Задание 1.

Исходные данные:

Используя данные о хоккеистах проверьте, является ли среди хоккеистов из Финляндии, Норвегии и Дании значимым отличие: а) роста, б) BMI.

Решение:

```
Python 3.8.10 (default, Jun  2 2021, 10:49:15)
[GCC 9.4.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import numpy as np
>>> import pandas as pd
>>> def anova(*ys: np.ndarray) -> float:
...     y = np.concatenate(ys)
...     ss_b = sum((yi.mean() - y.mean()) ** 2 * yi.size for yi in ys)
...     ss_w = sum(((yi - yi.mean()) ** 2).sum() for yi in ys)
...     sigma_b = ss_b / (len(ys) - 1)
...     sigma_w = ss_w / (y.size - len(ys))
...     return sigma_b / sigma_w
...
>>> df = pd.read_csv('hockey.csv')
>>> df.head()
   year country  no                name  ...                 club        age  cohort        bmi
0  2001     RUS  10     tverdovsky oleg  ...   anaheim mighty ducks  24.952772    1976  24.543462
1  2001     RUS   2   vichnevsky vitali  ...   anaheim mighty ducks  21.119781    1980  24.332277
2  2001     RUS  26  petrochinin evgeni  ...  severstal cherepovetal  25.229295    1976  28.680111
3  2001     RUS  28     zhdan alexander  ...          ak bars kazan  29.675565    1971  26.827421
4  2001     RUS  32     orekhovsky oleg  ...          dynamo moscow  23.490760    1977  28.734694

[5 rows x 13 columns]
>>>
>>> df['country'].unique()
array(['RUS', 'AUT', 'BLR', 'CAN', 'CZE', 'FIN', 'GER', 'ITA', 'JPN',
       'LAT', 'NOR', 'SUI', 'SVK', 'SWE', 'UKR', 'USA', 'POL', 'SLO',
       'DEN', 'FRA', 'KAZ', 'HUN'], dtype=object)
>>>
>>> countries = ['FIN', 'NOR', 'DEN']
>>> ys = [df.loc[df['country'] == country, 'height'] for country in countries]
>>> f = anova(*ys)
>>> f
4.273207343917213
>>>
>>> k = len(countries)
>>> n = sum(y.size for y in ys)
>>> k1 = k - 1
>>> k2 = n - k
>>> print(k1, k2)
2 1023
>>> alpha = 0.05
>>> import scipy
>>> from scipy import stats
>>> t = stats.f.ppf(1 - alpha, k1, k2)
>>> print(t)
```

```
3.0045220661840073
>>>
>>> ys = [df.loc[df['country'] == country, 'bmi'] for country in countries]
>>> f = anova(*ys)
>>> f
36.058167277438855
>>>
>>>
```