**Лабораторная работа № 4**

«Корреляционный анализ»

студента Когановского Григория группы Б22-534. Дата сдачи: 10.12.2024

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

Вариант №**7**

*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* |
|  |  |  |  |  | 100 |
|  |  |  |  |  |

*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* | -1.15 | 3.54 | 0.03 | 0.00 | 0.01 |
| *Y* | -0.97 | 0.32 |

Проверка значимости коэффициентов корреляции:

|  |  |  |  |
| --- | --- | --- | --- |
| Статистическая гипотеза, *H*0 | *p-value* | Статистическое решение при | Ошибка стат. решения |
|  | 0.73 | принимается | Нет |
|  | 0.97 | принимается | Нет |
|  | 0.94 | 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:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | [-1.95; -1.56) | [-1.56; -1.17) | [-1.17; -0.78) | [-0.78; -0.39) | [-0.39; 0.00] |
| = [-5.37; -3.61) | 2 | 2 | 2 | 3 | 0 |
| = [-3.61; -1.84) | 4 | 6 | 2 | 2 | 10 |
| = [-1.84; -0.07] | 6 | 12 | 7 | 14 | 5 |
| = [-0.07; 1.69] | 5 | 2 | 0 | 5 | 1 |
| = [1.69; 3.46] | 2 | 2 | 1 | 2 | 3 |

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

Теоретическая таблица сопряженности:

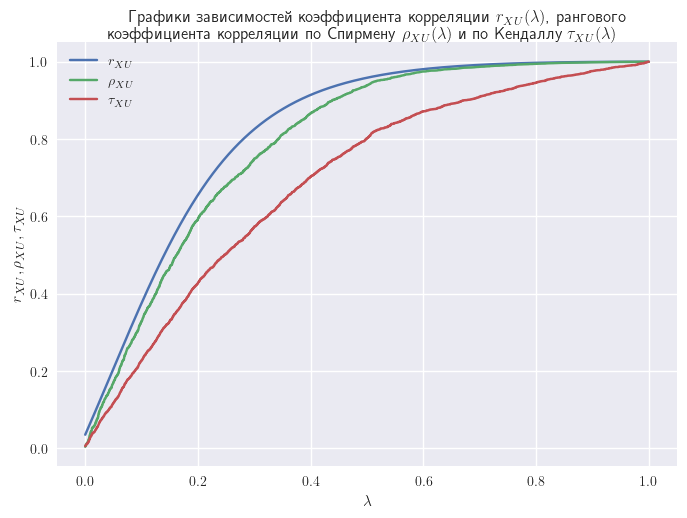
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | [-1.95; -1.56) | [-1.56; -1.17) | [-1.17; -0.78) | [-0.78; -0.39) | [-0.39; 0.00] |
| = [-5.37; -3.61) | 1.71 | 2.16 | 1.08 | 2.34 | 1.71 |
| = [-3.61; -1.84) | 4.56 | 5.76 | 2.88 | 6.24 | 4.56 |
| = [-1.84; -0.07] | 8.36 | 10.56 | 5.28 | 11.44 | 8.36 |
| = [-0.07; 1.69] | 2.47 | 3.12 | 1.56 | 3.38 | 2.47 |
| = [1.69; 3.46] | 1.90 | 2.40 | 1.20 | 2.60 | 1.90 |

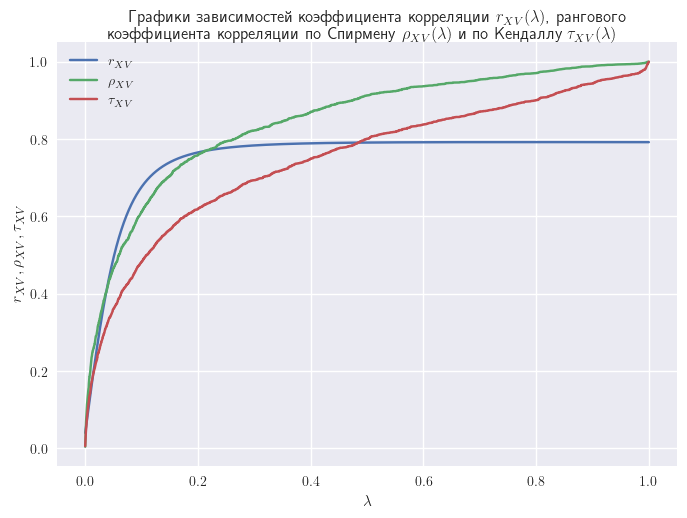
|  |  |  |  |
| --- | --- | --- | --- |
| Выборочное значение статистики критерия | *p-value* | Статистическое решение при | Ошибка стат. решения |
| 22.89 | 0.12 | принимается | Нет |

*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]

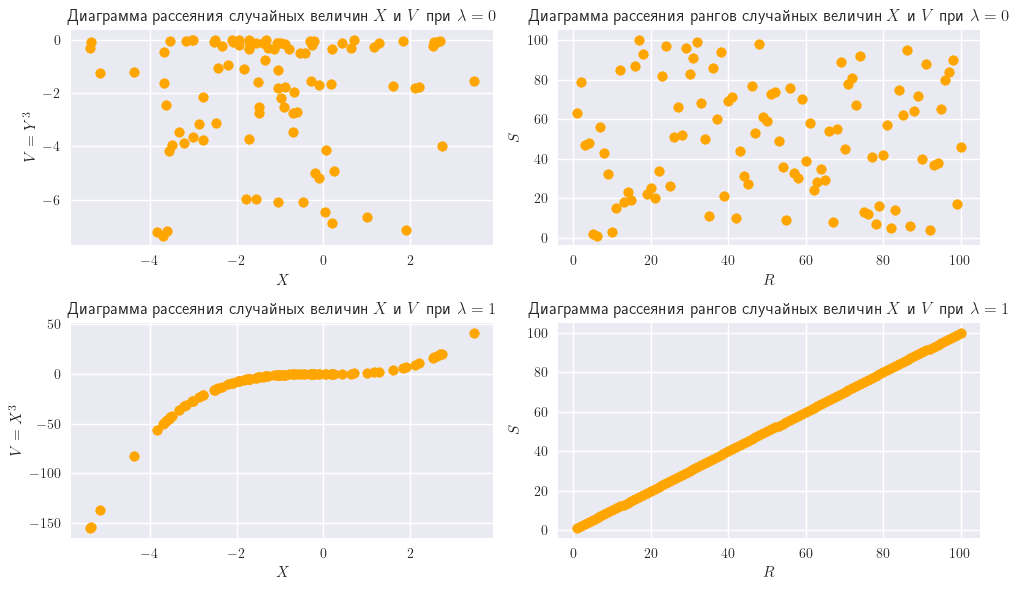
Случайная величина *V* = l*X*3 + (1–l)*Y*3 lÎ[0; 1]



*Выводы*:

По первому графику: при все коэффициенты корреляции стремятся к 0, что свидетельствует об отсутствии линейной и монотонной 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.