|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра прикладной математики (ПМ)**

**ОТЧЕТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ**

по дисциплине «Технологии и инструментарий анализа больших данных»

**Практическое занятие №4**

|  |  |  |  |
| --- | --- | --- | --- |
| Студент группы ИНБО-01-17 | *ИНБО-07-21, Веригин Никита Алексеевич* | (подпись) | |
| Преподаватель | *Царёв Роман Юрьевич* | (подпись) | |
| Отчет представлен | «\_\_\_»\_\_\_\_\_\_\_\_2024 г. | |  | |

Москва 2024 г.

Введение

Цели работы:

1. Определить два вектора, представляющие собой число автомобилей, припаркованных в течении 5 рабочих дней у бизнес центра на уличной стоянке и в подземном гараже.
   1. Найти и интерпретировать корреляцию между переменными «Улица» и «Гараж» (подсчитать корреляцию по Пирсону).
   2. Построить диаграмму рассеяния для вышеупомянутых переменных.
2. Найти и выгрузить данные. Вывести, провести предобработку и описать признаки.
   1. Построить корреляционную матрицу по одной целевой переменной. Определить наиболее коррелирующую переменную, продолжить с ней работу в следующем пункте.
   2. Реализовать регрессию вручную, отобразить наклон, сдвиг и MSE.
   3. Визуализировать регрессию на графике.
3. Загрузить данные: 'insurance.csv'. Вывести и провести предобработку. Вывести список уникальных регионов.
   1. Выполнить однофакторный ANOVA тест, чтобы проверить влияние региона на индекс массы тела (BMI), используя первый способ, через библиотеку Scipy.
   2. Выполнить однофакторный ANOVA тест, чтобы проверить влияние региона на индекс массы тела (BMI), используя второй способ, с помощью функции anova\_lm() из библиотеки statsmodels.
   3. С помощью t критерия Стьюдента перебрать все пары. Определить поправку Бонферрони. Сделать выводы.
   4. Выполнить пост-хок тесты Тьюки и построить график.
   5. Выполнить двухфакторный ANOVA тест, чтобы проверить влияние региона и пола на индекс массы тела (BMI), используя функцию anova\_lm() из библиотеки statsmodels.
   6. Выполнить пост-хок тесты Тьюки и построить график. 4. Оформить отчет о проделанной работе, написать выводы

ХОД РАБОТЫ

Код программ и результаты их выполнения представлены на Рисунках 1-13.

Изображение выглядит как текст, снимок экрана, число, программное обеспечение

Автоматически созданное описание

**Рисунок 1 *—* Задача 1**

**Изображение выглядит как текст, снимок экрана, программное обеспечение

Автоматически созданное описание**

**Рисунок 2 *—* Задача 2**

**Изображение выглядит как текст, снимок экрана, программное обеспечение, Значок на компьютере

Автоматически созданное описание**

**Рисунок 3 *—* Задача 2.1**

**Изображение выглядит как текст, снимок экрана, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание**

**Рисунок 4 *—* Код задачи 2.2**

**Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание**

**Рисунок 5 *—* Задача 2.2**

**Изображение выглядит как текст, снимок экрана, программное обеспечение, Значок на компьютере

Автоматически созданное описание**

**Рисунок 6 *—* Задача 2.3**

**Изображение выглядит как текст, снимок экрана, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание**

**Рисунок 7 *—* Задача 3**

**Изображение выглядит как текст, снимок экрана, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание**

**Рисунок 8 *—* Задача 3.1**

**Изображение выглядит как текст, снимок экрана, программное обеспечение, Шрифт

Автоматически созданное описание**

**Рисунок 9 *—* Задача 3.2**

**Изображение выглядит как текст, снимок экрана

Автоматически созданное описание**

**Рисунок 10 *—* Задача 3.3**

**Изображение выглядит как текст, снимок экрана, число, программное обеспечение

Автоматически созданное описание**

**Рисунок 11 *—* Задача 3.4**

**Изображение выглядит как текст, снимок экрана, Шрифт, программное обеспечение

Автоматически созданное описание**

**Рисунок 12 *—* Задача 3.5**

**Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание**

**Рисунок 13 *—* Задача 3.6**

ЗАКЛЮЧЕНИЕ

В данной работе были проделаны вычисления корреляции по Пирсону, исследованы углы наклона и диаграмма рассеяния, методы линейной регрессии, однофакторный ANOVA-тест, определена поправка Бонферрони, выполнены пост-хок тесты Тьюки и построены соответствующие графики.