

In [ ]:

```
import pandas as pd
import numpy as np
from statsmodels.stats.weightstats import _tconfint_generic
from google.colab import drive
```

In [ ]:

```
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

In [ ]:

```
df=pd.read_csv('/content/drive/MyDrive/data_python/AGU_rules/water.txt', sep = '\\t')
```

In [ ]:

```
df.head()
```

Out[ ]:

	location	town	mortality	hardness
0	South	Bath	1247	105
1	North	Birkenhead	1668	17
2	South	Birmingham	1466	5
3	North	Blackburn	1800	14
4	North	Blackpool	1609	18

# Используя метод .describe() вычислите описательные статистики для северных и южных городов. Сравните средние значения смертности в северных и южных городах и значения концентрации кальция в питьевой воде.

In [ ]:

```
df.loc[df.location == 'South'].describe()
```

Out[ ]:

	mortality	hardness
count	26.000000	26.000000
mean	1376.807692	69.769231
std	140.269175	40.360682
min	1096.000000	5.000000
25%	1259.250000	40.250000
50%	1364.000000	75.500000
75%	1485.750000	99.750000
max	1627.000000	138.000000

In [ ]:

```
df.loc[df.location == 'North'].describe()
```

Out[ ]:

	mortality	hardness
count	35.000000	35.000000
mean	1633.600000	30.400000
std	136.936911	26.134494
min	1378.000000	6.000000
25%	1557.500000	12.500000
50%	1637.000000	17.000000
75%	1718.000000	44.000000
max	1987.000000	94.000000

Средние значения смертности в северных городах *выше*, чем в южных.

Средние значения жесткости воды в северных городах *ниже*, чем в южных.

# Постройте 95% доверительные интервалы для средней годовой смертности по всем южным и северным городам. Отличаются ли границы интервалов?

In [ ]:

```
df_s=df.loc[df.location == 'South']
df_n=df.loc[df.location == 'North']
```

In [ ]:

```
mort_s_mean = df_s.mortality.mean()
mort_n_mean = df_n.mortality.mean()
```

In [ ]:

```
mort_s_mean_std = df_s.mortality.std() / np.sqrt(df_s.mortality.shape[0])
mort_n_mean_std = df_n.mortality.std() / np.sqrt(df_n.mortality.shape[0])
```

In [ ]:

```
print('95% interval of mortality in South towns:', _tconfint_generic(mort_s_mean, mort_s_mean_std, df_s.mortality.shape[0] - 1, 0.05, 'two-sided'))
```

95% interval: (1320.1517462936238, 1433.463638321761)

In [ ]:

```
print('95% interval of mortality in North towns:', _tconfint_generic(mort_n_mean, mort_n_mean_std, df_n.mortality.shape[0] - 1, 0.05, 'two-sided'))
```

95% interval: (1586.5605251961385, 1680.6394748038613)

Да, отличаются. Разброс смертности у Северных городов меньше, хотя абсолютные значения больше

# Постройте 95% доверительные интервалы для средней концентрации кальция в питьевой воде для южных и северных городов. Отличаются ли границы интервалов?

In [ ]:

```
hard_s_mean = df_s.hardness.mean()
hard_n_mean = df_n.hardness.mean()
```

In [ ]:

```
hard_s_mean_std = df_s.hardness.std() / np.sqrt(df_s.hardness.shape[0])
hard_n_mean_std = df_n.hardness.std() / np.sqrt(df_n.hardness.shape[0])
```

In [ ]:

```
print('95% interval of hardness in South towns:', _tconfint_generic(hard_s_mean, hard_s_mean_std, df_s.hardness.shape[0] - 1, 0.05, 'two-sided'))
```

95% interval of hardness in South towns: (53.467198692036106, 86.07126284642544)

In [ ]:

```
print('95% interval of hardness in North towns:', _tconfint_generic(hard_n_mean, hard_n_mean_std, df_n.hardness.shape[0] - 1, 0.05, 'two-sided'))
```

95% interval of hardness in North towns: (21.42248728572426, 39.37751271427574)

Да, отличаются. И не пересекаются.

Даже не знаю, какой тут можно сделать вывод... Пейте жесткую воду, живите дольше.

Храни вас Топ!

In [ ]: