

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ  
«БРЕСТСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

ФАКУЛЬТЕТ ЭЛЕКТРОННО-ИНФОРМАЦИОННЫХ СИСТЕМ

Кафедра интеллектуальных информационных технологий

## Отчет по лабораторной работе №1

Специальность АС-65

Выполнил:  
Д. А. Чмель,  
студент группы АС-65  
Проверил  
А. А. Крощенко,  
ст. преп. кафедры ИИТ,  
«    » \_\_\_\_\_ 2025 г.

Брест 2025

Цель работы: Получить практические навыки работы с данными с использованием библиотек Pandas для манипуляции и Matplotlib для визуализации. Научиться выполнять основные шаги предварительной обработки данных, такие как очистка, нормализация и работа с различными типами признаков.

## Задание 2. Загрузите данные и проверьте, есть ли в них пропущенные значения.

Выполнение:

Код программы

```
from google.colab import files
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler
import io

pd.set_option('display.max_columns', None)
plt.rcParams['figure.figsize'] = (10, 6)
plt.style.use('seaborn-v0_8-whitegrid')

print("АНАЛИЗ ДАННЫХ IRIS")

uploaded = files.upload()

#читаем данные
file_name = list(uploaded.keys())[0]
df = pd.read_csv(io.BytesIO(uploaded[file_name]))

print(df.head(50))

print("\nПроверка пропущенных значений:")
missing_values = df.isnull().sum()
print(missing_values)
```

Проверка пропущенных значений:

```
sepal.length    0
sepal.width     0
petal.length    0
petal.width     0
variety         0
dtype: int64
```

Основные статистические характеристики:

	sepal.length	sepal.width	petal.length	petal.width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

## Задание 2. 2. Выведите количество образцов каждого вида ириса.

```
print("\n" + "=" * 60)
print("2. КОЛИЧЕСТВО ОБРАЗЦОВ КАЖДОГО ВИДА ИРИСА")
print("-" * 50)

species_count = df['variety'].value_counts()
print("Количество образцов по видам:")
print(species_count)
```

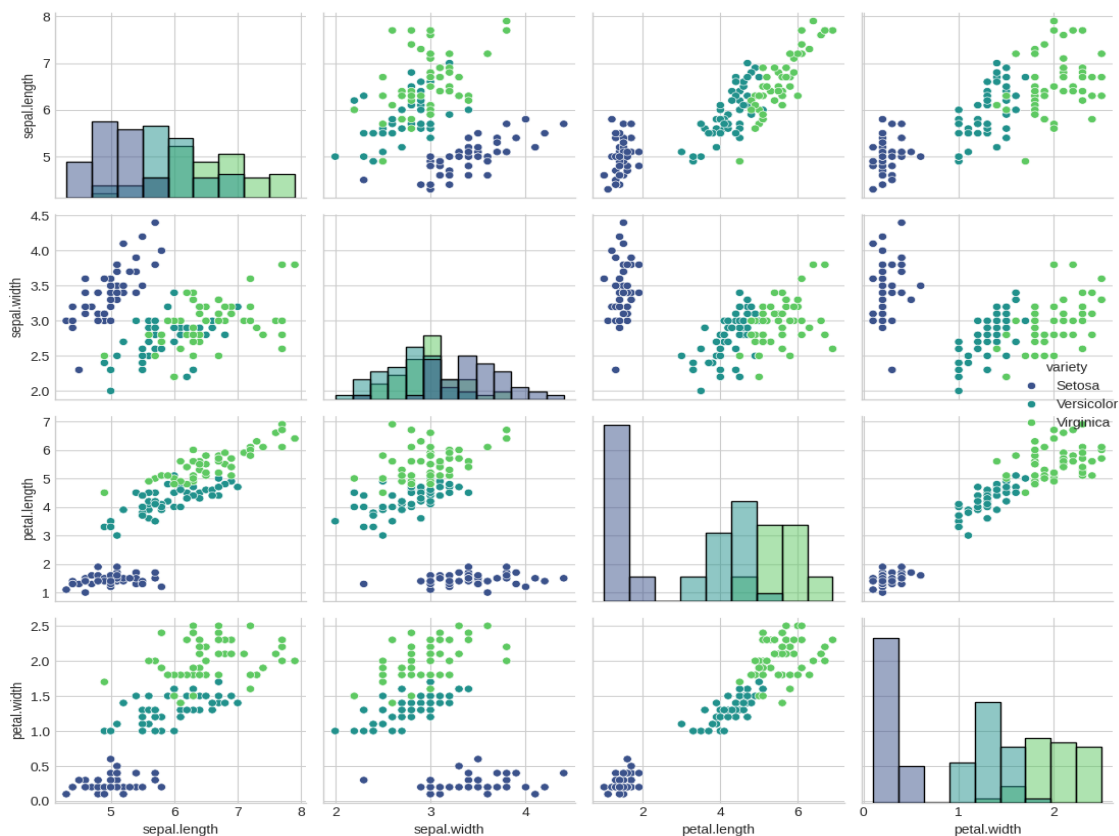
```
=====
2. КОЛИЧЕСТВО ОБРАЗЦОВ КАЖДОГО ВИДА ИРИСА
-----
Количество образцов по видам:
variety
Setosa      50
Versicolor  50
Virginica   50
Name: count, dtype: int64
```

## Задание 3. Постройте парные диаграммы рассеяния (pair plot) для всех признаков, чтобы визуально оценить их разделимость.

```
print("Построение парных диаграмм рассеяния")

# Создание pair plot
sns.pairplot(df, hue='variety', palette='viridis', diag_kind='hist', height=2.5)
plt.suptitle('Парные диаграммы рассеяния для признаков ирисов', y=1.02)
plt.tight_layout()
plt.show()
```

Парные диаграммы рассеяния для признаков ирисов



**Задание 4.** Для каждого вида ириса рассчитайте среднее значение по каждому из четырех признаков.

```
print("4. СРЕДНИЕ ЗНАЧЕНИЯ ПО ПРИЗНАКАМ ДЛЯ КАЖДОГО ВИДА")

numeric_columns = ['sepal.length', 'sepal.width', 'petal.length', 'petal.width']

mean_by_species = df.groupby('variety')[numeric_columns].mean()
std_by_species = df.groupby('variety')[numeric_columns].std()

print("Средние значения по видам:")
print(mean_by_species)

print("\nСтандартные отклонения по видам:")
print(std_by_species)
```

```
4. СРЕДНИЕ ЗНАЧЕНИЯ ПО ПРИЗНАКАМ ДЛЯ КАЖДОГО ВИДА
Средние значения по видам:
```

	sepal.length	sepal.width	petal.length	petal.width
variety				
Setosa	5.006	3.428	1.462	0.246
Versicolor	5.936	2.770	4.260	1.326
Virginica	6.588	2.974	5.552	2.026

```
Стандартные отклонения по видам:
```

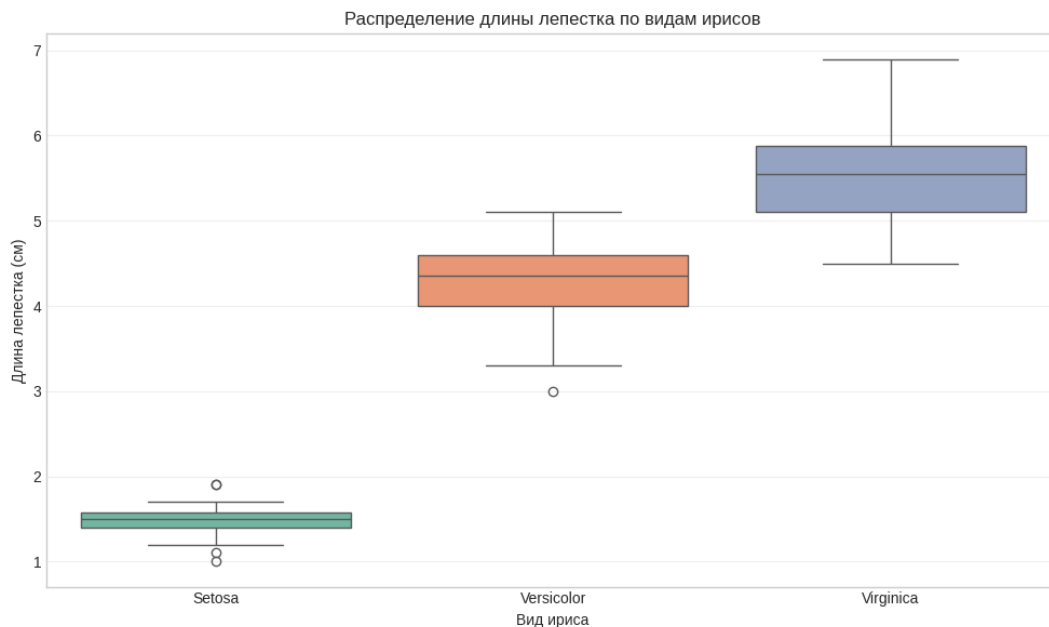
	sepal.length	sepal.width	petal.length	petal.width
variety				
Setosa	0.352490	0.379064	0.173664	0.105386
Versicolor	0.516171	0.313798	0.469911	0.197753
Virginica	0.635880	0.322497	0.551895	0.274650

**Задание 5.** Создайте "ящик с усами" (box plot) для признака Petal Length (cm), чтобы сравнить его распределение по разным видам ирисов.

```
print("5. ЯЩИК С УСАМИ ДЛЯ PETAL LENGTH")

plt.figure(figsize=(10, 6))
sns.boxplot(data=df, x='variety', y='petal.length', palette='Set2')
plt.title('Распределение длины лепестка по видам ирисов')
plt.xlabel('Вид ириса')
plt.ylabel('Длина лепестка (см)')
plt.grid(axis='y', alpha=0.3)
plt.tight_layout()
plt.show()

petal_stats = df.groupby('variety')['petal.length'].describe()
print(petal_stats)
```



**Задание 6. Стандартизируйте данные (приведите к нулевому среднему и единичному стандартному отклонению).**

```
print("\n6. СТАНДАРТИЗАЦИЯ ДАННЫХ")
print("-" * 50)

numeric_columns = ['sepal.length', 'sepal.width', 'petal.length', 'petal.width']

# Сохраняем исходные данные
original_data = df[numeric_columns].copy()

# Стандартизация
scaler = StandardScaler()
df_standardized = df.copy()
df_standardized[numeric_columns] = scaler.fit_transform(df[numeric_columns])

print("Результат стандартизации:")
print("-" * 60)
for col in numeric_columns:
    orig_mean, orig_std = original_data[col].mean(), original_data[col].std()
    std_mean, std_std = df_standardized[col].mean(), df_standardized[col].std()
    print(f"{col:15} |  $\mu$ : {orig_mean:6.2f}  $\rightarrow$  {std_mean:7.2f} |  $\sigma$ : {orig_std:5.2f}  $\rightarrow$  {std_std:5.2f}")
```

Результат стандартизации:

sepal.length	$\mu$ : 5.84 $\rightarrow$ -0.00   $\sigma$ : 0.83 $\rightarrow$ 1.00
sepal.width	$\mu$ : 3.06 $\rightarrow$ -0.00   $\sigma$ : 0.44 $\rightarrow$ 1.00
petal.length	$\mu$ : 3.76 $\rightarrow$ -0.00   $\sigma$ : 1.77 $\rightarrow$ 1.00
petal.width	$\mu$ : 1.20 $\rightarrow$ -0.00   $\sigma$ : 0.76 $\rightarrow$ 1.00

```

print("\n7. ВИЗУАЛИЗАЦИЯ СТАНДАРТИЗАЦИИ")
print("-" * 50)

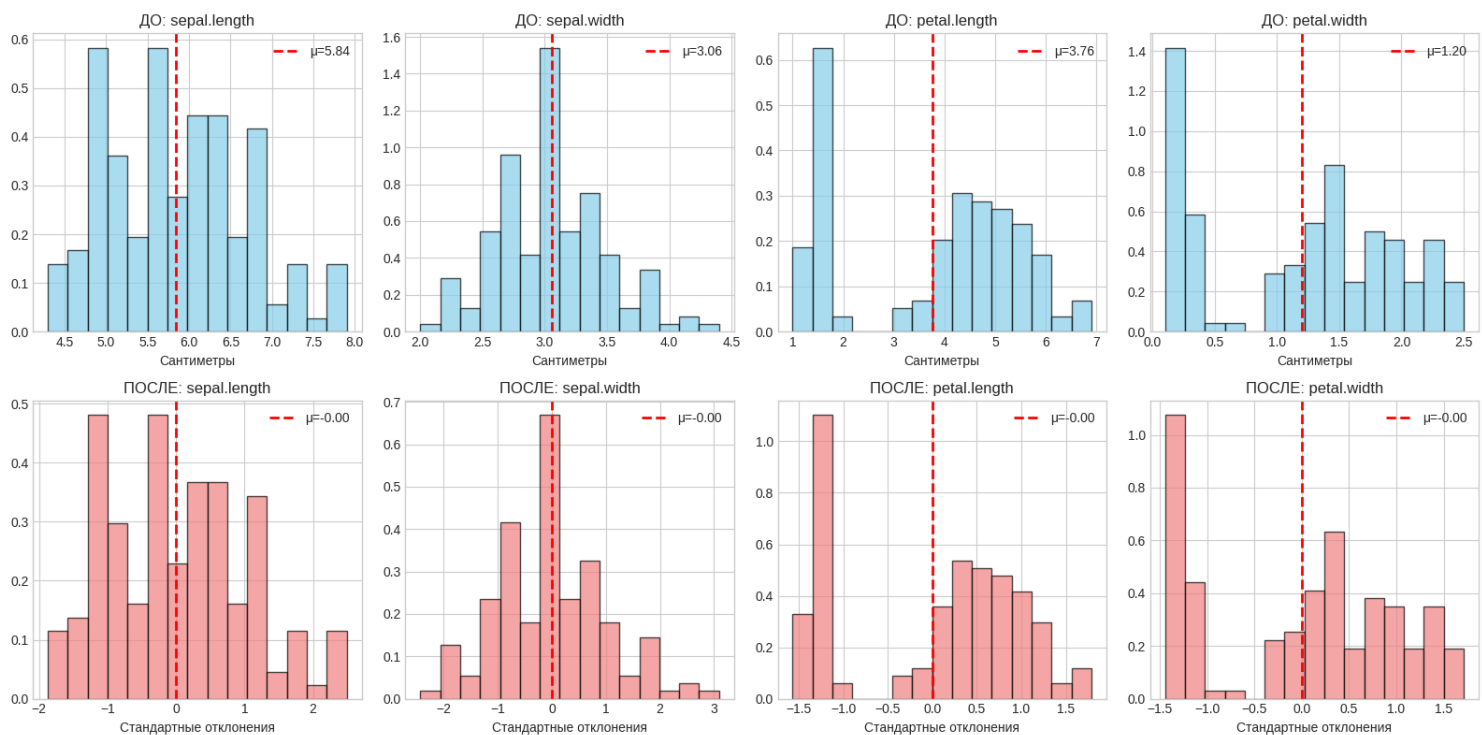
fig, axes = plt.subplots(2, 4, figsize=(16, 8))

for i, col in enumerate(numeric_columns):
    # До стандартизации
    axes[0, i].hist(original_data[col], alpha=0.7, color='skyblue', bins=15,
                    edgecolor='black', density=True)
    axes[0, i].set_title(f'ДО: {col}')
    axes[0, i].set_xlabel('Сантиметры')
    axes[0, i].axvline(original_data[col].mean(), color='red', linestyle='--',
                       linewidth=2, label=f' $\mu$ ={original_data[col].mean():.2f}')
    axes[0, i].legend()

    # После стандартизации
    axes[1, i].hist(df_standardized[col], alpha=0.7, color='lightcoral', bins=15,
                   edgecolor='black', density=True)
    axes[1, i].set_title(f'ПОСЛЕ: {col}')
    axes[1, i].set_xlabel('Стандартные отклонения')
    axes[1, i].axvline(df_standardized[col].mean(), color='red', linestyle='--',
                       linewidth=2, label=f' $\mu$ ={df_standardized[col].mean():.2f}')
    axes[1, i].legend()

plt.tight_layout()
plt.show()

```



## Выводы:

- 1) Качество данных: Набор данных Iris не содержит пропущенных значений и готов к дальнейшему анализу.
- 2) Разделимость классов: Виды ирисов хорошо разделяются по признакам, особенно по параметрам лепестков (petal.length и petal.width)
- 3) Статистические закономерности:
  - а) Setosa имеет наименьшие размеры лепестков
  - б) Virginica обладает наибольшими размерами цветка
  - в) Versicolor занимает промежуточное положение
- 4) Визуализация: Построенные графики эффективно демонстрируют различия между видами и распределения признаков.