

# Критерии дисперсионного анализа в R

## Критерий Стьюдента (нормальные выборки)

```
1 ## Default S3 method:
2 t.test(x, y = NULL,
3       alternative = c("two.sided", "less", "greater"),
4       mu = 0, paired = FALSE, var.equal = FALSE,
5       conf.level = 0.95, ...)
6
7 ## S3 method for class 'formula'
8 t.test(formula, data, subset, na.action, ...)
```

### Параметры

#### *для независимых выборок*

- `x` и `y` -- выборки, могут быть разной длины;
- `mu` -- разница средних, соответствующая основной гипотезе (обычно 0);
- `paired = FALSE` -- выборки независимые;

#### *для связанных выборок*

- - `x` -- выборка разностей;
- или
- - `x` и `y` выборки одинаковой длины, используются только разности;
- `mu` -- разница средних, соответствующая основной гипотезе (обычно 0);
- `paired = TRUE` -- выборки связанные;

### Общие параметры

- `alternative` -- тип альтернативной гипотезы (двусторонняя, односторонняя). Например, `alternative = "greater"` соответствует случаю, что среднее `x` больше чем среднее `y`;
- `var.equal` -- известно ли равенство дисперсий;
- `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` -- число степеней свободы распределения Стьюдента;
- `p.value` -- p-value критерия;
- `estimate` -- оценка разности средних;
- `conf.int` -- доверительный интервал для разности средних.

### Примеры:

In [8]:

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

Если неизвестно, равны ли дисперсии

In [9]:

```
1 t.test(x, y, conf.level = 0.90)
```

Welch Two Sample t-test

```
data: x and y
t = -4.7001, df = 6.9808, p-value = 0.002225
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
 -6.314673 -2.685327
sample estimates:
mean of x mean of y
    3.0      7.5
```

Если известно равенство дисперсий

In [10]:

```
1 t.test(x, y, conf.level = 0.90, var.equal = TRUE)
```

Two Sample t-test

```
data: x and y
t = -4.5826, df = 7, p-value = 0.002536
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
 -6.360439 -2.639561
sample estimates:
mean of x mean of y
    3.0      7.5
```

Если выборки парные

In [11]:

```
1 t.test(1:5, 10:6, conf.level = 0.90, paired = TRUE)
```

#### Paired t-test

```
data: 1:5 and 10:6
t = -3.5355, df = 4, p-value = 0.02411
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
 -8.014887 -1.985113
sample estimates:
mean of the differences
          -5
```

Одновыборочный критерий о равенстве среднего нулю

In [12]:

```
1 t.test(x, conf.level = 0.90)
```

#### One Sample t-test

```
data: x
t = 4.2426, df = 4, p-value = 0.01324
alternative hypothesis: true mean is not equal to 0
90 percent confidence interval:
 1.492557 4.507443
sample estimates:
mean of x
          3
```

Датасет mtcars встроен в R

In [13]:

```
1 head(mtcars)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
<b>Mazda RX4</b>	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
<b>Mazda RX4 Wag</b>	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
<b>Datsun 710</b>	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
<b>Hornet 4 Drive</b>	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
<b>Hornet Sportabout</b>	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
<b>Valiant</b>	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 [14]:

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

Welch Two Sample t-test

```
data: mpg by am
t = -3.7671, df = 18.332, p-value = 0.001374
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
 -10.576623  -3.913256
sample estimates:
mean in group 0 mean in group 1
    17.14737      24.39231
```

---

Прикладная статистика и анализ данных, 2019

Никита Волков

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