

## Практическое занятие №2

**Тема:** Модель парной линейной регрессии

**Цель работы:** освоить методику построения и анализа модели парной линейной регрессии с использованием метода наименьших квадратов (МНК).

**Алгоритм выполнения работы:**

1. Сначала выполните расчеты в Python для получения точных результатов.
2. Затем повторите ключевые расчеты в Excel для углубленного понимания.
3. Сравните результаты из обеих сред.
4. Сделайте выводы о качестве моделей и интерпретируйте результаты.

### 2. Исходные данные

**Правило выбора варианта:**

- **Вариант 1:** если последняя цифра пароля  $\in \{0, 1, 2, 3, 7, 9\}$
- **Вариант 2:** если последняя цифра пароля  $\in \{4, 5, 6, 8\}$

#### Вариант 1

**Вариант 1: Анализ зависимости успеваемости студентов**

**Исходные данные:**

исследуется зависимость между количеством часов самостоятельной работы в неделю (X) и средним баллом успеваемости (Y) для 15 студентов:

Студент	Часы самостоятельной работы (X)	Средний балл (Y)
1	5	3.2
2	8	3.5
3	10	3.7
4	12	3.9
5	15	4.1
6	18	4.3
7	20	4.4
8	22	4.6
9	25	4.7
10	28	4.8
11	30	4.9
12	32	5.0
13	35	5.1

14	38	5.2
15	40	5.3

## Вариант 2. Анализ зависимости цены квартиры от площади

### Исходные данные:

исследуется зависимость между площадью квартиры ( $X$ , м<sup>2</sup>) и ее ценой ( $Y$ , тыс. \$) на рынке недвижимости:

квартира	Площадь (X)	Цена (Y)
1	30	80
2	40	110
3	50	140
4	60	170
5	70	200
6	80	230
7	90	260
8	100	290
9	110	320
10	120	350
11	130	380
12	140	410
13	150	440

## Методические указания

### ЧАСТЬ 1: ВЫПОЛНЕНИЕ В PYTHON

```

"""
ЛАБОРАТОРНАЯ РАБОТА №2: ПАРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ
Этот код строит и анализирует модель парной линейной
регрессии
"""

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm
from scipy import stats

# БЛОК 1: Подготовка данных
# Пояснение: Замените данные на свои в соответствии с
вариантом
data = {
    'X': [12, 15, 18, 20, 22, 25, 28, 30, 32, 35, 38, 40],
    'Y': [56, 65, 74, 82, 88, 95, 104, 112, 118, 126, 135,
142]
}
df = pd.DataFrame(data)

```

```

# БЛОК 2: Предварительный анализ
# Пояснение: Изучаем взаимосвязь переменных перед
построением модели
print("=== ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ===")
print(f"Корреляция: {df['X'].corr(df['Y']):.4f}")

plt.figure(figsize=(10, 6))
plt.scatter(df['X'], df['Y'], alpha=0.7)
plt.title('Диаграмма рассеяния')
plt.xlabel('X (Независимая переменная)')
plt.ylabel('Y (Зависимая переменная)')
plt.grid(True)
plt.show()

# БЛОК 3: Построение регрессионной модели
# Пояснение: Используем МНК для оценки параметров модели
X = sm.add_constant(df['X']) # Добавляем константу
model = sm.OLS(df['Y'], X).fit()

print("\n=== РЕЗУЛЬТАТЫ РЕГРЕССИИ ===")
print(model.summary())

# БЛОК 4: Анализ остатков
# Пояснение: Проверяем выполнение предпосылок регрессии
residuals = model.resid
fitted = model.fittedvalues

plt.figure(figsize=(15, 5))

plt.subplot(1, 3, 1)
plt.scatter(fitted, residuals)
plt.axhline(y=0, color='red', linestyle='--')
plt.title('Остатки vs Предсказанные значения')
plt.xlabel('Предсказанные значения')
plt.ylabel('Остатки')

plt.subplot(1, 3, 2)
stats.probplot(residuals, dist="norm", plot=plt)
plt.title('Q-Q plot остатков')

plt.subplot(1, 3, 3)
plt.hist(residuals, bins=6, edgecolor='black')
plt.title('Гистограмма остатков')

plt.tight_layout()
plt.show()

```

## ЧАСТЬ 2: ВЫПОЛНЕНИЕ В EXCEL

### Шаг 1: Подготовка данных для расчетов

A (X)	B (Y)	C (X- X <sub>ср</sub> )	D (Y- Y <sub>ср</sub> )	E ((X- X <sub>ср</sub> )*(Y- Y <sub>ср</sub> )	F ((X- X <sub>ср</sub> ) <sup>2</sup> )	G (Предсказанные)	H (Остатки)
-------	-------	----------------------------	----------------------------	------------------------------------------------------	--------------------------------------------	-------------------	-------------

				Ycp))		Y)	

## Шаг 2: Расчет коэффициентов регрессии

I1: =CPЗНАЧ(A:A) # Среднее X

I2: =CPЗНАЧ(B:B) # Среднее Y

I3: =СУММ(E:E)/СУММ(F:F) # Коэффициент b1

I4: =I2-I3\*I1 # Коэффициент b0 (константа)

## Шаг 3: Расчет показателей качества модели

I5: =КВПИРСОН(B:B,A:A) # R-квадрат

I6: =СТОШУХ(B:B,A:A) # Стандартная ошибка регрессии

## Шаг 4: Расчет остатков

G2: =\$I\$4+\$I\$3\*A2 # Предсказанные значения

H2: =B2-G2 # Остатки

## ЧАСТЬ 3: СРАВНЕНИЕ РЕЗУЛЬТАТОВ

Параметр	Python	Excel	Совпадение
Коэффициент b0			
Коэффициент b1			
R-квадрат			
Стандартная ошибка			
t-статистика для b1			

## Справочные данные и формулы

### Основные формулы МНК:

- Наклон:  $b_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$
- Константа:  $b_0 = \bar{Y} - b_1 \bar{X}$
- Коэффициент детерминации:  $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$

#### **Проверка значимости:**

- Стандартная ошибка коэффициента:  $se(b_1) = \sqrt{\frac{s^2}{\sum(X_i - \bar{X})^2}}$
- t-статистика:  $t = \frac{b_1}{se(b_1)}$
- Число степеней свободы:  $df = n - 2$

#### **Критические значения t-распределения ( $\alpha=0.05$ ):**

- $df=10$ : 2.228
- $df=11$ : 2.201
- $df=12$ : 2.179

#### **Требования к отчету:**

1. **Титульный лист** (название работы, вариант, ФИО, группа)
2. **Постановка задачи** (цель анализа, описание данных)
3. **Исходные данные** (таблица)
4. **Результаты расчетов:**
  - Основные статистики.
  - Таблица сравнения результатов, полученных с применением Python и Excel.
  - Графики.
5. **Выводы** (интерпретация результатов, оценка качества модели)

#### **Требования к отчету для всех вариантов:**

1. **Титульный лист** (название работы, вариант, ФИО, группа)

2. **Постановка задачи** (формулировка исследуемой зависимости)
3. **Исходные данные** (таблица)
4. **Результаты расчетов:**
  - Основные статистики.
  - Таблица сравнения результатов, полученных с применением Python и Excel.
  - Графики.
5. **Выводы** (интерпретация результатов, оценка качества модели)

**Критерии оценки для получения зачета:**

**Зачет:**

- полностью выполнены все расчеты, построены все графики, дана грамотная интерпретация, ответы на контрольные вопросы
- выполнены основные расчеты, есть незначительные ошибки в интерпретации или оформлении графиков
- выполнены не все расчеты, имеются ошибки в методике анализа, возможно потребуется доработка

**Незачет:** работа не выполнена или выполнена неверно