

Критерии дисперсионного анализа в R

Критерий Уилкоксона-Манна-Уитни (независимые выборки) и критерий ранговых знаков Уилкоксона (связные выборки)

```
1 wilcox.test(x, y = NULL,  
2             alternative = c("two.sided", "less", "greater"),  
3             mu = 0, paired = FALSE, exact = NULL, correct = TRUE,  
4             conf.int = FALSE, conf.level = 0.95, ...)  
5  
6 wilcox.test(formula, data, subset, na.action, ...)
```

Параметры

для критерия Уилкоксона-Манна-Уитни

- `x` и `y` -- выборки, могут быть разной длины;
- `mu` -- смещение, соответствующее основной гипотезе (обычно 0). Предполагается, что распределения выборок отличаются сдвигом `mu` ;
- `paired = FALSE` -- выборки независимые;

для критерия ранговых знаков Уилкоксона

- - `x` -- выборка разностей;
- или
- - `x` и `y` выборки одинаковой длины, используются только разности;
- `mu` -- смещение, соответствующее основной гипотезе (обычно 0). Предполагается, что распределение выборки разностей (`x` или `x - y`) симметрично относительно `mu` ;
- `paired = TRUE` -- выборки связанные;

Общие параметры

- `alternative` -- тип альтернативной гипотезы (двусторонняя, односторонняя);
- `correct` -- следует ли применять коррекцию при аппроксимации нормальным распределением;
- `exact` -- использовать ли точные вычисления или же асимптотические;
- `conf.int` -- строить ли доверительный интервал;
- `conf.level` -- уровень доверия доверительного интервала;
- `formula` -- формула в виде `lhs ~ rhs` , где `lhs` -- числовой признак, а `rhs` -- фактор с двумя уровнями (бинарная переменная). Выборки получаются разделением числового признака по значению фактора;
- `data` -- данные (матрица или таблица);
- `na.action` -- функция, указывающая что делать с пропусками в данных.

Возвращают:

- `statistic` -- статистика критерия;
- `parameter` -- параметр распределения;
- `p.value` -- p-value критерия;
- `estimate` -- оценка параметра сдвига (только если `conf.int = TRUE`);
- `conf.int` -- доверительный интервал параметра сдвига (только если `conf.int = TRUE`).

Примеры:

In [1]:

```
1 x <- c(1, 2, 3, 4, 5)
2 y <- c(6, 7, 8, 9)
3 wilcox.test(x, y, conf.int = TRUE, conf.level = 0.90)
```

Wilcoxon rank sum test

```
data: x and y
W = 0, p-value = 0.01587
alternative hypothesis: true location shift is not equal to 0
90 percent confidence interval:
 -7 -2
sample estimates:
difference in location
      -4.5
```

In [2]:

```
1 x <- c(1, 2, 3, 4, 5)
2 y <- c(10, 9, 8, 7, 6)
3 wilcox.test(x, y, paired = TRUE, conf.int = TRUE, conf.level = 0.90)
```

Wilcoxon signed rank test

```
data: x and y
V = 0, p-value = 0.0625
alternative hypothesis: true location shift is not equal to 0
90 percent confidence interval:
 -9 -1
sample estimates:
(pseudo)median
      -5
```

Датасет mtcars встроен в R

In [3]:

```
1 head(mtcars)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

Величина am -- бинарный фактор. По нему любой столбец можно разбить на две независимые выборки. Вызов функции выглядит так:

In [4]:

```
1 wilcox.test(mpg ~ am, data = mtcars, conf.int = TRUE, conf.level = 0.90)
```

```
Warning message in wilcox.test.default(x = c(21.4, 18.7, 18.1, 14.3, 24.4, 22.8, :  
4.4, 22.8, :
```

```
"cannot compute exact p-value with ties"Warning message in wilcox.test.default(x = c(21.4, 18.7, 18.1, 14.3, 24.4, 22.8, :  
t.default(x = c(21.4, 18.7, 18.1, 14.3, 24.4, 22.8, :  
"cannot compute exact confidence intervals with ties"
```

Wilcoxon rank sum test with continuity correction

```
data: mpg by am
```

```
W = 42, p-value = 0.001871
```

```
alternative hypothesis: true location shift is not equal to 0
```

```
90 percent confidence interval:
```

```
-10.999924 -3.600056
```

```
sample estimates:
```

```
difference in location
```

```
-6.799963
```

Критерий знаков (связные выборки)

```
1 binom.test(x, n, p = 0.5,  
2           alternative = c("two.sided", "less", "greater"),  
3           conf.level = 0.95)
```

Параметры

- `x` -- вектор длины 2: количество единиц, количество нулей;
или
- `x` и `n` -- количество единиц, размер выборки;
- `p` -- вероятность единицы (обычно 0.5);
- `alternative` -- тип альтернативной гипотезы (двусторонняя, односторонняя);
- `conf.level` -- уровень доверия доверительного интервала;

Возвращают:

- `statistic` -- число единиц;
- `parameter` -- число наблюдений;
- `p.value` -- p-value критерия;
- `estimate` -- оценка вероятности единицы;
- `conf.int` -- доверительный интервал вероятности единицы.

Примеры:

In [5]:

```
1 binom.test(5, 18)
```

Exact binomial test

data: 5 and 18

number of successes = 5, number of trials = 18, p-value = 0.09625

alternative hypothesis: true probability of success is not equal to 0.

5

95 percent confidence interval:

0.09694921 0.53480197

sample estimates:

probability of success

0.2777778

Критерий Колмогорова (критерии согласия) и критерий Смирнова (критерии однородности)

```
1 ks.test(x, y, ..., alternative = c("two.sided", "less", "greater"), exact =  
  NULL, tol=1e-8, simulate.p.value=FALSE, B=2000)
```

Параметры

для критерия Колмогорова

- `x` -- выборка;
- `y` -- функция распределения или ее название;

для критерия Смирнова

- - `x` и `y` -- выборки, могут быть разной длины;

Общие параметры

- `alternative` -- тип альтернативной гипотезы (двусторонняя, односторонняя);
- `correct` -- следует ли применять коррекцию при аппроксимации нормальным распределением;
- `exact` -- использовать ли точные вычисления или же асимптотические.

Возвращают:

- `statistic` -- статистика критерия;
- `p.value` -- p-value критерия.

Примеры:

In [6]:

```
1 x <- c(1, 2, 3, 4, 5)
2 y <- c(6, 7, 8, 9)
3 ks.test(x, y)
```

Two-sample Kolmogorov-Smirnov test

data: x and y

D = 1, p-value = 0.01587

alternative hypothesis: two-sided

In [7]:

```
1 ks.test(x, 'pnorm')
```

One-sample Kolmogorov-Smirnov test

data: x

D = 0.84134, p-value = 0.000201

alternative hypothesis: two-sided

Прикладная статистика и анализ данных, 2019

Никита Волков

<https://mipt-stats.gitlab.io/> (<https://mipt-stats.gitlab.io/>).