

Практическое занятие №1.

Тема: Описательный анализ исходных данных

Цель работы:

Освоить методы первичного статистического анализа данных, научиться вычислять основные описательные статистики, строить графики распределения и делать предварительные выводы о характере данных.

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.

Правило выбора варианта:

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

Вариант 1: Анализ данных о заработной плате

Исходные данные:

Ежемесячная заработная плата 25 сотрудников компании (в тыс. руб.):

Сотрудник	Зарплата
1	45
2	52
3	48
4	65
5	38
6	72
7	55
8	42
9	58
10	68
11	75
12	62
13	47
14	53

15	71
16	49
17	56
18	64
19	59
20	51
21	44
22	69
23	57
24	61
25	54

Вариант 2: Анализ успеваемости студентов

Исходные данные:

Средний балл успеваемости 30 студентов за семестр:

Студент	Средний балл
1	4.2
2	3.8
3	4.5
4	3.9
5	4.1
6	3.7
7	4.3
8	4.6
9	3.5
10	4.4
11	4.0
12	3.6
13	4.7
14	3.4
15	4.8
16	3.3
17	4.9
18	3.2
19	5.0
20	3.1
21	4.2
22	3.9
23	4.1
24	3.8
25	4.3
26	4.5
27	3.7
28	4.6
29	3.6
30	4.4

Вариант 3: Анализ времени доставки товара

Исходные данные:

Время доставки товара (в днях) для 35 заказов:

Заказ	Время доставки
1	3
2	5
3	4
4	6
5	2
6	7
7	4
8	3
9	5
10	8
11	6
12	4
13	5
14	3
15	7
16	9
17	4
18	6
19	5
20	3
21	8
22	4
23	6
24	5
25	7
26	10
27	4
28	6
29	5
30	3
31	8
32	4
33	7
34	5
35	6

Методические указания

ЧАСТЬ 1: ВЫПОЛНЕНИЕ В PYTHON

Пояснение: Данный код автоматически выполняет все необходимые расчеты. Вам нужно последовательно запускать блоки кода и анализировать результаты.

```
"""
ЛАБОРАТОРНАЯ РАБОТА №1: ОПИСАТЕЛЬНЫЙ АНАЛИЗ
Этот код выполняет полный описательный анализ данных
"""

# БЛОК 1: Импорт библиотек
# Пояснение: Эти библиотеки содержат готовые функции для
статистического анализа
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats

# БЛОК 2: Загрузка данных согласно вашему варианту
# Пояснение: Замените данные на свои в соответствии с
вариантом
data = [45, 52, 48, 65, 38, 72, 55, 42, 58, 68, 75, 62, 47,
53, 71,
        49, 56, 64, 59, 51, 44, 69, 57, 61, 54]

df = pd.DataFrame(data, columns=['Зарплата'])

print("=== ОСНОВНЫЕ СТАТИСТИКИ ===")
print(df.describe())

# БЛОК 3: Расчет дополнительных статистик
# Пояснение: Эти показатели помогают описать распределение
данных
print("\n=== ДОПОЛНИТЕЛЬНЫЕ СТАТИСТИКИ ===")
print(f"Медиана: {df['Зарплата'].median():.2f}")
print(f"Стандартное отклонение: {df['Зарплата'].std():.2f}")
print(f"Дисперсия: {df['Зарплата'].var():.2f}")
print(f"Коэффициент вариации:
{(df['Зарплата'].std()/df['Зарплата'].mean()*100):.2f}%")
print(f"Асимметрия: {stats.skew(df['Зарплата']):.4f}")
print(f"Эксцесс: {stats.kurtosis(df['Зарплата']):.4f}")

# БЛОК 4: Визуализация данных
# Пояснение: Графики помогают визуально оценить
распределение данных
plt.figure(figsize=(15, 5))

plt.subplot(1, 3, 1)
plt.hist(df['Зарплата'], bins=8, edgecolor='black',
alpha=0.7)
plt.title('Гистограмма распределения')
```

```
plt.xlabel('Зарплата')
plt.ylabel('Частота')

plt.subplot(1, 3, 2)
plt.boxplot(df['Зарплата'])
plt.title('Box-plot')
plt.ylabel('Зарплата')

plt.subplot(1, 3, 3)
stats.probplot(df['Зарплата'], dist="norm", plot=plt)
plt.title('Q-Q plot')

plt.tight_layout()
plt.show()
```

ЧАСТЬ 2: ВЫПОЛНЕНИЕ В EXCEL

Пояснение: в Excel мы повторим ключевые расчеты, чтобы понять "механику" методов.

Шаг 1. Подготовка данных

Создайте таблицу:

A (Данные)	B (Отклонение от среднего)	C (Квадрат отклонения)
45	=A2-\$E\$1	=B2^2

Шаг 2. Расчет основных статистик

E1: =СРЗНАЧ(A:A) # Среднее значение
E2: =МЕДИАНА(A:A) # Медиана
E3: =СТАНДОТКЛОН.В(A:A) # Стандартное отклонение
E4: =ДИСП.В(A:A) # Дисперсия
E5: =E3/E1*100 # Коэффициент вариации (%)
E6: =МИН(A:A) # Минимум
E7: =МАКС(A:A) # Максимум
E8: =СЧЁТ(A:A) # Количество наблюдений

Шаг 3. Построение графиков

- **Гистограмма:** Вставка → Диаграммы → Гистограмма
- **Box-plot:** Вставка → Диаграммы → Ящик с усами (в Excel 2016+)

ЧАСТЬ 3: СРАВНЕНИЕ РЕЗУЛЬТАТОВ

Заполните таблицу сравнения:

Параметр	Python	Excel	Совпадение
Среднее			
Медиана			
Стандартное отклонение			
Дисперсия			
Коэффициент вариации			

В столбце «Совпадение» поставить знак «+» или «-» в зависимости от результата.

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, характеристика распределения)
6. **Ответы на контрольные вопросы**

Контрольные вопросы:

1. В чем разница между средним арифметическим и медианой?
2. Когда коэффициент вариации считается высоким?
3. Что показывает межквартильный размах?
4. Как интерпретировать коэффициенты асимметрии и эксцесса?
5. При каких условиях распределение считается нормальным?
6. Как определяются выбросы в box-plot?

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