

**Лабораторная работа №5**  
**Выполнил: Кузнецов Павел М3207**  
**Вариант 1**

**Условие:**

Вариант 3. В файле приведены данные о мужской обуви.

1. Постройте линейную модель, где в качестве независимых переменных выступают расход в городе, расход на шоссе, мощность (вместе со свободным коэффициентом), зависимой – цена.

2. Проверьте следующие подозрения:

a) Чем больше мощность, тем больше цена

b) Цена изменяется в зависимости от расхода в городе

c) Проверьте гипотезу  $H_0$  о равенстве одновременно нулю коэффициентов при расходе в городе и расходе на шоссе против альтернативы  $H_1 = \bar{H}_0$

**Решение:**

Вычисляем оценки коэффициентов линейной модели:

```
1 import pandas as pd
2 import numpy
3 import scipy.stats as stats
4
5 data = pd.read_csv('/Users/frogge/proggs/mathstatLabs/lab-5/cars93.
    csv')
6
7
8 X = numpy.column_stack((numpy.ones(len(data)), data["MPG.city"],
    data["MPG.highway"], data["Horsepower"]))
9 Y = data['Price']
10
11 coefficients = numpy.linalg.inv(X.T.dot(X)).dot(X.T).dot(Y) #
    matrix method
```

Получившиеся коэффициенты:

overall: 6.68867892

city: -0.03860018  
highway: -0.17885863  
horsepower: 0.13131383

Остаточная дисперсия, доверительные интервалы для оценок коэффициентов:

```
1 resvariance = numpy.sum((Y - X.dot(coefficients))**2) / (len(data)
  - len(coefficients))
2
3 covariance = resvariance * numpy.linalg.inv(X.T.dot(X))
4 err = numpy.sqrt(numpy.diag(covariance))
5 t = stats.t.ppf(0.975, len(data) - len(coefficients))
6 dov_intervals = numpy.column_stack((coefficients - t * err,
  coefficients + t * err))
```

Получившиеся значения:

Остаточная дисперсия: 35.694685733186155

Доверительные интервалы для коэффициентов:

```
[[ -5.21020505  18.58756289]
 [ -0.74812245   0.67092209]
 [ -0.88350169   0.52578444]
 [  0.09931121   0.16331646]]
```

Вычисляем коэффициент детерминации:

```
1 NSM = numpy.sum((Y - numpy.mean(Y))**2)
2 ESM = numpy.sum((X.dot(coefficients) - numpy.mean(Y))**2)
3 R2 = ESM / NSM
```

Коэффициент детерминации: 0.6299138920082498

## 2. Проверка гипотез:

```
1 unneededv, p = stats.ttest_ind(Y, data["Horsepower"])
2 if p < 0.05:
3     print("first hypothesis incorrect")
4 else:
5     print("first hypothesis correct")
6
7
8 unneededv, p = stats.ttest_ind(Y, data["MPG.city"])
9 if p < 0.05:
10    print("second hypothesis incorrect")
11 else:
12    print("second hypothesis correct")
13
14 unneededv, pc = stats.f_oneway(Y, data["MPG.city"])
15 unneededv, ph = stats.f_oneway(Y, data["MPG.highway"])
16 if pc < 0.05 or ph < 0.05:
```

```
17     print("third hypothesis incorrect")
18 else:
19     print("third hypothesis correct")
```

Вывод:

Первая гипотеза отвергается: мощность влияет на цену

Вторая гипотеза отвергается: расход в городе влияет на цену

Третья гипотеза отвергается: цена зависит от расхода в городе и/или от расхода на шоссе