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 Toolbox™ MATLAB / Python SciPy.stats для проведения корреляционного анализа данных.

1. Исходные данные

Характеристики наблюдаемых случайных величин:

СВ	Распределение	Параметры	Математическое ожидание, m_i	Дисперсия, σ_i^2	Объем выборки, n_i
X	$R(2,6)$	$a_1 = 2, b_1 = 6$	$m_1 = \frac{a_1 + b_1}{2} = 4$	$\sigma_1^2 = \frac{(b_1 - a_1)^2}{12} = \frac{4}{3}$	50
Y	$R(2,6)$	$a_2 = 2, b_2 = 6$	$m_2 = \frac{a_2 + b_2}{2} = 4$	$\sigma_2^2 = \frac{(b_2 - a_2)^2}{12} = \frac{4}{3}$	

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

Выборочные характеристики:

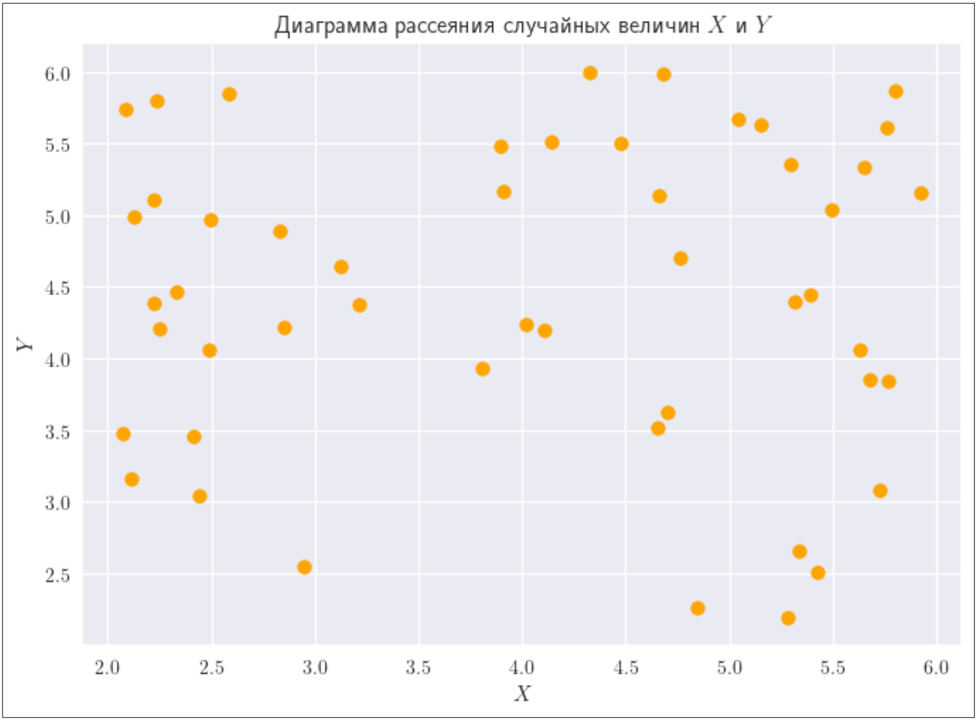
СВ	Среднее, $\overline{x_i}$	Оценка дисперсии, s_i^2	КК по Пирсону, r_{XY}	КК по Спирмену, ρ_{XY}	КК по Кендаллу, τ_{XY}
X	4.03	1.83	-0.01	0.02	0.01
Y	4.47	1.13			

Проверка значимости коэффициентов корреляции:

Статистическая гипотеза, H_0	*p-value*	Статистическое решение при $\alpha = 0.05$	Ошибка стат. решения
$H_0: \ r_{XY} = 0$ $H_1: \ r_{XY} \neq 0$	0.97	H_0 принимается	Нет
$H_0: \ \rho_{XY} = 0$ $H_1: \ \rho_{XY} \neq 0$	0.90	H_0 принимается	Нет
$H_0: \ \tau_{XY} = 0$ $H_1: \ \tau_{XY} \neq 0$	0.93	H_0 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:

$$H_0: \ F_y(y \mid X \in \Delta_1) = F_y(y \mid X \in \Delta_2) = \ldots = F_y(y \mid X \in \Delta_k) = F_y(y)$$
$$H': \ \exists i, j : F_Y(y \mid X \in \Delta_i) \neq F_Y(y \mid X \in \Delta_j)$$

Эмпирическая таблица сопряженности:

X \ Y	[2.20; 3.46)	[3.46; 4.73)	[4.73; 6.00]
Δ_1 = [2.07; 3.36)	4	8	7
Δ_2 = [3.36; 4.64)	0	3	5
Δ_3 = [4.64; 5.92]	5	8	10

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

Теоретическая таблица сопряженности:

X \ Y	[2.20; 3.46)	[3.46; 4.73)	[4.73; 6.00]
Δ_1 = [2.07; 3.36)	3.42	7.22	8.36
Δ_2 = [3.36; 4.64)	1.44	3.04	3.52
Δ_3 = [4.64; 5.92]	4.14	8.74	10.12

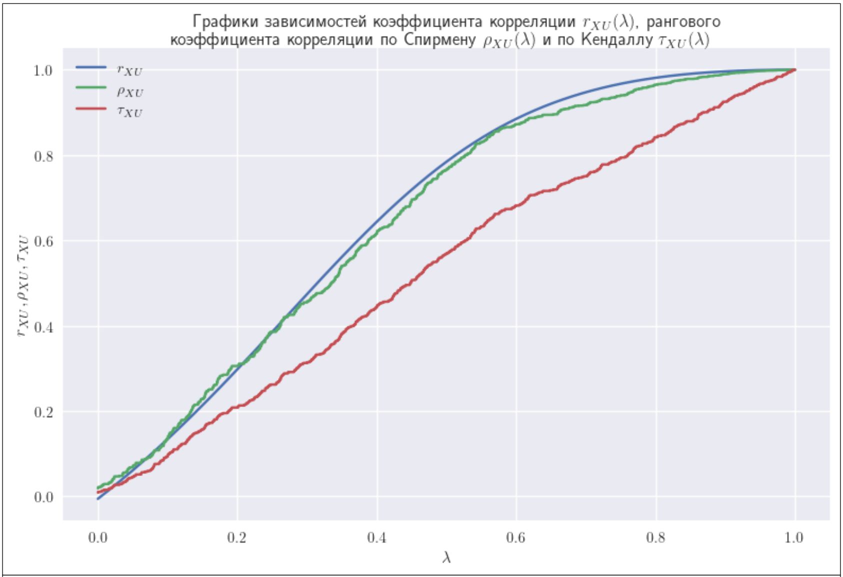
Выборочное значение статистики критерия	*p-value*	Статистическое решение при $\alpha = 0.05$	Ошибка стат. решения
2.71	0.61	H_0 принимается	Нет

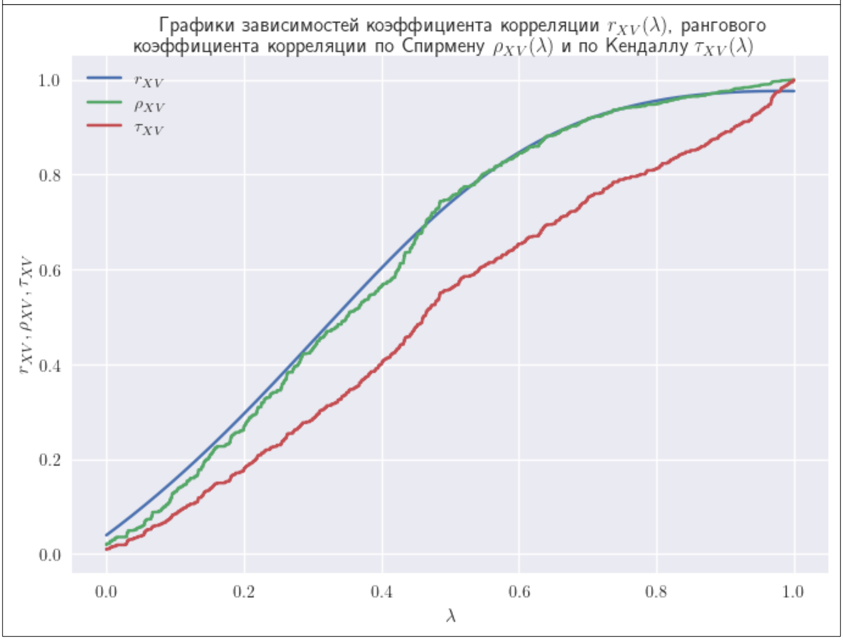
Примечание: для проверки гипотезы использовать функцию **crosstab (scipy.stats.chi2_contingency)**

4. Исследование корреляционной связи

Случайная величина $U = \lambda X + (1–\lambda)Y$, $\lambda \in [0; 1]$

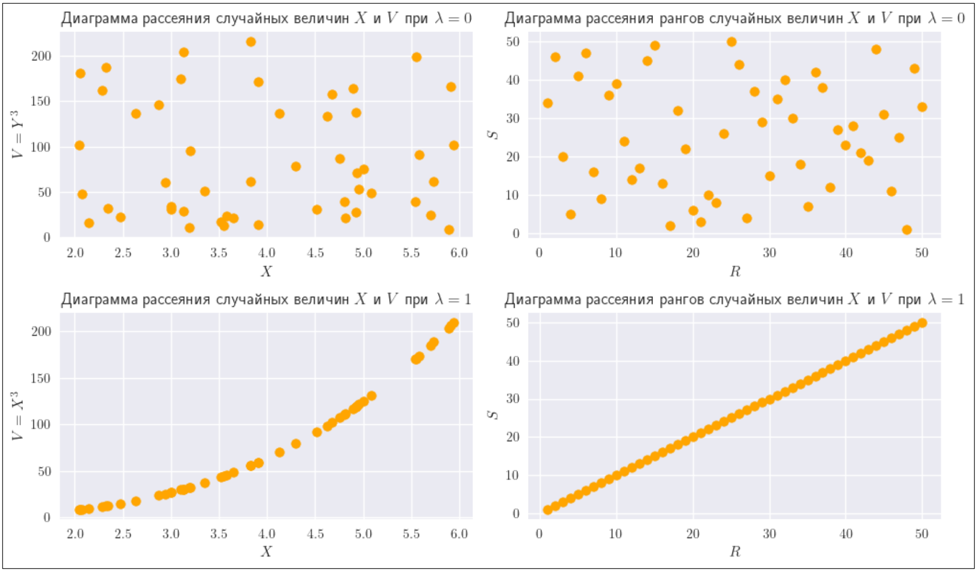
Случайная величина $V = \lambda X^3 + (1–\lambda)Y^3$ $\lambda \in [0; 1]$





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

По второму графику: r_{XV} никогда не принимает значений 1, что свидетельствует об отсутствии линейной функциональной зависимости между случайными величинами X и Y. Однако при $\lambda \to 1$ коэффициенты корреляции $\rho_{XV}, \tau_{XV} \to 1$, что свидетельствует о наличии монотонной функциональной зависимости между случайными величинами X и V. При $\lambda \to 0$ коэффициенты корреляции $r_{XV}, \rho_{XV}, \tau_{XV}$ близки к 0, что свидетельствует об отсутствии линейной и даже монотонной корреляционной связи между случайными величинами X и V.



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

Выводы:
Если X и V независимы, то и их ранги R и S также будут независимыми.

Если $V = \varphi(X)$, φ - монотонная функция, то переход к рангам "выпрямляет" монотонную зависимость исходных признаков.