```
import numpy as np
import pandas as pd
import seaborn as sns
```

```
!pip install arch
```

```
Requirement already satisfied: arch in /usr/local/lib/python3.10/dist-packages (6.2.0)
Requirement already satisfied: numpy>=1.19 in /usr/local/lib/python3.10/dist-packages (from arch) (1.23.5)
Requirement already satisfied: scipy>=1.5 in /usr/local/lib/python3.10/dist-packages (from arch) (1.11.3)
Requirement already satisfied: pandas>=1.1 in /usr/local/lib/python3.10/dist-packages (from arch) (1.5.3)
Requirement already satisfied: statsmodels>=0.12 in /usr/local/lib/python3.10/dist-packages (from arch) (0.14.0)
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.1->arch) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.1->arch) (2023.3.post1)
Requirement already satisfied: patsy>=0.5.2 in /usr/local/lib/python3.10/dist-packages (from statsmodels>=0.12->arch) (0.5.3)
Requirement already satisfied: packaging>=21.3 in /usr/local/lib/python3.10/dist-packages (from statsmodels>=0.12->arch) (23.2)
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from patsy>=0.5.2->statsmodels>=0.12->arch) (1.16.0)
```

```
from arch.bootstrap import IIDBootstrap, IndependentSamplesBootstrap
```

```
rng = np.random.default_rng(121123)
x= rng.normal(loc=5, scale=4, size=20)
```

```
x
```

```
array([ 7.44448501,  1.32863533,  3.71716326,  0.81397629, 16.45224597,
        2.33720653, -0.1369632 ,  1.39384377,  2.37056083,  1.47463987,
        5.9776267 ,  8.19359627,  9.16280614,  8.02632743,  9.94009093,
        4.09321809,  5.01238673,  6.51929058, 11.4023354 ,  9.83357337])
```

$E(x_1) = 5$ $Var(X_i) = 4^2$, X_i независимы.

```
np.mean(x)
```

```
5.767852265342644
```

Поиграем в настоящего иссоелователя, сделаем вид, что мы не знаем $E(X_i)$. Мы можем получить точечную оценку для математического ожидания. Естественная формула, $\bar{x} = \frac{x_1 + \ldots + x_n}{n}$

```
mu_hat = np.mean(x)
mu_hat
```

```
5.767852265342644
```

Как построит доверительный интервал для неизвестного $\mu = E(X_i)$? Хочу точечную оценку превратить в интервальную. Достаточно "размножить" точечную оценку.

Вывод: из наших $n = 20$ наблюдений случайно выберем 20 с возможностью повтора.

```
x_star1 = rng.choice(x, size=len(x))
```

```
x_star1
```

```
array([ 3.71716326,  9.94009093,  5.01238673,  9.16280614,  9.94009093,
        4.09321809,  1.47463987,  9.94009093,  2.37056083,  8.02632743,
        0.81397629,  9.83357337, 11.4023354 ,  1.39384377,  6.51929058,
        5.9776267 ,  2.37056083,  1.39384377,  9.83357337, 16.45224597])
```

```
np.mean(x_star1)
```

```
6.483412259925059
```

```
x_star2 = rng.choice(x, size=len(x))
np.mean(x_star2)
```

```
5.3713085435961085
```

```
x_star2
```

```
array([ 9.83357337,  8.02632743,  3.71716326,  9.16280614,  0.81397629,
        2.37056083,  8.02632743,  7.44448501,  8.19359627,  3.71716326,
        1.47463987,  8.19359627, 11.4023354 ,  2.33720653,  5.9776267 ,
        7.44448501,  1.32863533,  1.47463987,  5.01238673,  1.47463987])
```

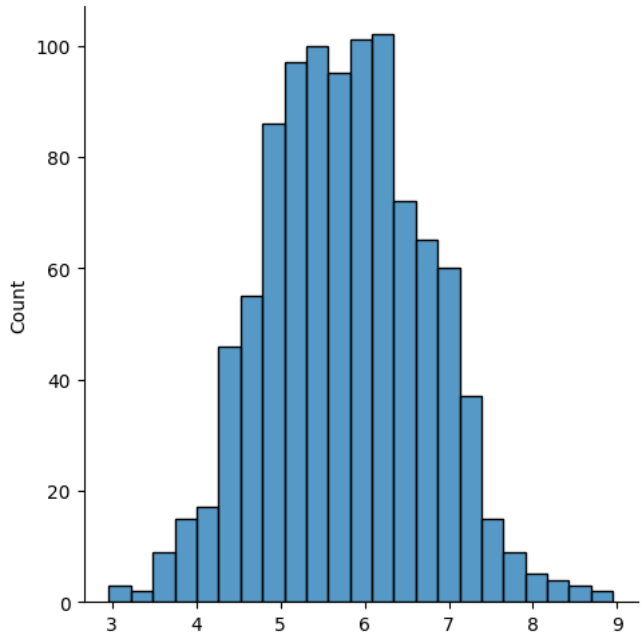
```
n_boot = 1000
mu_hat_star = [np.mean(rng.choice(x, size=len(x))) for i in range(n_boot)]
```

```
mu_hat_star[1:10]
```

```
[6.824113638919632,
 3.8279361494632624,
 6.608281676683572,
 5.084468555215613,
 6.784637792141162,
 4.65010663942731,
 6.745647633827874,
 5.177139612667885,
 5.824707420796941]
```

```
sns.displot(x=np.array(mu_hat_star))
```

```
<seaborn.axisgrid.FacetGrid at 0x788a788f43d0>
```



Наивный бутстрэп доверительный интервал для $\mu = E(X_i)$

```
[np.quantile(mu_hat_star, 0.025), np.quantile(mu_hat_star, 0.975)]
```

```
[3.949405203828274, 7.627184457355352]
```

Плюсы:

1. Мы можем не знать формулы для дисперсии оценки.

Мы не использовали $Var(\hat{\mu})$

2. В большинстве случаев формулы верны при больших n. И бутстрэп тоже требует больших n.

Часто оказывается, что большое n, нужное для "правильного" варианта бутстрэпа, гораздо меньше, чем больное n, нужное для центральной предельной теоремы. Здесь мы строим бутстрэп доверительный интервал для μ;

```
boot_x = IIDBootstrap(x, seed=121123)
boot_x.conf_int(np.mean, method='basic', reps=10000, size=0.95)
```

```
array([[3.89545113],
       [7.56242836]])
```

А здесь доверительный интервал для σ:

```
boot_x.conf_int(np.std, method='basic', reps=10000, size=0.95)
```

```
array([[3.09927118],
       [5.51846084]])
```

Tim Hestergerg, What teacher should know about bootstrap?

```
w = x + rng.normal(loc=1, scale=3, size=len(x))
```

```
np.corrcoef(x, w)
```

```
array([[1.        , 0.80815957],
       [0.80815957, 1.        ]])
```

```
def corr(x, y):
  corr_mat = np.corrcoef(x,y)
  return corr_mat[0,1]
```

```
corr(x, w)
```

```
0.8081595731806452
```

```
x
```

```
array([ 7.44448501,  1.32863533,  3.71716326,  0.81397629, 16.45224597,
        2.33720653, -0.1369632 ,  1.39384377,  2.37056083,  1.47463987,
        5.9776267 ,  8.19359627,  9.16280614,  8.02632743,  9.94009093,
        4.09321809,  5.01238673,  6.51929058, 11.4023354 ,  9.83357337])
```

```
w
```

```
array([ 7.63551032,  0.17151301,  4.0014103 ,  1.56258164, 16.84370869,
        2.88645306,  7.32445231,  1.57226989,  0.83172915,  4.15409999,
        6.96362435,  4.13350742, 11.63971414,  7.76525885,  5.79197105,
        1.06104007,  3.29572243, 11.06854471, 12.47158076, 14.31680673])
```

```
obs_id = rng.choice(range(len(x)), size=len(x))
obs_id
```

```
array([12,  3, 12, 16, 16, 10,  6,  0, 14, 15,  5, 18, 15, 17,  0, 15, 15,
        9,  5,  3])
```

```
x_star1 = x[obs_id]
x_star1
```

```
array([ 3.71716326,  9.94009093,  5.01238673,  9.16280614,  9.94009093,
        4.09321809,  1.47463987,  9.94009093,  2.37056083,  8.02632743,
        0.81397629,  9.83357337, 11.4023354 ,  1.39384377,  6.51929058,
        5.9776267 ,  2.37056083,  1.39384377,  9.83357337, 16.45224597])
```

```
w_star1 = w[obs_id]
w_star1
```

```
array([11.63971414,  1.56258164, 11.63971414,  3.29572243,  3.29572243,
        6.96362435,  7.32445231,  7.63551032,  5.79197105,  1.06104007,
        2.88645306, 12.47158076,  1.06104007, 11.06854471,  7.63551032,
        1.06104007,  1.06104007,  4.15409999,  2.88645306,  1.56258164])
```

```
corr(x_star1, w_star1)
```

```
-0.3036262548751661
```

```
boot_xw = IIDBootstrap(x, w, seed=121123)
boot_xw.conf_int(corr, method='basic', reps=10000, size=0.95)
```

```
array([[0.67662832],
       [1.08580312]])
```

```
4 - 3 == 1
```

```
True
```

```
0.4 - 0.3 == 0.1
```

```
False
```

```
y = rng.normal(loc=9, scale=3, size=25)
```

```
y
```

```
array([ 7.96933213,  5.31228744,  7.50993883, 13.21543022, 13.06881094,
       11.87431443,  9.50095962, 12.36131984,  6.73232622,  4.49216658,
        6.33703441,  9.23152675,  7.77628348, 11.64648968,  9.54274251,
       10.34804977,  3.91267876,  9.70893065,  8.49246366, 11.48938513,
       11.24965738,  7.89351279, 12.87663473,  9.98524934,  9.54815162])
```

```
np.mean(y) - np.mean(x)
```

```
3.5151748111511116
```

```
def mean_diff(x, y):
  return np.mean(y) - np.mean(x)
```

```
boot_xy = IndependentSamplesBootstrap(x, y, seed=121123)
boot_xy.conf_int(mean_diff, reps=10000, size=0.95, method='basic')
```

```
array([[1.42627125],
       [5.64971081]])
```