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 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	$N(-1,2)$	$m_1=-1, \sigma_1=2$	$m_1=-1$	$\sigma_1^2=4$	100
Y	$R(-2,0)$	$a_2=-2, b_2=0$	$m_2=\frac{a_2+b_2}{2}=-1$	$\sigma_2^2=\frac{(b_2-a_2)^2}{12}=\frac{1}{3}$	

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	-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 $\alpha = 0.05$	Ошибка стат. решения
$H_0: \ r_{XY} = 0$ $H_1: \ r_{XY} \neq 0$	0.73	H_0 принимается	Нет
$H_0: \ \rho_{XY} = 0$ $H_1: \ \rho_{XY} \neq 0$	0.97	H_0 принимается	Нет
$H_0: \ \tau_{XY} = 0$ $H_1: \ \tau_{XY} \neq 0$	0.94	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	[-1.95; -1.56)	[-1.56; -1.17)	[-1.17; -0.78)	[-0.78; -0.39)	[-0.39; 0.00]
Δ_1 = [-5.37; -3.61)	2	2	2	3	0
Δ_2 = [-3.61; -1.84)	4	6	2	2	10
Δ_3 = [-1.84; -0.07]	6	12	7	14	5
Δ_4 = [-0.07; 1.69]	5	2	0	5	1
Δ_5 = [1.69; 3.46]	2	2	1	2	3

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

Теоретическая таблица сопряженности:

X \ Y	[-1.95; -1.56)	[-1.56; -1.17)	[-1.17; -0.78)	[-0.78; -0.39)	[-0.39; 0.00]
Δ_1 = [-5.37; -3.61)	1.71	2.16	1.08	2.34	1.71
Δ_2 = [-3.61; -1.84)	4.56	5.76	2.88	6.24	4.56
Δ_3 = [-1.84; -0.07]	8.36	10.56	5.28	11.44	8.36
Δ_4 = [-0.07; 1.69]	2.47	3.12	1.56	3.38	2.47
Δ_5 = [1.69; 3.46]	1.90	2.40	1.20	2.60	1.90

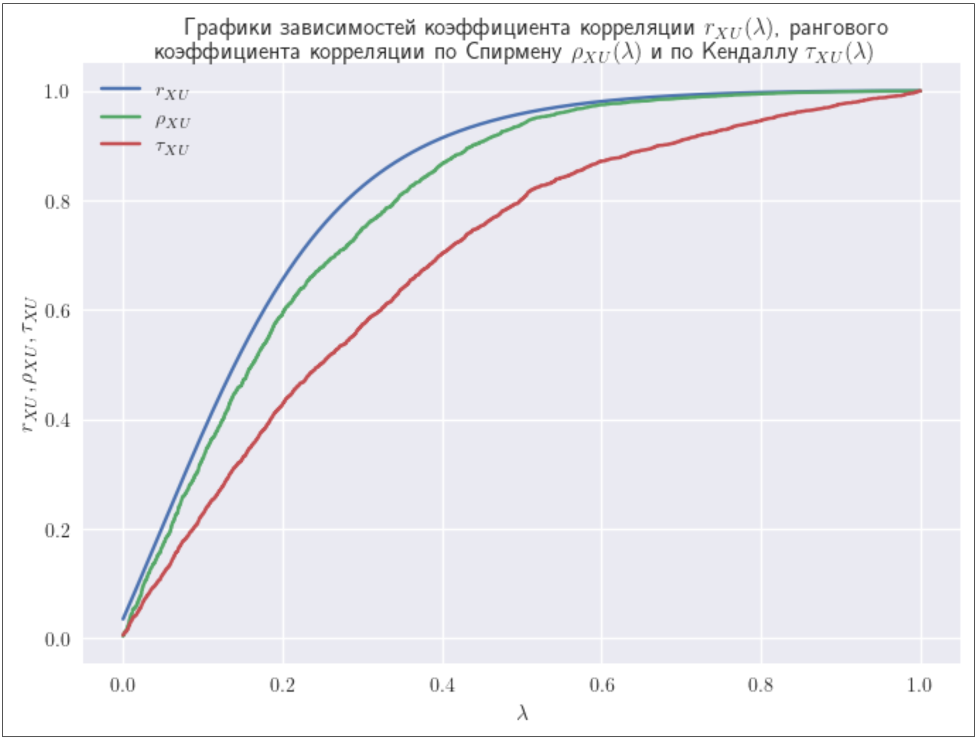
Выборочное значение статистики критерия	*p-value*	Статистическое решение при $\alpha = 0.05$	Ошибка стат. решения
22.89	0.12	H_0 принимается	Нет

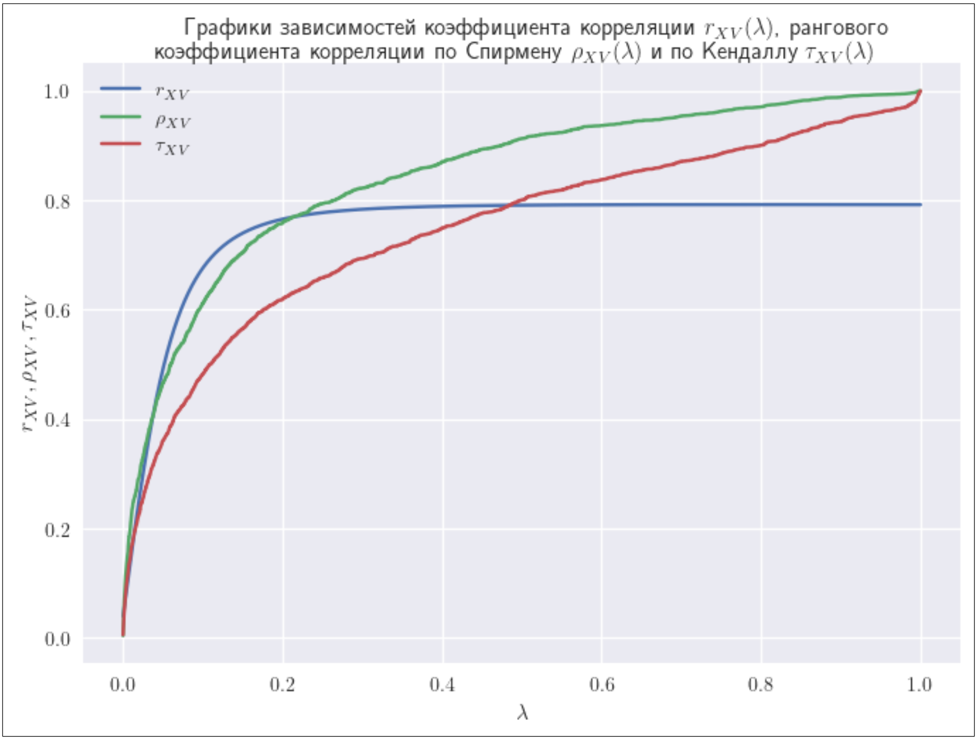
Примечание: для проверки гипотезы использовать функцию **crosstab (scipy.stats.chi2_contingency)**

<u>4. Исследование корреляционной связи</u>

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

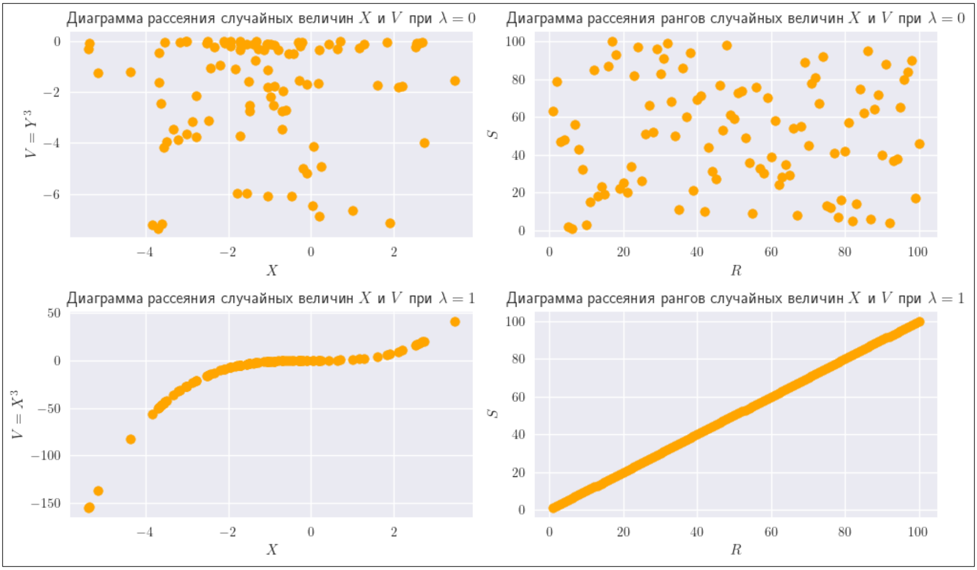
Случайная величина $V = \lambda X^3 + (1–\lambda)Y^3$ $\lambda \in [0; 1]$





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