

Задание.

1. Необходимо выбрать метрику и привести аргументацию.

По условию задачи целевая метрика равна какой-то из колонок, и тест заключается в том, чтобы проверить увеличение дохода. За доход как раз отвечает колонка NPV, поэтому метрика — среднее NPV.

```
In [ ]: import csv
import pandas as pd
df = pd.read_csv('hist_telesales.csv',
                 index_col='ID',
                 sep=',')
df
```

```
Out [ ]:
```

	Флаг дозвола	Флаг продажи	Расходы	PV	NPV
ID					
0	1	0	90	0	-90
1	0	0	5	0	-5
2	0	0	68	0	-68
3	1	0	22	0	-22
4	1	0	22	0	-22
...
72156	1	1	577	1346	769
72157	0	0	8	0	-8
72158	0	0	23	0	-23
72159	0	0	4	0	-4
72160	1	1	132	1385	1253

72161 rows x 5 columns

2. Альтернатива в критерии.

Альтернатива H_1 — уменьшение цены продукта позволит суммарно увеличить доходность продукта. То есть $H_1 : NPV_{\text{тест}} > NPV_{\text{контроль}}$

3.1. Каков размер выборки? Привести аргументацию и написать как получилось то или иное число.

```
In [ ]: from statistics import pvariance
import scipy.stats

a = 0.05
b = 0.2
sigma2 = pvariance(df.NPV.tolist())
```

```

z_1_a = scipy.stats.norm.ppf(1 - a)
z_b = scipy.stats.norm.ppf(b)
k = 0.5
MDE = 0.05 * df.NPV.mean()
N = sigma2 * (z_1_a - z_b)**2 / (k * MDE**2)
N

```

Out[]: 28671.429203969576

следовательно, минимальное N равно 28671

4. Принятие решения. Расписать подробно с аргументами.

Так как нужно сравнить средние, стоит выбор между применением критерия Стьюдента и критерием Манна-Уитни. Из документации к `scipy.stats.mannwhitneyu`: *The Mann-Whitney U test is a non-parametric version of the t-test for independent samples. When the means of samples from the populations are normally distributed, consider `scipy.stats.ttest_ind`.*

Поэтому проверим, являются ли обе выборки нормально распределёнными. Посмотрим на график:

```

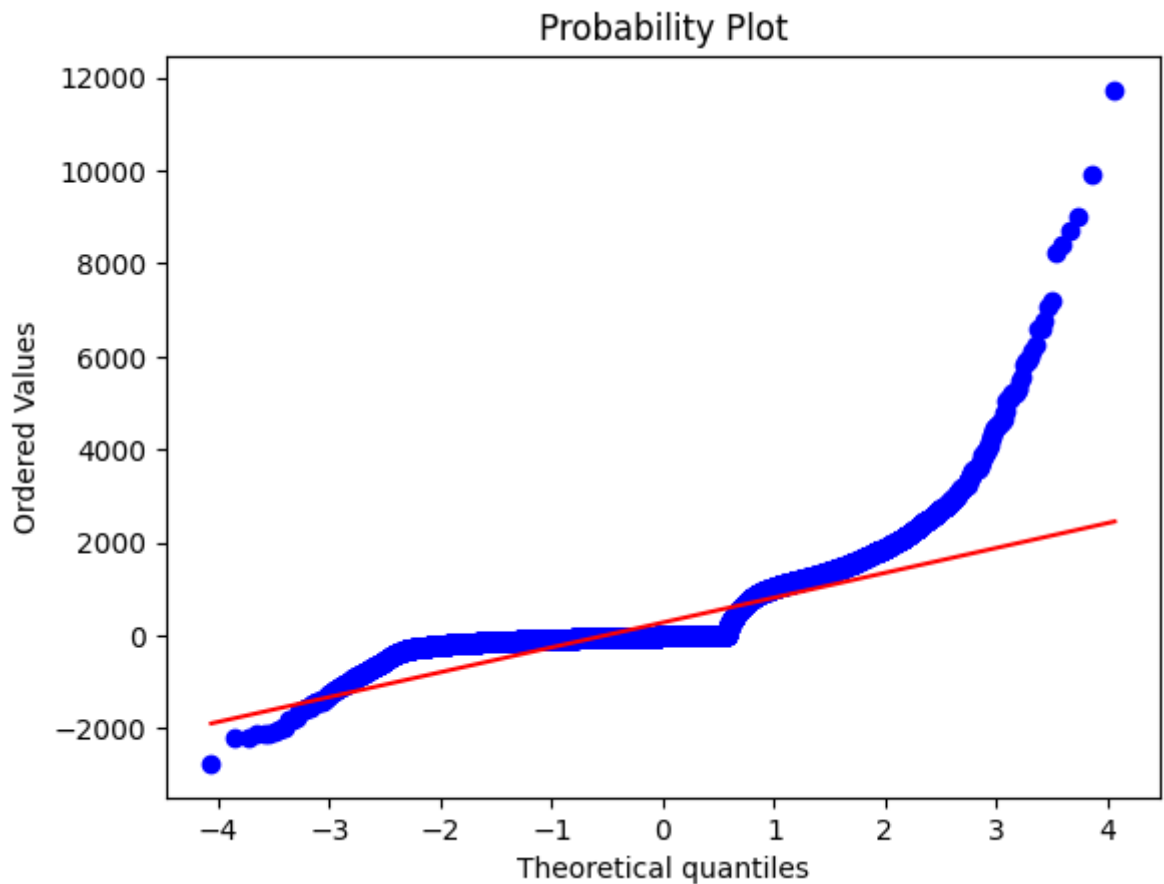
In [ ]: import pylab

df_control = pd.read_csv('Контроль.csv',
                        index_col='ID',
                        sep=',')

df_test = pd.read_csv('Тест.csv',
                    index_col='ID',
                    sep=',')

scipy.stats.probplot(df_test.NPV.to_numpy(), dist="norm", plot=pylab)
pylab.show()

```



Действительно, тестовая выборка не имеет нормальное распределение, поэтому используем критерий Манна Уитни

```
In [ ]: from scipy.stats import mannwhitneyu  
  
mannwhitneyu(df_control.NPV.tolist(), df_test.NPV.tolist(),  
              alternative = 'greater').pvalue
```

```
Out[ ]: 0.917100825637646
```

$pvalue > 0.05$, значит, мы не можем отвергнуть нулевую гипотезу теста.
Следовательно, мы не можем утверждать, что уменьшение цены продукта позволит суммарно увеличить доходность продукта.