

Выборка, которая используется в задании 4, представляет собой суточную температуру воздуха в городе Санкт-Петербург в октябре месяце 2020 года.

```
Ввод [1]: 1 import pandas as pd
          2 import numpy as np
          3 import matplotlib.pyplot as plt
          4 import statistics
          5 import scipy.stats as sps
          6

Ввод [3]: 1 data = np.array([12.7, 14.3, 13.8, 13.2, 12.6, 15.1, 14.2, 12.9, 13.3, 12.8, 13.5, 13.8, 12.7, 12.9, 13.1, 14.5, 13.3, 13.7, 1
          2
          3 data
          4
          5
          6

Out[3]: array([12.7, 14.3, 13.8, 13.2, 12.6, 15.1, 14.2, 12.9, 13.3, 12.8, 13.5, 13.8, 12.7, 12.9, 13.1, 14.5, 13.3, 13.7, 13.1, 13.8, 13.1, 12.1, 13.6, 12.8, 12.7, 11.7, 12.9])
```

После создания массива numpy вычислим:

Среднее значение

Дисперсию

Стандартное отклонение.

```
Ввод [17]: 1 print ("This is the minimum number:", min(data))
          2 print ("This is the maximum number:", max(data))
          3 print ("This is the mean number:", sum(data)/len(data))
          4

This is the minimum number: 11.7
This is the maximum number: 15.1
This is the mean number: 13.253333333333336
```

```
Ввод [18]: 1 mean=np.mean(data)
          2 var=np.var(data)
          3 std=np.sqrt(var)
```

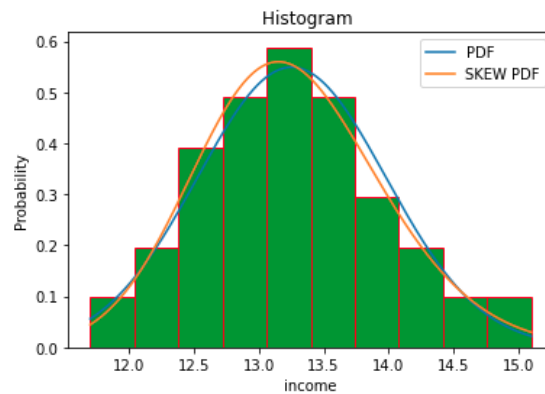
```
Ввод [28]: 1 print("This is the standard deviation:", std)
          2 print("This is var:", var)
          3 print("This is mean:", mean)
          4

This is the standard deviation: 0.7265137453773481
This is var: 0.5278222222222223
This is mean: 13.253333333333332
```

Далее построим гистограмму и функцию плотности вероятности (PDF):

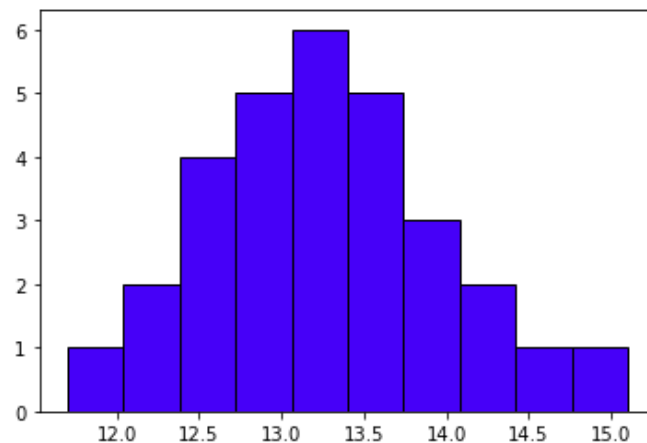
```
Ввод [20]: 1 x=np.linspace(min(data), max(data),1000)
```

```
Ввод [21]: 1 y_pdf=sps.norm.pdf(x,mean,std)
2 y_skew_pdf=sps.skewnorm.pdf(x,*sps.skewnorm.fit(data))
3 l1=plt.plot(x,y_pdf, label='PDF')
4 l2=plt.plot(x,y_skew_pdf, label='SKEW PDF')
5 n, bins,patches=plt.hist(data, 10,density=True,facecolor='g',edgecolor='red', alpha=0.85)
6 plt.xlabel('income')
7 plt.ylabel('Probability')
8 plt.title('Histogram ')
9 plt.legend((l1,l2),(l1.get_label(), l2.get_label()), loc='upper right')
10
11 plt.show()
12
```



```
Ввод [23]: 1 plt.hist(data,color = 'blue', edgecolor = 'black',
2             bins = 10)
```

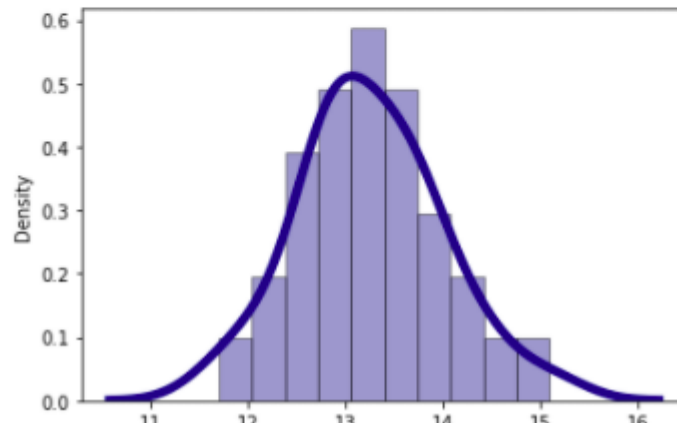
```
Out[23]: (array([1., 2., 4., 5., 6., 5., 3., 2., 1., 1.]),
array([11.7 , 12.04, 12.38, 12.72, 13.06, 13.4 , 13.74, 14.08, 14.42,
14.76, 15.1 ]),
<BarContainer object of 10 artists>)
```



```
Ввод [25]: 1 sns.distplot(data, hist=True, kde=True,
2             bins=10, color = 'darkblue',
3             hist_kws={'edgecolor':'black'},
4             kde_kws={'linewidth': 5})
```

C:\Users\maks-\anaconda3\envs\my\lib\site-packages\seaborn\distplot.py:100: FutureWarning: The distplot function is deprecated and will be removed in a future version. Please use 'kdeplot' (for kernel density estimation), 'histplot' (for histogram with similar flexibility) or 'histplot' (an axes-level function) instead. (https://seaborn.pydata.org/tutorial/axes.html#faceting)

Out[25]: <AxesSubplot:ylabel='Density'>



После анализа графика можно предположить, что распределение случайной величины является нормальным.

Гипотеза H_0 заключается в том, что выборка подчиняется нормальному закону. Альтернативная гипотеза в настоящем тесте будет звучать так: выборка не подчиняется нормальному закону.

Ниже представлены результат теста нашей гипотезы с помощью функции `normaltest`, которая проверяет нулевую гипотезу о том, что выборка происходит из нормального распределения. Он основан на тесте Д'Агостино и хи квадрат.

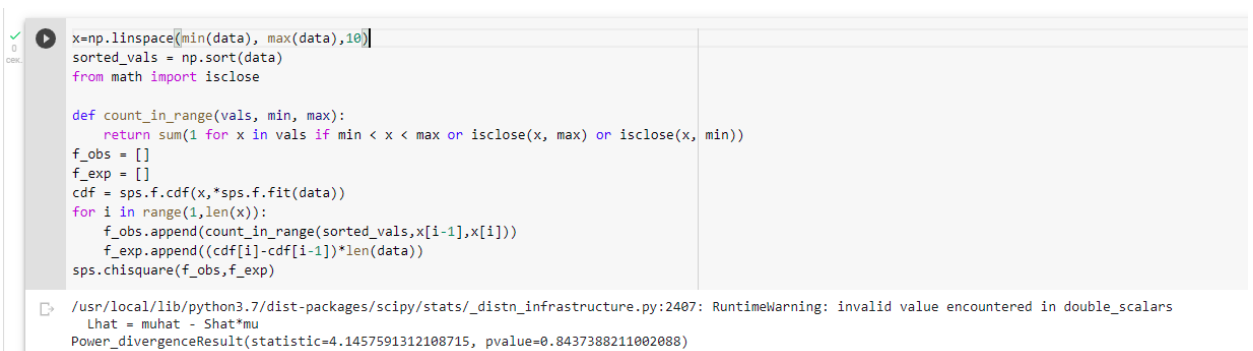
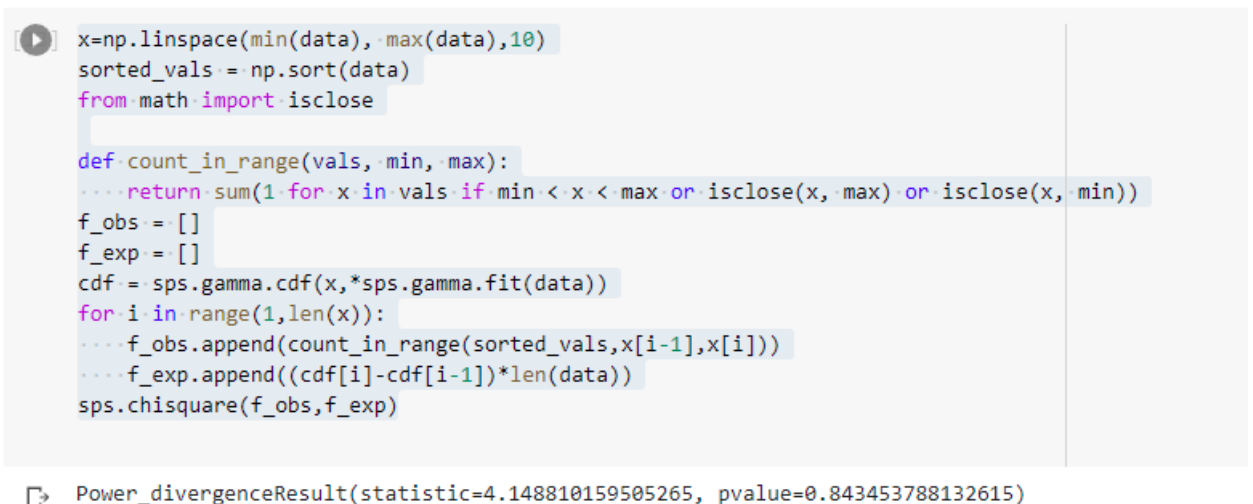
```
Ввод [26]: 1 print('normality =', sps.normaltest(data))
2
normality = NormaltestResult(statistic=0.992241343246913, pvalue=0.6088881611166078)
```

Исходя из значения вероятности (P-Value), для набора данных мы не можем отклонить нулевую гипотезу ($p > 0,05$). Следовательно, можно сделать вывод, что выборка подчиняется нормальному закону распределения.

Проверим два других распределения. Гамма и Фишера.



Получим значения P-value для двух распределений



Исходя из значения (P-Value), мы не можем отклонить нулевую гипотезу (Распределение Фишера / Распределение Гамма) ($p > 0,05$). Наиболее вероятностное распределение = Фишера