

Необходимые импорты

In []:

```
import pandas as pd
from scipy import stats
from statsmodels.formula.api import ols
```

Загрузка данных

- функция `read_csv()` из `pandas`;
- функция `read_excel()` из `pandas`.

Работа с датафреймами

Пример готового датафрейма `df`, упоминается в коде ниже:

	group	x	y	z
0	A	1	10	1
1	B	2	24	0
2	B	4	38	0
3	A	8	17	1
4	B	5	15	0

Описание данных

- `df.info()`
- `df.describe()`
- `df["x"].describe()`
- `df["group"].value_counts()`

Группировка и описание

- `df.groupby("group")["x"].describe()`

Графики

- `df["x"].hist();`
- `df["x"].plot.box();`

Критерий Стьюдента для двух независимых выборок

- если `df_a` и `df_b` – датафреймы со строками, отобранными для группы А и группы В соответственно, то `stats.ttest_ind(df_a["x"], df_b["x"])`

Линейные регрессионные модели

- пример 1:

```
model = ols("y ~ x", data = df).fit()
model.summary()
```

- пример 2:

```
model = ols("y ~ x + z", data = df).fit()
model.summary()
```

- пример 3:

```
model = ols("y ~ x + group", data = df).fit()
model.summary()
```