

Задание 5

Борисов Д.А.

Вариант 1

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib
```

In [2]:

```
df = pd.read_excel('data_5.xlsx', header=0)
df.head()
```

Out[2]:

	1	2	3	4	5	6	7
0	40	25	45	30	35	40	60
1	80	35	70	40	45	30	50
2	55	70	60	70	65	45	80
3	60	110	90	70	80	75	70

In [3]:

```
# k - число столбцов df.shape[1]
# n - число строк df.shape[0]
k = df.shape[1]
n = df.shape[0]
```

In [10]:

```
df2 = pd.DataFrame()
df2 = df**2
# сумма квадратов всех x
Q1 = df2.sum().sum()
print('Q1 =', Q1)
df2.head()
```

Q1 = 105825

Out[10]:

	1	2	3	4	5	6	7
0	1600	625	2025	900	1225	1600	3600
1	6400	1225	4900	1600	2025	900	2500
2	3025	4900	3600	4900	4225	2025	6400
3	3600	12100	8100	4900	6400	5625	4900

In [18]:

```
# сумма по столбца
```

```
df_sum_column = df.sum()
df_2sum_column = df_sum_column**2
df_sum_column.head()
df_2sum_column.head()

# Сумма квадратов суммируемых столбцов деленных на всего строк
Q2 = df_2sum_column.sum() / n
print('Q2 =', round(Q2, 3))
```

Q2 = 95368.75

In [19]:

```
# квадрат суммы средних значений деленных на k и n
Q3 = (df_sum_column.sum() ** 2) / (n * k)
print('Q3 =', round(Q3, 3))
```

Q3 = 94308.036

In [21]:

```
S02 = (Q1 - Q2) / (k * (n - 1))
SA2 = (Q2 - Q3) / (k - 1)

print('S02 =', round(S02, 3), 'SA2 =', round(SA2, 3))
F_imper = SA2 / S02
print('F_imper =', round(F_imper, 3))
```

S02 = 497.917 SA2 = 176.786

F_imper = 0.355

In [22]:

```
# критические значения
from scipy.stats import f
alpha = 0.95
F_crit = f.sf(alpha, k - 1, k * (n - 1))

print('F_imper=', round(F_imper, 3), '\nF_crit=', round(F_crit, 3))
```

F_imper= 0.355

F_crit= 0.481

Вывод: так как **F**импер меньше **F**крити влияние фактора признается незначимым. Выборку считаем однородной

In [23]:

```
S2_ = (Q1 - Q3) / (k * n - 1)
print('дисперсия однородно выборки = ', round(S2_, 3))
```

дисперсия однородно выборки = 426.554

In []: