**Лабораторная работа № 4**

**«Корреляционный анализ»**

студента Кевролетина Алексея группы Б19-504 . Дата сдачи: 17.12.2021

Ведущий преподаватель: Трофимов А.Г. оценка: подпись:\_\_\_\_\_\_\_

Вариант № 10

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

1. Исходные данные

Характеристики наблюдаемых случайных величин:

| СВ | Распределение | Параметры | Математическое ожидание, *mi* | Дисперсия, | Объем выборки, *n* |
| --- | --- | --- | --- | --- | --- |
| *X* |  | k | 5 | 10 | 300 |
| *Y* |  | k | 5 | 10 |

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

Выборочные характеристики:

| СВ | Среднее, | Оценка дисперсии, | КК по Пирсону, | КК по Спирмену, | КК по Кендаллу, |
| --- | --- | --- | --- | --- | --- |
| *X* | 4.964 | 9.732 | -0.033 | -0.009 | -0.003 |
| *Y* | 4.785 | 8.995 |

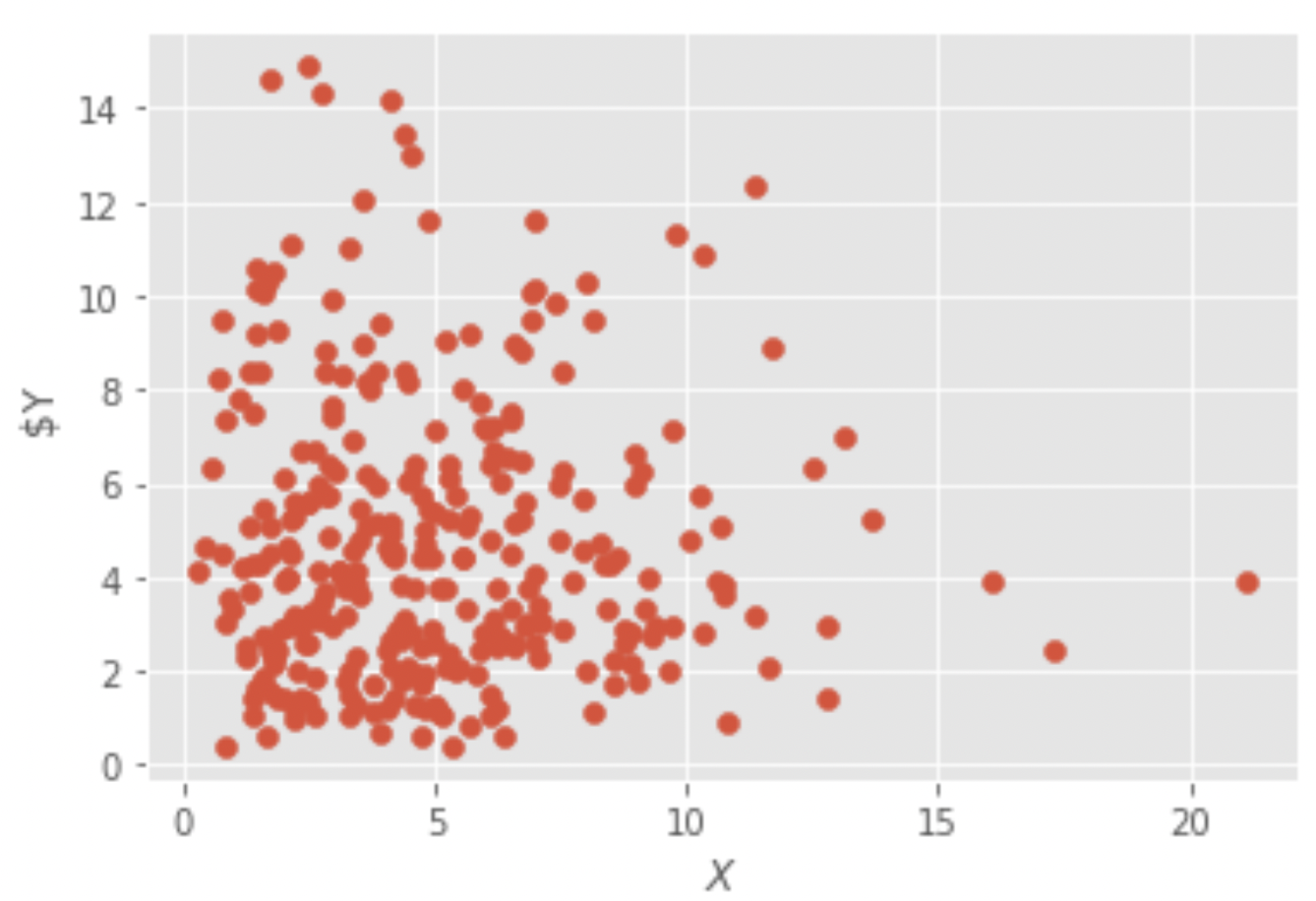
Проверка значимости коэффициентов корреляции:

| Статистическая гипотеза, *H*0 | *p-value* | Статистическое решение при  α = \_0.05\_ | Ошибка стат. решения |
| --- | --- | --- | --- |
| *H*0: *rXY* = 0 | 0.569 |  | рода |
| *H*0: ρ*XY* = 0 | 0.878 |  |  |
| *H*0: τ*XY* = 0 | 0.941 |  | 2 рода |

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

2. Визуальное представление двумерной выборки

Диаграмма рассеяния случайных величин *X* и *Y*:



3. Проверка независимости методом таблиц сопряженности

Статистическая гипотеза:



Эмпирическая таблица сопряженности:

| X \ Y | [0.397; 3.296) | [3.296; 6.196) | [6.196; 9.096) | [9.096; 11.995) | [11.995; 14.895) |
| --- | --- | --- | --- | --- | --- |
| = [0.296; 4.455) | 57 | 51 | 24 | 12 | 6 |
| = [4.455; 8.613) | 43 | 40 | 19 | 9 | 1 |
| = [8.613; 12.771) | 13 | 10 | 5 | 2 | 1 |
| = [12.771; 16.929) | 2 | 2 | 1 | 0 | 0 |
| =[16.929; 21.086] | 1 | 1 | 0 | 0 | 0 |

Теоретическая таблицы сопряженности:

| X \ Y | [0.397; 3.296) | [3.296; 6.196) | [6.196; 9.096) | [9.096; 11.995) | [11.995; 14.895) |
| --- | --- | --- | --- | --- | --- |
| = [0.296; 4.455) | 58 | 52 | 24.5 | 11.5 | 4 |
| = [4.455; 8.613) | 43.3 | 38.8 | 18.3 | 8.59 | 2.99 |
| = [8.613; 12.771) | 11.99 | 10.75 | 5.06 | 2.38 | 0.83 |
| = [12.771; 16.929) | 1.93 | 1.73 | 0.82 | 0.38 | 0.13 |
| =[16.929; 21.086] | 0.77 | 0.69 | 0.33 | 0.15 | 0.05 |

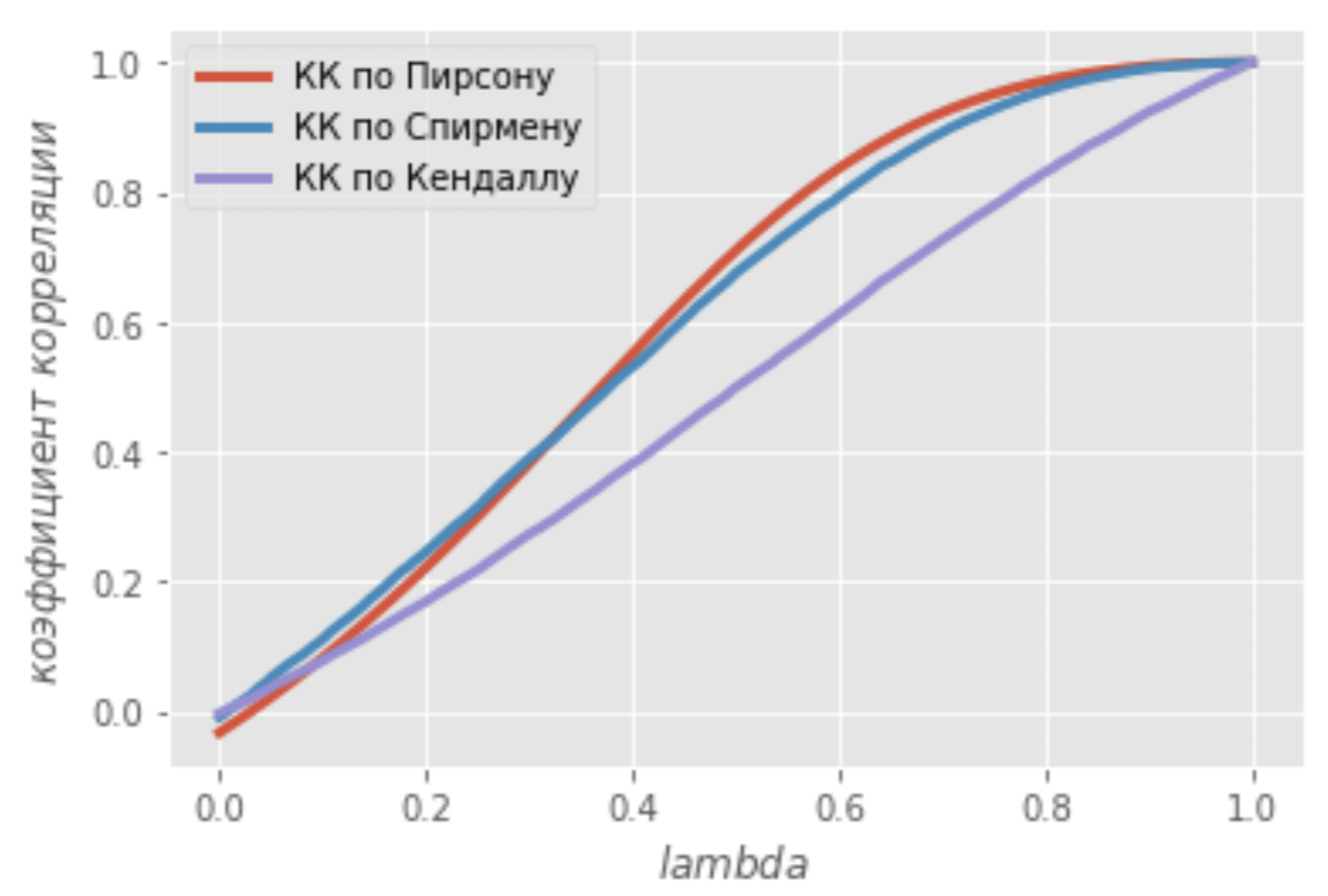
| Выборочное значение статистики критерия | *p-value* | Статистическое решение при  α = 0.05 | Ошибка стат. решения |
| --- | --- | --- | --- |
| 4.0457 | 0.9988 |  | нет |

4. Исследование корреляционной связи

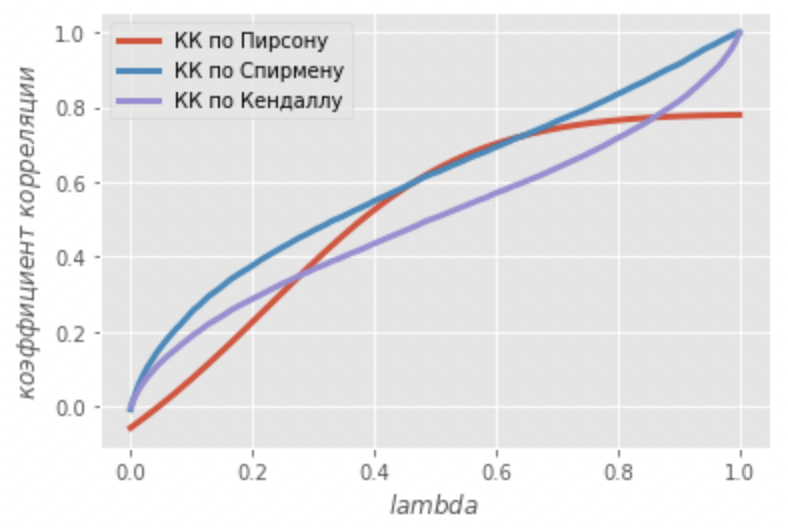
Случайная величина *U* = λ*X* + (1–λ)*Y*, λ∈[0; 1]

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

Графики зависимостей коэффициента корреляции , рангового коэффициента корреляции по Спирмену , по Кендаллу



Графики зависимостей , ,



*Выводы*: С увеличением значения коэффициент корреляции , ранговый коэффициент корреляции по Спирмену и по Кендаллу стремятся к единице. При коэффициенты корреляции равны нулю и статистическая связь между случайными величинами отсутствует, а при увеличении значения теснота статистической свзи между случайными величинами увеличивается, и при между случайными величинами имеется линейная функциональная связь.

| Диаграмма рассеяния случайных величин *X* и *V* при λ = 0: | Диаграмма рассеяния **рангов** случайных величин *X* и *V* при λ = 0: |
| --- | --- |

| Диаграмма рассеяния случайных величин *X* и *V* при λ = 1: | Диаграмма рассеяния **рангов** случайных величин *X* и *V* при λ = 1: |
| --- | --- |

*Выводы*: Из диаграммы рассеяния случайных величин X и V при видно, что статистическая связь между данными случайными величинами отсутствует, при этом ранги случайных величин X и V при рассеяны практически равномерно внутри квадрата. На диаграмме рассеяния случайных величин X и V при прослеживается монотонная зависимость между случайными величинами, при этом переход к рангам выпрямляет данную зависимость.