностям $d_i = X_i - Y_i$:

$$d_1, \ldots, d_n \ \sim \ \text{i.i.d.}$$

Предполагаем, что d_i распределены симметрично.

В этом случае $\operatorname{Med}\left(X_i - Y_i\right) = \mathrm{E}\left(X_i - Y_i\right)$, если $\mathrm{E}\left(X_i - Y_i\right)$ существует.

Наблюдения, где $X_i = Y_i$, опять выкидываются: n — число оставшихся наблюдений.

Критерий знаковых рангов

(Signed rank test)

Гипотезы: $$\mathrm{H}_0:\ \mathrm{Med}\left(X_i - Y_i\right) = 0$$

$$\begin{cases} \mathrm{H_A}:\ \mathrm{Med}\left(X_i - Y_i\right) \neq 0 \\ \mathrm{H_A}:\ \mathrm{Med}\left(X_i - Y_i\right) > 0 \\ \mathrm{H_A}:\ \mathrm{Med}\left(X_i - Y_i\right) < 0 \end{cases}$$

Переходим к разностям: $$d_1, \dots, d_n \sim \text{ i.i.d.}$$

Предполагаем, что d_i распределены симметрично.

Ранжируем наблюдения по модулю d_i:

Person	*Before(X)*	*After(Y)*	*Diff.*	*Sign*	$rank\left(\lvert d_i \rvert\right)$
1	212	237	-25	-	5
2	282	291	-9	-	1
3	203	191	12	+	2
4	327	341	-14	-	3
5	165	192	-27	-	6
6	198	180	18	+	4

Сумма рангов с положительными знаками: $$SR^+ = \sum_{i=1}^{n} rank\left(\lvert d_i \rvert\right) \cdot \left[d_i > 0\right] = 2 + 4 = 6.$$

С отрицательными знаками: $$SR^- = \sum_{i=1}^{n} rank\left(\lvert d_i \rvert\right) \cdot \left[d_i < 0\right] = 5 + 1 + 3 + 6 = 15.$$

Критерий знаковых рангов
(Signed rank test)

Основная гипотеза: $$\mathrm{H}_0:\ \mathrm{Med}\left(X_i - Y_i\right) = 0$$

Альтернатива:

$\mathrm{H_A}:\ \mathrm{Med}\left(X_i - Y_i\right) > 0$	$\mathrm{H_A}:\ \mathrm{Med}\left(X_i - Y_i\right) < 0$	$\mathrm{H_A}:\ \mathrm{Med}\left(X_i - Y_i\right) \neq 0$
Статистика:		
$T = SR^-$	$T = SR^+$	$T = \min\left(SR^+, SR^-\right)$
распределение?		
Крит. область:		
$T \le T_{n,\alpha}$	$T \le T_{n,\alpha}$	$T \le T_{n,\alpha/2}$

Это другие $T_{n,\alpha}$ *- не те, что для критерия знаков!*
Они приведены в особых таблицах.

В примере с продавцами: $T_{n,\alpha} = T_{6,0.1} = 3.$

$T = SR^+ = 6 > 3 \quad \Rightarrow$ Нет оснований отвергнуть основную гипотезу и считать, что курсы способствуют повышению продаж

Односторонние критические значения $T_{n,\alpha}$ для критерия знаковых рангов

$n \setminus \alpha$	**0.005**	**0.010**	**0.025**	**0.050**	**0.100**
4					0
5				0	2
6			0	2	3
7		0	2	3	5
8	0	1	3	5	8
9	1	3	5	8	10
10	3	5	8	10	14
11	5	7	10	13	17
12	7	9	13	17	21
13	9	12	17	21	26
14	12	15	21	25	31
15	15	19	25	30	36
16	19	23	29	35	42
17	23	27	34	41	48
18	27	32	40	47	55
19	32	37	46	53	62
20	37	43	52	60	69

Откуда берутся эти значения?

Можно пойти в лоб. Если перебрать все возможные комбинации знаков и рангов для выборки в 6 наблюдений, то поймём:

- всего комбинаций 64;

- «критические» комбинации (в которых $T = SR^+ \leq 4 = T_{6,0.1}$):

$$\left.\begin{array}{llllllll} -1 & -2 & -3 & -4 & -5 & -6 & \Rightarrow & SR^+ = 0 \\ +1 & -2 & -3 & -4 & -5 & -6 & \Rightarrow & SR^+ = 1 \\ -1 & +2 & -3 & -4 & -5 & -6 & \Rightarrow & SR^+ = 2 \\ +1 & +2 & -3 & -4 & -5 & -6 & \Rightarrow & SR^+ = 3 \\ -1 & -2 & +3 & -4 & -5 & -6 & \Rightarrow & SR^+ = 3 \end{array}\right\} \text{5 комбинаций из 64}$$

Здесь я пренебрегаю возможностью одинаковых рангов...

Вероятность ошибки первого рода: $\frac{5}{64} \approx 0.078 \leq 0.1$.

При критической области $T \leq 4$ эта вероятность была бы равна $\frac{7}{64} \approx 0.109 > 0.1$.

Нормальное приближение для критерия знаковых рангов

При верной $\mathrm{H}_0: \mathrm{Med}\left(X_i - Y_i\right) = 0$

$$\mu^{+} = \mathrm{E}\left(SR^{+}\right) = \frac{n\left(n+1\right)}{4};$$

$$\sigma^{2+} = \mathrm{D}\left(SR^{+}\right) = \frac{n\left(n+1\right)\left(2n+1\right)}{24}.$$

Статистика:

$$Z = \frac{SR^{+} - \mu^{+}}{\sigma^{+}} \underset{asy}{\overset{\mathrm{H}_0}{\sim}} N\left(0,1\right).$$

Критическая область?

Догадайтесь сами.

Сопоставление критериев для сравнения связанных пар

Достоинства	*Недостатки*
t-критерий	
► Мощность	► Требует нормальность ► Чувствителен к выбросам
Критерий знаков	
► Не предполагает определённого распределения ► Устойчив к выбросам ► Применим к порядковым признакам	► Низкая мощность
Критерий знаковых рангов	
► Мало зависит от распределения ► Устойчив к выбросам	► При нормальности мощность чуть ниже, чем у t-критерия ► Требует симметричность

Замечания:

► Критерии знаков и знаковых рангов могут применяться не только для сравнения связанных пар. Они годятся для проверки гипотезы о медиане простого (одномерного) статистического признака.

► Критерий знаковых рангов часто называют критерием Уилкоксона (Wilcoxon test). Так же называют и существенно другой критерий ранговых сумм (rank sum test). Не путайте.

Критерий ранговых сумм разберёте сами

Сейчас приятнее поболтать о параметрических и непараметрических методах вообще

Следующая лекция:

Однофакторный дисперсионный анализ

Критерий Краскелла-Уоллиса