**Лабораторная работа № 4**

«Корреляционный анализ»

студента Баранова Александра группы Б22-534. Дата сдачи: 10.12.2024

Ведущий преподаватель: Новиков М.А. оценка: подпись:\_\_\_\_\_\_\_

Вариант №**2**

*Цель работы*: изучение функций Statistics and Machine Learning ToolboxÔ MATLAB / Python SciPy.stats для проведения корреляционного анализа данных.

1. Исходные данные

Характеристики наблюдаемых случайных величин:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| СВ | Распределение | Параметры | Математическое ожидание, *mi* | Дисперсия, image.pdf | Объем выборки, *ni* |
| *X* |  |  |  |  | 50 |
| *Y* |  |  |  |  |

*Примечание*: для генерации случайных чисел использовать функции **rand**, **randn, chi2rnd (****s****c****ipy.stats: uniform.rvs, norm.rvs, chi2.rvs)**

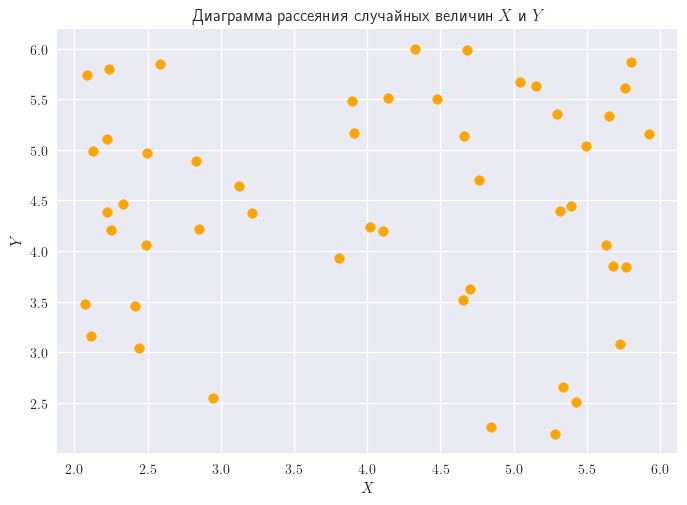
Выборочные характеристики:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| СВ | Среднее, | Оценка дисперсии, | КК по Пирсону, | КК по Спирмену, | КК по Кендаллу, |
| *X* | 4.03 | 1.83 | -0.01 | 0.02 | 0.01 |
| *Y* | 4.47 | 1.13 |

Проверка значимости коэффициентов корреляции:

|  |  |  |  |
| --- | --- | --- | --- |
| Статистическая гипотеза, *H*0 | *p-value* | Статистическое решение при | Ошибка стат. решения |
|  | 0.97 | принимается | Нет |
|  | 0.90 | принимается | Нет |
|  | 0.93 | принимается | Нет |

*Примечание*: для проверки гипотез использовать функцию **corr (scipy.stats.pearsonr)**

2. Визуальное представление двумерной выборки

*Примечание*: для построения диаграммы использовать функции **plot, scatter (matplotlib.pyplot.scatter)**

3. Проверка независимости методом таблиц сопряженности

Статистическая гипотеза:

Эмпирическая таблица сопряженности:

|  |  |  |  |
| --- | --- | --- | --- |
|  | [2.20; 3.46) | [3.46; 4.73) | [4.73; 6.00] |
| = [2.07; 3.36) | 4 | 8 | 7 |
| = [3.36; 4.64) | 0 | 3 | 5 |
| = [4.64; 5.92] | 5 | 8 | 10 |

*Примечание*: для группировки использовать функцию **hist3 (matplotlib.pyplot.hist2d)**

Теоретическая таблица сопряженности:

|  |  |  |  |
| --- | --- | --- | --- |
|  | [2.20; 3.46) | [3.46; 4.73) | [4.73; 6.00] |
| = [2.07; 3.36) | 3.42 | 7.22 | 8.36 |
| = [3.36; 4.64) | 1.44 | 3.04 | 3.52 |
| = [4.64; 5.92] | 4.14 | 8.74 | 10.12 |

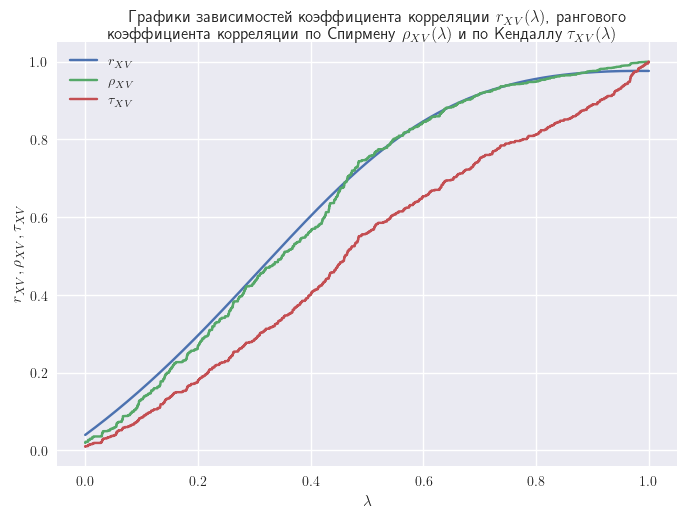
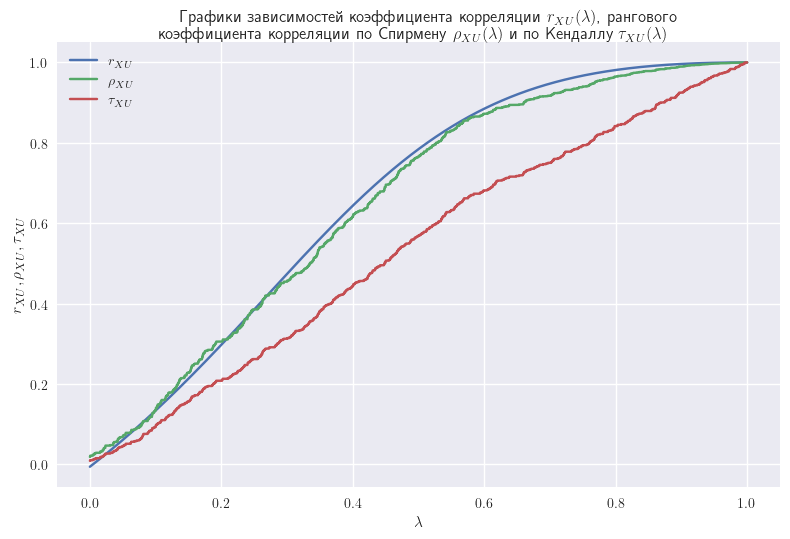
|  |  |  |  |
| --- | --- | --- | --- |
| Выборочное значение статистики критерия | *p-value* | Статистическое решение при | Ошибка стат. решения |
| 2.71 | 0.61 | принимается | Нет |

*Примечание*: для проверки гипотезы использовать функцию **crosstab (scipy.stats.chi2\_contingency)**

4. Исследование корреляционной связи

Случайная величина *U* = l*X* + (1–l)*Y*, lÎ[0; 1]

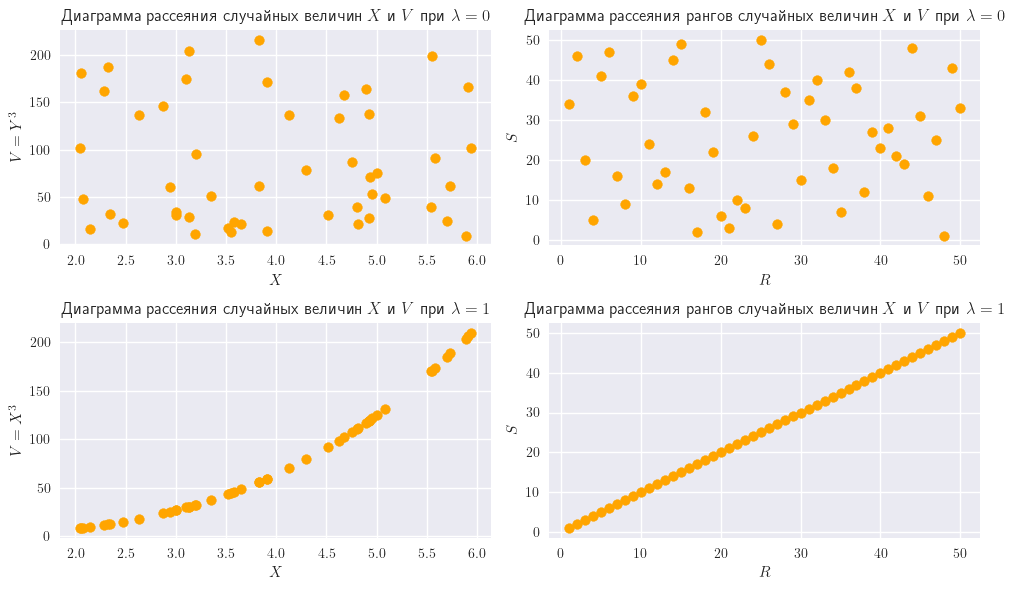
Случайная величина *V* = l*X*3 + (1–l)*Y*3 lÎ[0; 1]



*Выводы*:

По первому графику: при все коэффициенты корреляции стремятся к 0, что свидетельствует об отсутствии линейной корреляционной связи между случайными величинами и . При все коэффициенты корреляции стремятся к 1, что свидетельствует о наличии линейной функциональной зависимости между случайными величинами и .

По второму графику: никогда не принимает значений 1, что свидетельствует об отсутствии линейной функциональной зависимости между случайными величинами и . Однако при коэффициенты корреляции , что свидетельствует о наличии монотонной функциональной зависимости между случайными величинами и . При коэффициенты корреляции близки к 0, что свидетельствует об отсутствии линейной и даже монотонной корреляционной связи между случайными величинами и .



*Примечание*: для расчёта рангов использовать функцию **tiedrank (scipy.stats.rankdata)**

*Выводы*:

Если и независимы, то и их ранги и также будут независимыми.

Если , - монотонная функция, то переход к рангам "выпрямляет" монотонную зависимость исходных признаков.