

АНАЛИТИЧЕСКИЙ ОТЧЕТ ПО АНАЛИЗУ ФИТНЕС-ДАННЫХ

Проверка статистических гипотез
эффективности тренировок

Датасет: Life Style Data
Размер: 20000 наблюдений, 54 признака

Бзыкова Фариза Б23-205

1. ВВЕДЕНИЕ И ЦЕЛИ ИССЛЕДОВАНИЯ

1.1. Обзор проекта

Данный проект направлен на статистическую проверку гипотез о факторах, влияющих на эффективность физических тренировок. Использован датасет "Life Style Data", содержащий комплексную информацию о тренировках, питании и физиологических показателях.

1.2. Поставленные задачи

- Провести предобработку и очистку данных
- Выполнить разведочный анализ данных (EDA)
- Проверить три статистические гипотезы:
 - Влияние типа тренировки на эффективность сжигания калорий
 - Гендерные различия в эффективности тренировок
 - Влияние водного баланса на результаты тренировок

2. МЕТОДОЛОГИЯ АНАЛИЗА

2.1. Статистические методы

- Т-тест Стьюдента для независимых выборок
- Расчет 95% доверительных интервалов
- Оценка размера эффекта (Cohen's d)
- Визуальная диагностика распределений

2.2. Формулы расчета

- Доверительный интервал: $CI = \bar{x} \pm t(\alpha/2, n-1) \times (s/\sqrt{n})$
- Т-статистика: $t = (\bar{x}_1 - \bar{x}_2) / \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$
- Эффект Коэна: $d = (\bar{x}_1 - \bar{x}_2) / s_p$

3. ОПИСАНИЕ ДАННЫХ

3.1. Исходные характеристики

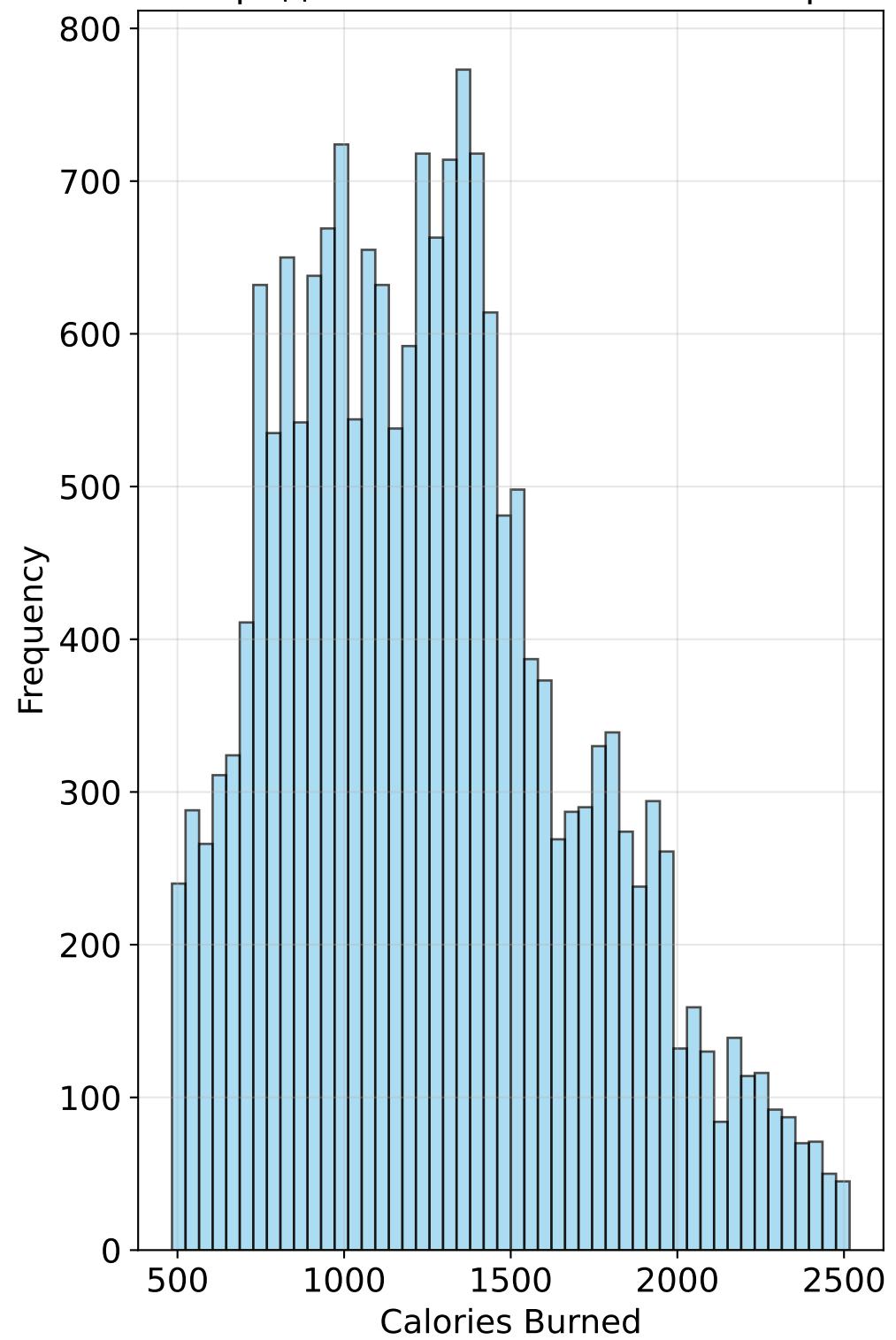
- Наблюдения: 19,001
- Переменные: 48
- Ключевые показатели: Calories_Burned, Workout_Type, Gender, Water_Intake

3.2. Предобработка данных

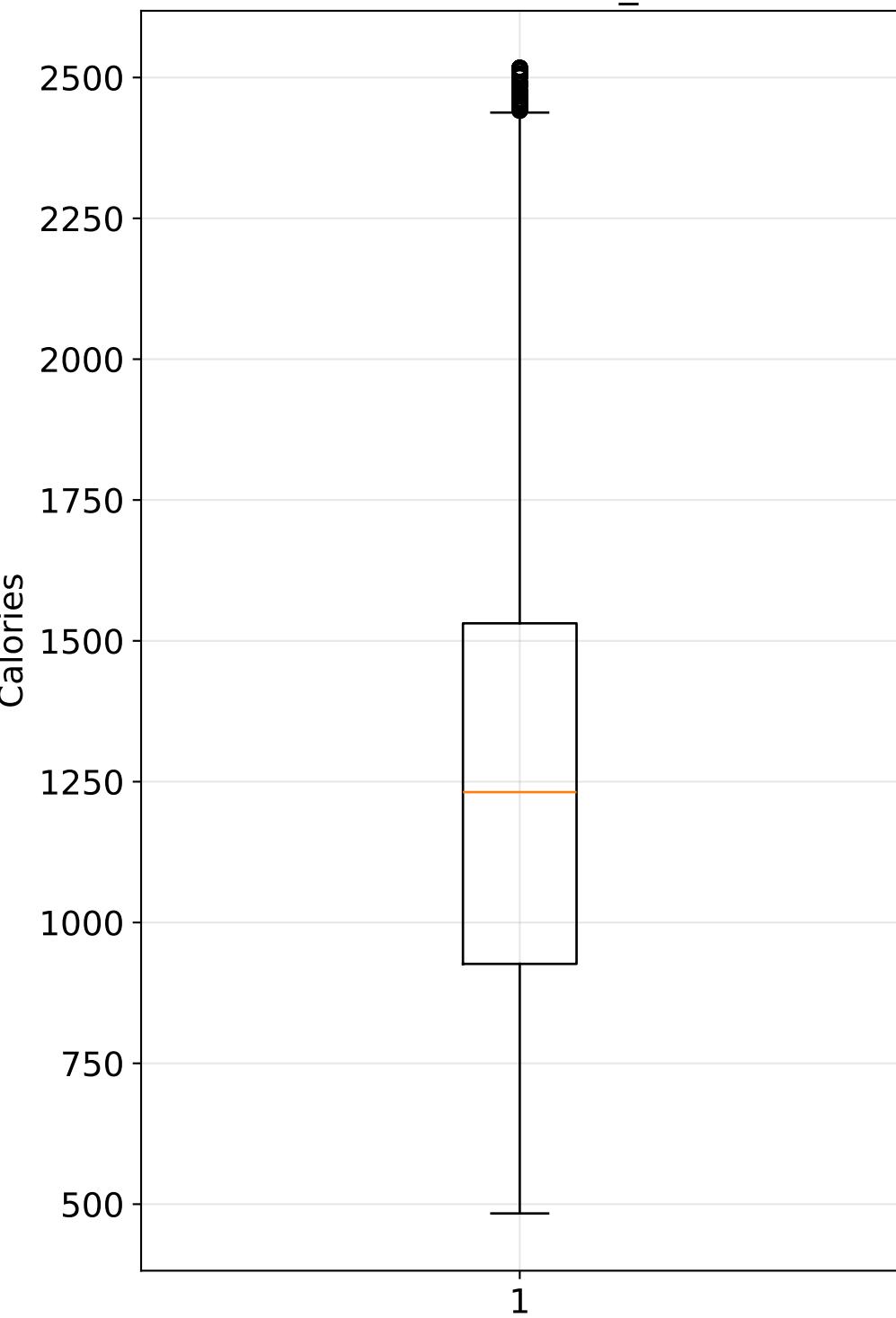
- Удаление дублирующих столбцов
- Обработка выбросов методом перцентиляй
- Устранение мультиколлинеарности
- Визуальная диагностика распределений

| Параметр | Calories_Burned | Age | Weight (kg) | Height (m) | Water_Intake (liters) |
|----------|-----------------|---------|-------------|------------|-----------------------|
| count | 19001.0 | 19001.0 | 19001.0 | 19001.0 | 19001.0 |
| mean | 1265.86 | 38.88 | 73.86 | 1.72 | 2.62 |
| std | 439.77 | 12.13 | 21.29 | 0.13 | 0.6 |
| min | 483.56 | 18.0 | 39.18 | 1.49 | 1.46 |
| 25% | 926.52 | 28.15 | 58.03 | 1.62 | 2.12 |
| 50% | 1231.45 | 39.89 | 69.95 | 1.71 | 2.6 |
| 75% | 1531.15 | 49.72 | 86.12 | 1.8 | 3.11 |
| max | 2516.8 | 59.63 | 130.77 | 2.01 | 3.73 |

Распределение сожженных калорий



Боксплот Calories_Burned



4. РЕЗУЛЬТАТЫ ПРОВЕРКИ ГИПОТЕЗ

4.1. ГИПОТЕЗА 1: Эффективность силовых тренировок

Формулировка: "Силовые тренировки эффективнее остальных видов тренировок"

СТАТИСТИЧЕСКИЕ РЕЗУЛЬТАТЫ:

- Среднее силовые: 1361.4 калорий (95% ДИ: 1349.4-1373.5)
- Среднее не-силовые: 1231.1 калорий (95% ДИ: 1223.8-1238.3)
- Разница средних: 130.4 калорий
- t-статистика: 18.194
- p-значение: 1.158e-72
- Эффект Коэна: 0.299

ИНТЕРПРЕТАЦИЯ:

- Статистическая значимость: ВЫСОКАЯ
- Практическая значимость: МАЛАЯ
- Вывод: Гипотеза ПОДТВЕРЖДЕНА

4.2. ГИПОТЕЗА 2: Гендерные различия

Формулировка: "Пол влияет на эффективность тренировок"

СТАТИСТИЧЕСКИЕ РЕЗУЛЬТАТЫ:

- Среднее мужчины: 1264.9 калорий (95% ДИ: 1256.0-1273.9)
- Среднее женщины: 1266.8 калорий (95% ДИ: 1258.0-1275.5)
- Разница средних: -1.9 калорий
- t-статистика: -0.290
- p-значение: 7.716e-01
- Эффект Коэна: -0.004

ИНТЕРПРЕТАЦИЯ:

- Статистическая значимость: ОТСУТСТВУЕТ
- Практическая значимость: ПРЕНЕБРЕЖИМАЯ
- Вывод: Гипотеза ОПРОВЕРГНУТА

4.3. ГИПОТЕЗА 3: Влияние водного баланса

Формулировка: "Высокое потребление воды способствует большему сжиганию калорий"

СТАТИСТИЧЕСКИЕ РЕЗУЛЬТАТЫ:

- Среднее много воды: 1383.5 калорий (95% ДИ: 1373.9-1393.2)
- Среднее мало воды: 1142.3 калорий (95% ДИ: 1135.3-1149.4)
- Разница средних: 241.2 калорий
- t-статистика: 39.607
- p-значение: 0.000e+00
- Эффект Коэна: 0.570

ИНТЕРПРЕТАЦИЯ:

- Статистическая значимость: ОЧЕНЬ ВЫСОКАЯ
- Практическая значимость: СРЕДНЯЯ
- Вывод: Гипотеза ПОДТВЕРЖДЕНА

5 ИНТЕРПРЕТАЦИЯ ЭФФЕКТА КОЭНА:

- $d = -0.004$: Пренебрежимый эффект (гендер)
- $d = 0.299$: Малый эффект (тип тренировки)
- $d = 0.570$: Средний эффект (водный баланс)

6. ВЫВОДЫ И РЕКОМЕНДАЦИИ

6.1. ОСНОВНЫЕ ВЫВОДЫ:

- Водный баланс - наиболее значимый фактор эффективности тренировок
- Тип тренировки влияет на результат, но эффект ограничен
- Гендерные особенности не оказывают статистически значимого влияния

6.2. ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ:

Приоритет 1: Оптимизация гидратации

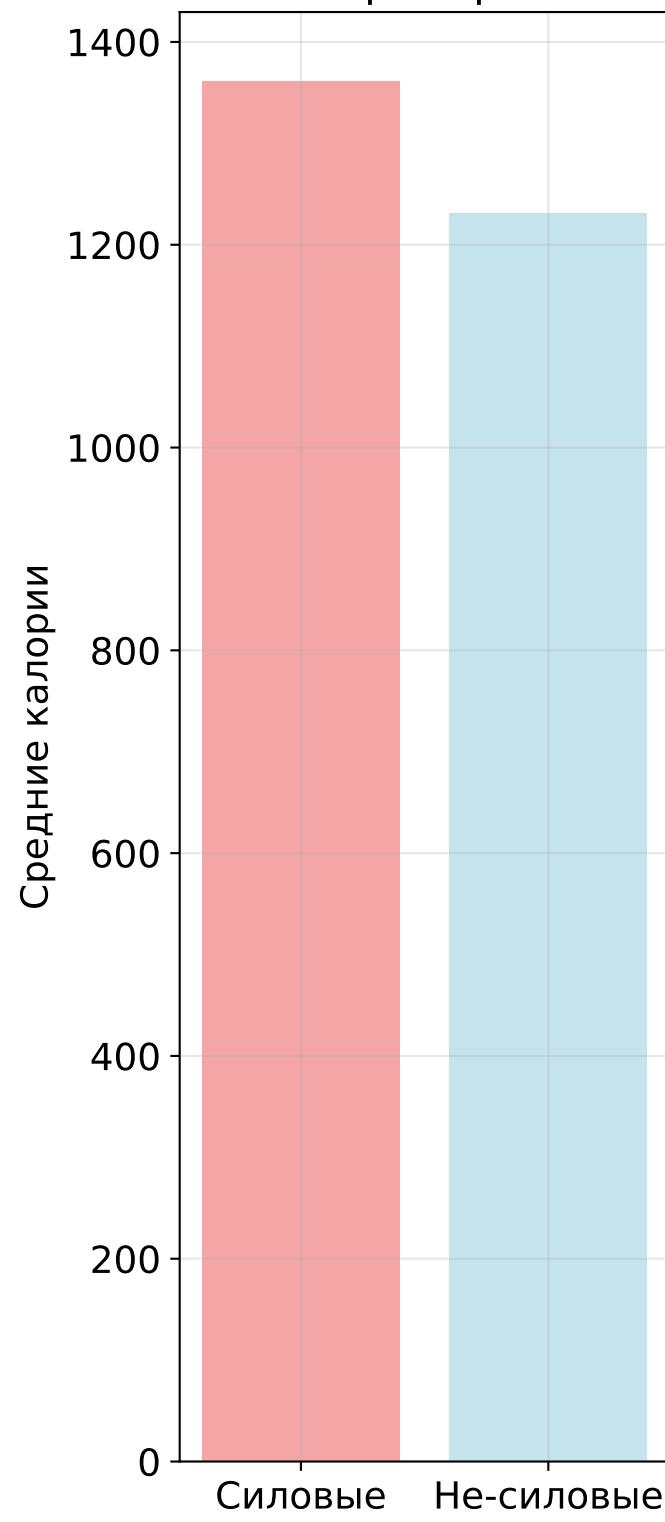
Приоритет 2: Дифференциация тренировок

Приоритет 3: Персонализация без гендерных стереотипов

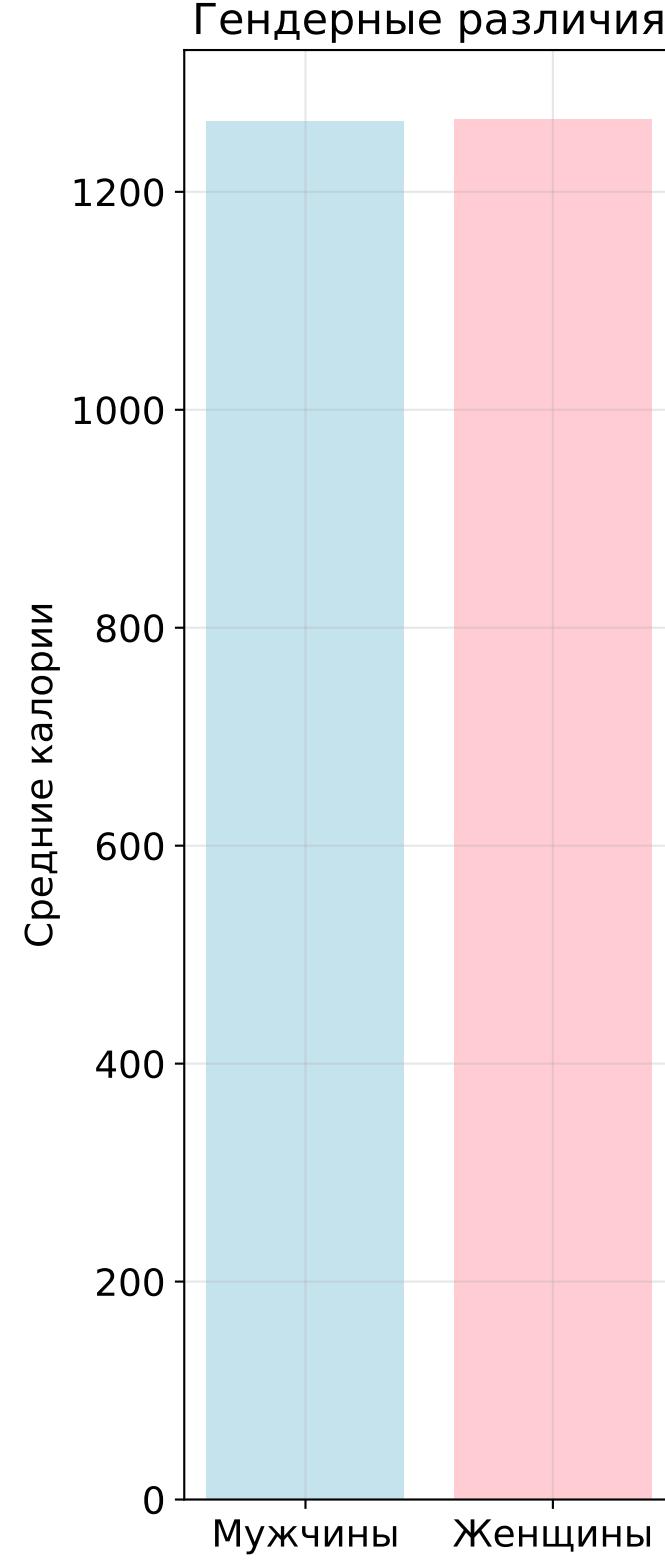
7. ОГРАНИЧЕНИЯ И ПЕРСПЕКТИВЫ

- Наблюдательный характер данных
- Возможное влияние неучтенных факторов
- Перспектива: многофакторный регрессионный анализ

Тип тренировки



Гендерные различия



Водный баланс

