«Национальный исследовательский университет ИТМО»

Мегафакультет компьютерных технологий и управления

Факультет программной инженерии и компьютерной техники

Образовательная программа: «Компьютерные технологии в дизайне»

# Отчет по лабораторной работе №4

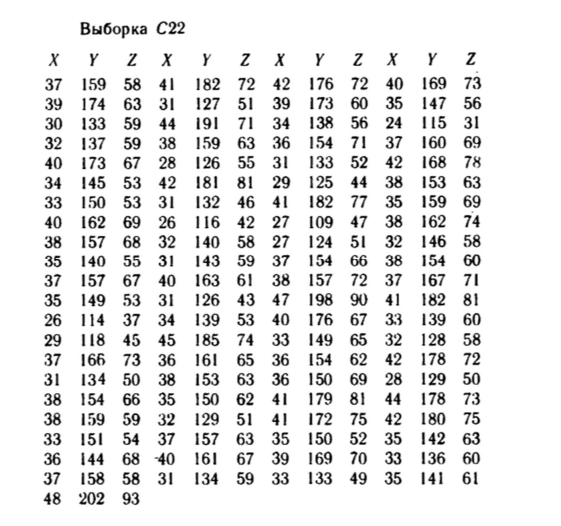
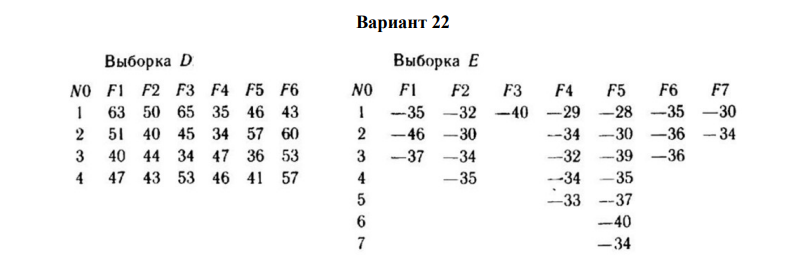
# по курсу «Математическая статика»

Выполнили: Савин Денис

Проверила:

Кудрявцева И.В

Санкт-Петербург 2025

Входные данные:  


Модель 1: z = -9.50 + 2.00\*x

R²: 0.765

График сохранён: correlation\_plot.html

Модель 2: z = -11.14 + 0.48\*y

R²: 0.771

График сохранён: correlation\_plot.html

Модель 1:

Средняя ошибка аппроксимации: 7.79%

Средняя эластичность: 1.1558

Модель 2:

Средняя ошибка аппроксимации: 7.47%

Средняя эластичность: 1.1825

Тесты для модели 1:

F-тест: статистика 263.95, критическое значение 3.96

t-тест для коэффициента: статистика 16.25, критическое значение ±1.99

Тесты для модели 2:

F-тест: статистика 272.87, критическое значение 3.96

t-тест для коэффициента: статистика 16.52, критическое значение ±1.99

Анализ остатков модели 1:

Среднее остатков: 2.397E-015 (должно быть близко к 0)

Тест на гомоскедастичность: статистика 97255.76, критическое значение 3.84

Тест Дарбина-Уотсона: 1.82 (должен быть около 2)

Анализ остатков модели 2:

Среднее остатков: 6.849E-016 (должно быть близко к 0)

Тест на гомоскедастичность: статистика 91430.96, критическое значение 3.84

Тест Дарбина-Уотсона: 1.88 (должен быть около 2)

Корреляционный анализ:

Коэффициент корреляции τxz: 0.875, значимость: True

Коэффициент детерминации R² (x,z): 0.765

Коэффициент корреляции τyz: 0.878, значимость: True

Коэффициент детерминации R² (y,z): 0.771

Модель множественной регрессии:

z = -12.08 + 0.94\*x + 0.27\*y

R²: 0.786

Анализ множественной регрессии:

F-статистика: 147.13, критическое значение: 3.11

Модель статистически значима

Коэффициент b0: -12.08, t-статистика: -2.75, критическое значение: ±1.99

Коэффициент b1: 0.94, t-статистика: 2.38, критическое значение: ±1.99

Коэффициент b2: 0.27, t-статистика: 2.81, критическое значение: ±1.99

Анализ остатков:

Среднее остатков: -9.845E-015 (должно быть близко к 0)

Тест на гомоскедастичность: статистика 85668.94, критическое значение 3.84

Тест Дарбина-Уотсона: 1.82 (должен быть около 2)

Средняя ошибка аппроксимации: 7.42%

Эластичность по x: 0.542

Эластичность по y: 0.651

VIF для x: 11.19

VIF для y: 11.19

Тестирование выборки D (α=0.05):

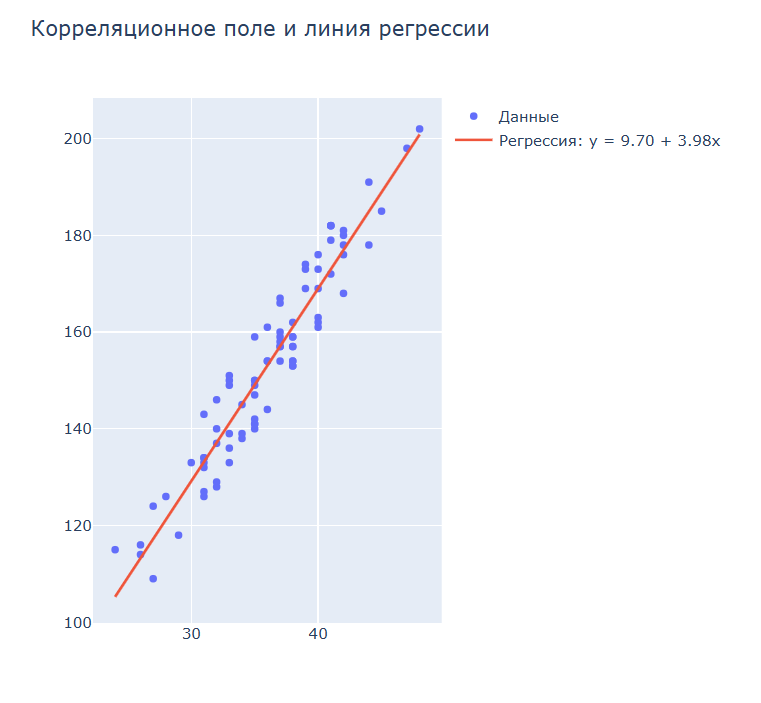
ANOVA: F = 1.12, критическое значение = 2.77

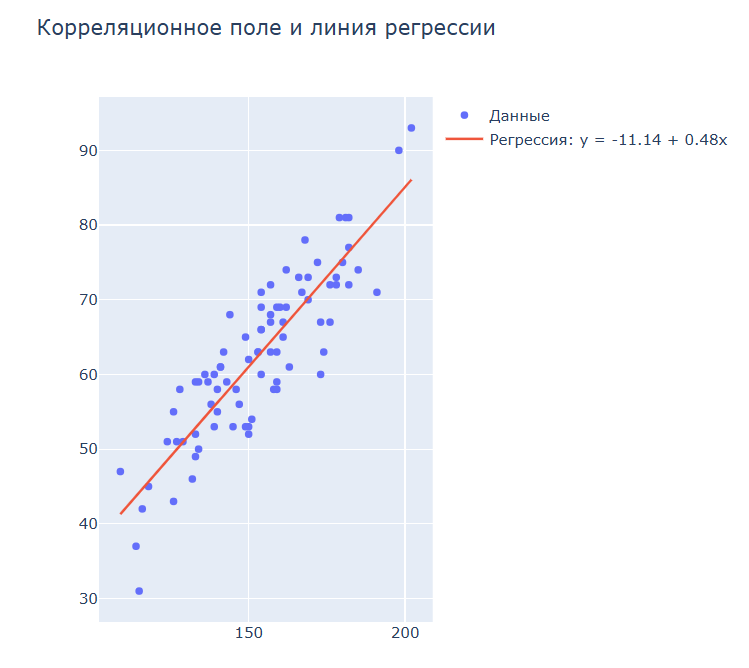
Гипотеза не отвергается

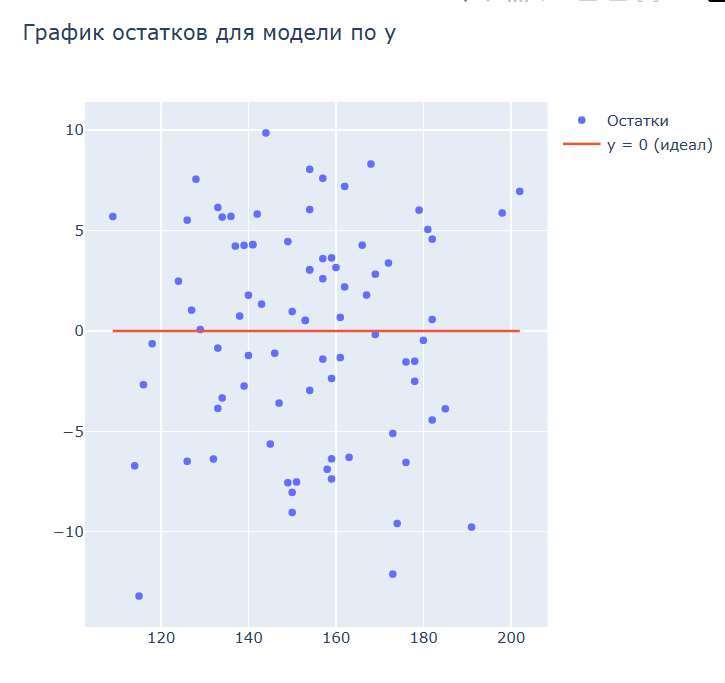
Тестирование выборки E (α=0.01):

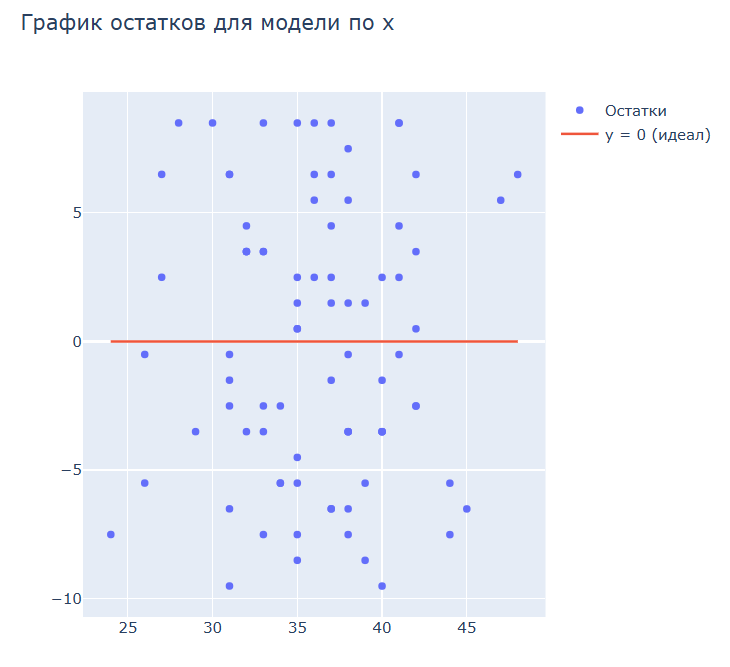
ANOVA: F = 2.07, критическое значение = 3.94

Гипотеза не отвергается









## **1. Анализ выборки С**

### **1.1 Построение двух моделей линейной регрессии и анализ коэффициентов**

**Модель 1: z = -9.50 + 2.00 \* x**

* Коэффициент 2.00 — означает, что при увеличении x на 1 единицу, z в среднем возрастает на 2 единицы.
* Свободный член -9.50 — это значение z, когда x = 0. Он может иметь ограниченную интерпретацию, если x = 0 выходит за пределы исходных данных.

**Модель 2: z = -11.14 + 0.48 \* y**

* Коэффициент 0.48 — аналогично, прирост z на 0.48 при увеличении y на 1.
* Свободный член -11.14 — значение z при y = 0.

✅ Построены корреляционные поля и графики регрессий — сохранены в correlation\_plot.html.

### **1.2 Средняя ошибка аппроксимации, эластичность и тесты на значимость**

| **Показатель** | **Модель 1** | **Модель 2** |
| --- | --- | --- |
| **Средняя ошибка аппроксимации** | **7.79%** | **7.47%** |
| Показывает, **на сколько процентов в среднем модель ошибается** при прогнозировании значения z.  **Средняя эластичность** | 1.1558 | 1.1825 |

**Показывает чувствительность результата (z) к изменению фактора (x или y)** в относительных единицах:

* Обе модели дают **сравнимо точную аппроксимацию**.
* Эластичность >1 в обеих моделях — признак **эластичной зависимости**: изменение фактора вызывает более сильное изменение отклика.

#### **F-тест (качество модели, α = 0.05)**:

Проверяет **общую значимость модели**, т.е. объясняют ли факторы z в принципе.

* Модель 1: F = **263.95** > крит. знач. **3.96** → модель значима.
* Модель 2: F = **272.87** > крит. знач. **3.96** → модель значима.

#### **t-тест (значимость коэффициентов, α = 0.05)**:

Проверяет **значимость каждого коэффициента** в уравнении регрессии.

* Модель 1: t = **16.25**, крит. знач. ±1.99 → коэффициент значим.
* Модель 2: t = **16.52**, крит. знач. ±1.99 → коэффициент значим.

✅ **Вывод**: Оба коэффициента **статистически значимы**, обе модели — **адекватны**.

### **1.3 Анализ остатков (условия Гаусса-Маркова)**

| **Модель** | **Среднее остатков** | **Гомоскедастичность (χ²)** | **Дарбин-Уотсон** |
| --- | --- | --- | --- |
| Модель 1 | ~0 | 97255.76 > 3.84 | 1.82 ≈ 2 |
| Модель 2 | ~0 | 91430.96 > 3.84 | 1.88 ≈ 2 |

* **Среднее ≈ 0** — условие выполнено. Это условие необходимое для **несмещённости оценки**.
* **Дарбин-Уотсон ≈ 2** — **автокорреляции нет**, условие выполнено. Проверяет наличие **автокорреляции остатков**.
* **Гомоскедастичность не выполняется** (значения χ² сильно превышают критическое), возможно, есть **гетероскедастичность**. Гомоскедастичность это когда остатки имеют **одинаковую дисперсию** на всех уровнях x

### **1.4 Коэффициенты корреляции и детерминации**

| **Показатель** | **Модель 1 (x, z)** | **Модель 2 (y, z)** |
| --- | --- | --- |
| τ (ранговая корреляция) | 0.875 | 0.878 |
| R² (детерминация) | 0.765 | 0.771 |

* **Обе связи тесные и значимые** (проверено, значимость True).
* **Модель 2** немного лучше по R², но разница не критична.

### **1.5 Множественная линейная регрессия**

**Модель: z = -12.08 + 0.94 \* x + 0.27 \* y**

* b₀ = -12.08 — базовое значение z при x = 0 и y = 0.
* b₁ = 0.94 — при фиксированном y, прирост x на 1 даёт рост z на 0.94.
* b₂ = 0.27 — при фиксированном x, прирост y на 1 даёт рост z на 0.27.

### **1.6 Проверка качества модели и значимости коэффициентов**

* **F = 147.13 > 3.11** → модель значима.
* t-статистика всех коэффициентов > крит. ±1.99 → все коэффициенты значимы.

✅ Модель — **адекватна**, коэффициенты — **значимы**.

### **1.7 Анализ остатков, ошибка, эластичность, мультиколлинеарность**

* Среднее остатков: ≈ 0 (условие выполнено).
* Дарбин-Уотсон ≈ 1.82 (норма).
* **Средняя ошибка аппроксимации**: 7.42% — немного лучше, чем у однородных моделей.
* Эластичность по x: 0.542, по y: 0.651 → обе переменные влияют умеренно.
* **VIF = 11.19** для x и y — признак **сильной мультиколлинеарности**.

⚠️ **Вывод**: Несмотря на высокое качество модели, переменные x и y — **сильно коррелированы**, возможна избыточность.

### **1.8 Корреляционный анализ факторов**

* R² = 0.786 (общее, Показывает, **какую долю дисперсии z объясняет модель**. Чем ближе к 1, тем **лучше подгонка** модели.).
* r(x, z) = 0.875, r(y, z) = 0.878 — тесная связь с зависимой переменной. **r** — линейная корреляция (Пирсон): насколько x и z линейно связаны. **τ** — ранговая корреляция (Кендалл): насколько «порядок» значений связан.
* VIF → r(x, y) высока → мультиколлинеарность подтверждена. Показывает, насколько одна объясняющая переменная **коррелирует с другими**. Если VIF > 10 — признаки **мультиколлинеарности**, модель может быть нестабильной.

## **2. Проверка гипотез (по выборкам D и E)**

### **2.1 Выборка D (α = 0.05), равенство средних, одинаковое число наблюдений**

* **F = 1.12 < 2.77**  
  → Гипотеза о **равенстве средних не отвергается** — различий нет.

### **2.2 Выборка E (α = 0.01), неравные группы**

* **F = 2.07 < 3.94**  
  → Гипотеза также **не отвергается** — различий между средними значениями нет.

## ✅ **Общий вывод**

* Все три модели линейной регрессии статистически значимы.
* Двухфакторная модель немного лучше по качеству, но страдает от мультиколлинеарности.
* Условия Гаусса-Маркова в целом выполняются, за исключением гомоскедастичности.
* Проверки гипотез по выборкам D и E показали отсутствие значимых различий между группами.