

Лабораторная работа № 3 «Однофакторный дисперсионный анализ»

студента Когановского Григория группы Б22-534. Дата сдачи: 29.11.2024
Ведущий преподаватель: Новиков М.А. оценка: _____ подпись: _____

Вариант №7

Цель работы: изучение функций Statistics and Machine Learning Toolbox™ MATLAB / Python SciPy.stats для проведения однофакторного дисперсионного анализа (*One-Way ANOVA*).

1. Исходные данные

Характеристики наблюдаемых случайных величин:

СВ	Распределение	Параметры	Математическое ожидание, m_i	Дисперсия, σ_i^2	Объем выборки, n_i
X_1	$N(-1, 2)$	$m_1 = -1, \sigma_1 = 2$	$m_1 = -1$	$\sigma_1^2 = 4$	100
X_2	$R(-2, 0)$	$a_2 = -2, b_2 = 0$	$m_2 = \frac{a_2 + b_2}{2} = -1$	$\sigma_2^2 = \frac{(b_2 - a_2)^2}{12} = \frac{1}{3}$	100
X_3	$N(2, 1)$	$m_3 = 2, \sigma_3 = 1$	$m_3 = 2$	$\sigma_3^2 = 1$	200

Количество случайных величин $k = 3$

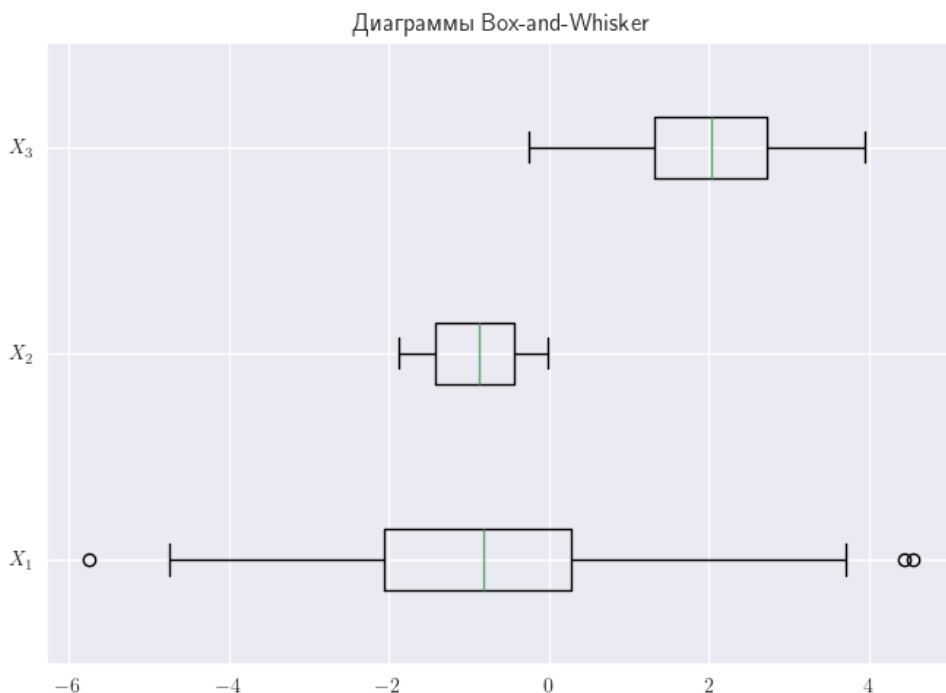
Примечание: для генерации случайных чисел использовать функции **rand**, **randn**, **chi2rnd** (**scipy.stats: uniform.rvs, norm.rvs, chi2.rvs**)

Выборочные характеристики:

СВ	Среднее, \bar{x}_i	Оценка дисперсии, s_i^2	Оценка с.к.о., s_i
X_1	-0.87	3.88	1.97
X_2	-0.92	0.33	0.57
X_3	2.01	0.86	0.93
<i>Pooled</i>	0.07	3.56	1.89

2. Визуальное представление выборок

Диаграммы *Box-and-Whisker*:



Примечание: для построения диаграмм использовать функции **boxplot**, **vartestn** (**matplotlib.pyplot.boxplot**)

3. Проверка условия применимости дисперсионного анализа

Статистическая гипотеза: $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$

Критерий Бартлетта:

Выборочное значение статистики критерия	p -value	Статистическое решение при $\alpha = 0.05$	Ошибка стат. решения
145.96	0.00	H_0 отклоняется	Нет

Примечание: для проверки гипотезы использовать функцию **vartestn** (**scipy.stats.bartlett**)

4. Однофакторный дисперсионный анализ

Таблица дисперсионного анализа:

Источник вариации	Показатель вариации	Число степеней свободы	Несмещённая оценка
Группировочный признак	$\tilde{D}_{\text{межгр}} = 1.88$	$k - 1 = 2$	$\frac{n}{k - 1} \tilde{D}_{\text{межгр}} = 281.77$
Остаточные признаки	$\tilde{D}_{\text{внутр}} = 1.67$	$n - k = 297$	$\frac{n}{n - k} \tilde{D}_{\text{внутр}} = 1.69$
Все признаки	$\tilde{D}_{\text{общ}} = 3.55$	$n - 1 = 299$	$\frac{n}{n - 1} \tilde{D}_{\text{общ}} = 3.56$

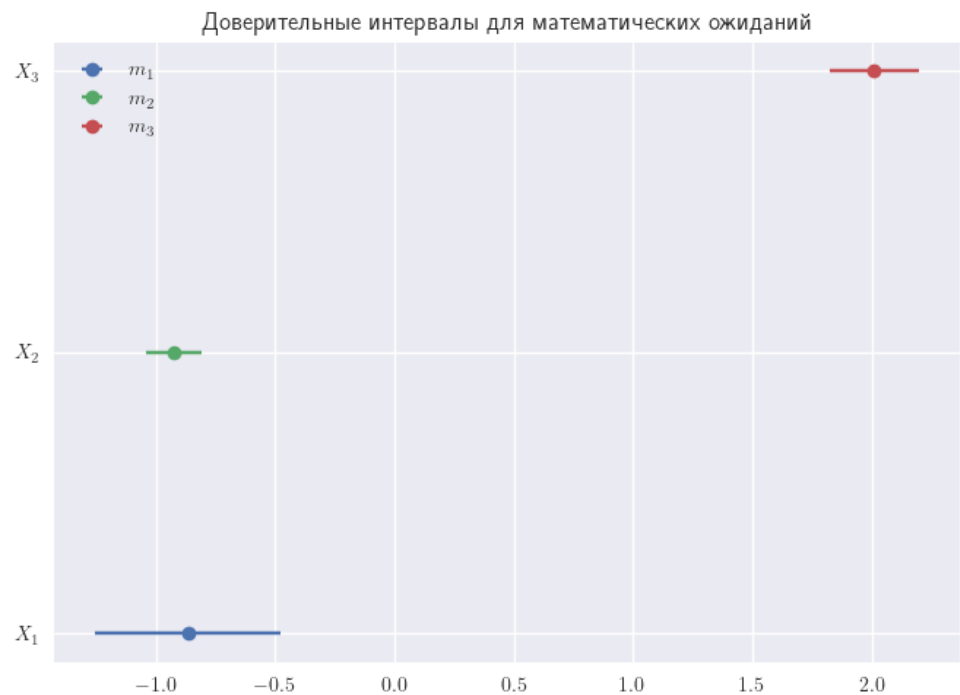
Эмпирический коэффициент детерминации $\eta^2 = 0.53$ Эмпирическое корреляционное отношение $\eta = 0.73$ Статистическая гипотеза: $H_0 : m_1 = m_2 = m_3$

Выборочное значение статистики критерия	<i>p-value</i>	Статистическое решение при $\alpha = 0.05$	Ошибка стат. решения
166.84	0.00	H_0 отклоняется	Нет

Примечание: при расчетах использовать функцию **anova1** (**scipy.stats.f_oneway**)

5. Метод линейных контрастов

Доверительные интервалы для m_1, \dots, m_k :



Попарные сравнения m_i и m_j :

Гипотеза	Выборочное значение статистики критерия	<i>p-value</i>	Статистическое решение при $\alpha = 0.05$	Ошибка стат. решения
$H_0 : m_1 = m_2$	-0.06	0.94	H_0 принимается	Нет
$H_0 : m_1 = m_3$	2.88	0.00	H_0 отклоняется	Нет
$H_0 : m_2 = m_3$	2.94	0.00	H_0 отклоняется	Нет

Примечание: при расчетах использовать функцию **multcompare** (statsmodels.stats.multicomp.pairwise_tukeyhsd)