

Добрый день! Данные для задания были взяты из платформы Kaggle, из базы данных, посвященной факторам, влияющим на стресс у студентов. Я решила с помощью бутстрэп построить доверительный интервал для уровня тревожности, который оценивался по шкале от 0 до 21.

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
```

```
In [2]: !pip install arch
```

```
Requirement already satisfied: arch in c:\users\ekaterina\anaconda3\lib\site-packages (6.2.0)
Requirement already satisfied: pandas>=1.1 in c:\users\ekaterina\anaconda3\lib\site-packages (from arch) (1.4.4)
Requirement already satisfied: scipy>=1.5 in c:\users\ekaterina\anaconda3\lib\site-packages (from arch) (1.9.1)
Requirement already satisfied: statsmodels>=0.12 in c:\users\ekaterina\anaconda3\lib\site-packages (from arch) (0.13.2)
Requirement already satisfied: numpy>=1.19 in c:\users\ekaterina\anaconda3\lib\site-packages (from arch) (1.21.5)
Requirement already satisfied: pytz>=2020.1 in c:\users\ekaterina\anaconda3\lib\site-packages (from pandas>=1.1->arch) (2022.1)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\ekaterina\anaconda3\lib\site-packages (from pandas>=1.1->arch) (2.8.2)
Requirement already satisfied: patsy>=0.5.2 in c:\users\ekaterina\anaconda3\lib\site-packages (from statsmodels>=0.12->arch) (0.5.2)
Requirement already satisfied: packaging>=21.3 in c:\users\ekaterina\anaconda3\lib\site-packages (from statsmodels>=0.12->arch) (21.3)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\users\ekaterina\anaconda3\lib\site-packages (from packaging>=21.3->statsmodels>=0.12->arch) (3.0.9)
Requirement already satisfied: six in c:\users\ekaterina\anaconda3\lib\site-packages (from patsy>=0.5.2->statsmodels>=0.12->arch) (1.16.0)
```

```
In [3]: from arch.bootstrap import IIDBootstrap, IndependentSamplesBootstrap
```

```
In [4]: data = pd.read_csv('StressLevelDataset.csv')
stress = data["anxiety_level"]
stress
```

```
Out[4]: 0      14
1      15
2      12
3      16
4      16
..
1095   11
1096    9
1097    4
1098   21
1099   18
Name: anxiety_level, Length: 1100, dtype: int64
```

```
In [5]: np.mean(stress)
```

```
Out[5]: 11.063636363636364
```

```
In [6]: rng = np.random.default_rng(111111)
n_boot = 1000
mu_stress = [np.mean(rng.choice(stress, size=len(stress))) for i in range(n_boot)]
[ np.quantile(mu_stress, 0.025), np.quantile(mu_stress, 0.975) ]
```

```
Out[6]: [10.722704545454546, 11.411886363636365]
```

```
In [7]: boot_x = IIDBootstrap(stress, seed=111111)
boot_x.conf_int(np.mean, method='percentile', reps=1000, size=0.95)
```

```
Out[7]: array([[10.72270455],
               [11.41188636]])
```

С 95% уверенностью среднее значение для уровня тревожности всех студентов находится в интервале от 10.72 до 11.41, как для метода процентилей, так и для базового метода

```
In [8]: boot_x.conf_int(np.median, method='percentile', reps=10000, size=0.95)
```

```
Out[8]: array([[11.],
               [12.]])
```

```
In [9]: boot_x.conf_int(np.std, method='percentile', reps=10000, size=0.95)
```

```
Out[9]: array([[5.9383271 ],
               [6.28352486]])
```

Для оценки различий между двумя средними значениями я решила разделить группу студентов на две подгруппы: одну с историей психических заболеваний (значение 1), а другую без такой истории (значение 0).

```
In [10]: health = data["mental_health_history"]
health_history_1 = stress[data["mental_health_history"] == 1].tolist()
health_history_0 = stress[data["mental_health_history"] == 0].tolist()
```

```
In [11]: health_history_1_array = np.array(health_history_1)
health_history_0_array = np.array(health_history_0)
```

```
In [12]: def mean_diff(health_history_1, health_history_0):
         return np.mean(health_history_0) - np.mean(health_history_1)
```

```
In [13]: boot_sh = IndependentSamplesBootstrap(health_history_1_array, health_history_0_array, se
boot_sh.conf_int(mean_diff, reps=10000, size=0.95, method='basic')
```

```
Out[13]: array([[ -8.32768222],
               [ -7.21201825]])
```

С 95% уверенностью можно утверждать, что разница в среднем уровне тревожности между группой студентов с историей психических заболеваний и группой без такой истории составляет от -8.33 до -7.21. Наличие отрицательных значений указывает на то, что уровень тревожности склонен быть выше у студентов с историей психических заболеваний по сравнению со студентами без такой истории.

```
In [ ]:
```