**Лабораторная работа № 2**

«Критерии согласия и однородности выборок»

студента Баранова Александра группы Б22-534. Дата сдачи: 17.11.2024

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

Вариант №**2**

*Цель работы*: изучение функций Statistics and Machine Learning ToolboxÔ MATLAB / Python SciPy.stats для проверки критериев согласия (*goodness-of-fit tests*) и однородности выборок.

1. Исходные данные

Характеристики наблюдаемой случайной величины *X*:

|  |  |  |  |
| --- | --- | --- | --- |
| Распределение | Параметры | Математическое ожидание, *m* | Дисперсия, s2 |
|  |  |  |  |

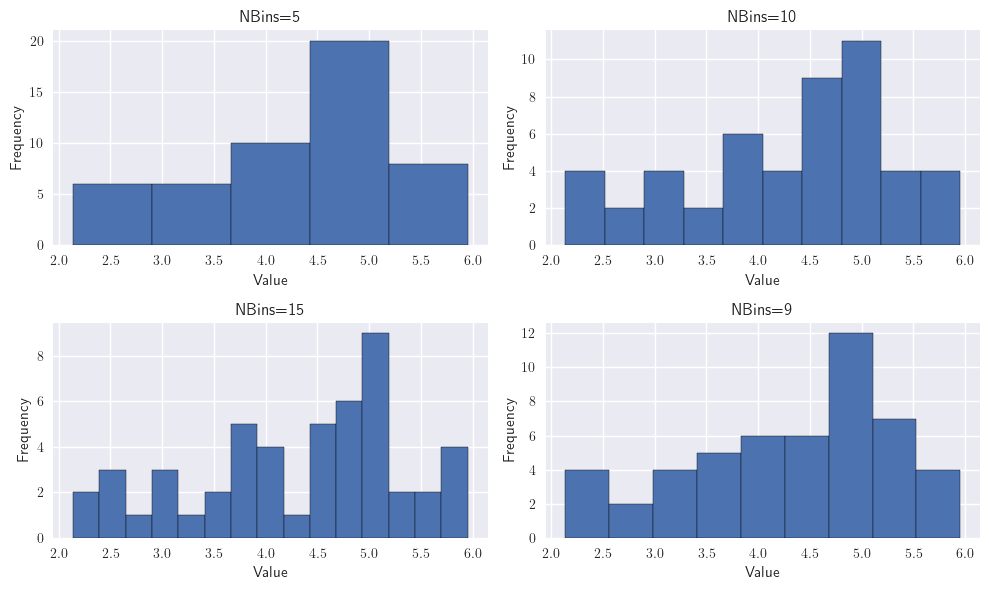
Объём выборки *n*1 =**50**

*Примечание*: для генерации случайных чисел использовать функции **r****a****nd**, **randn, chi2rnd (****s****c****ipy.stats: uniform.rvs, norm.rvs, chi2.rvs)**

Выборочные характеристики:

|  |  |  |
| --- | --- | --- |
| Среднее, image.pdf | Оценка дисперсии, s2 | Оценка с.к.о., s |
| 4.31 | 1.04 | 1.02 |

2. Визуальное представление выборки

Гистограммы частот:

*Примечание*: для построения гистограмм использовать функцию **hist (scipy.stats: histogram; matplotlib.pyplot: hist)**

3. Критерий хи-квадрат

*а*) Статистическая гипотеза: *H*0: *X* ~ *N(m,* s)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Число интервалов группировки | Выборочное значение статистики критерия | *p-value* | Статистическое решение при  a = 0.05 | Ошибка стат. решения |
| 5 | 8.14 | 0.02 | отклоняется | Нет |
| 10 | 16.29 | 0.02 | отклоняется | Нет |
| 15 | 23.50 | 0.02 | отклоняется | Нет |
| 9 | 11.48 | 0.07 | принимается | Да |

*б*) Статистическая гипотеза: *H*0: *X* ~ *R*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Число интервалов группировки | Выборочное значение статистики критерия | *p-value* | Статистическое решение при  a = 0.05 | Ошибка стат. решения |
| 5 | 13.60 | 0.00 | отклоняется | Да |
| 10 | 15.20 | 0.03 | отклоняется | Да |
| 15 | 20.80 | 0.05 | принимается | Нет |
| 9 | 11.56 | 0.07 | принимается | Нет |

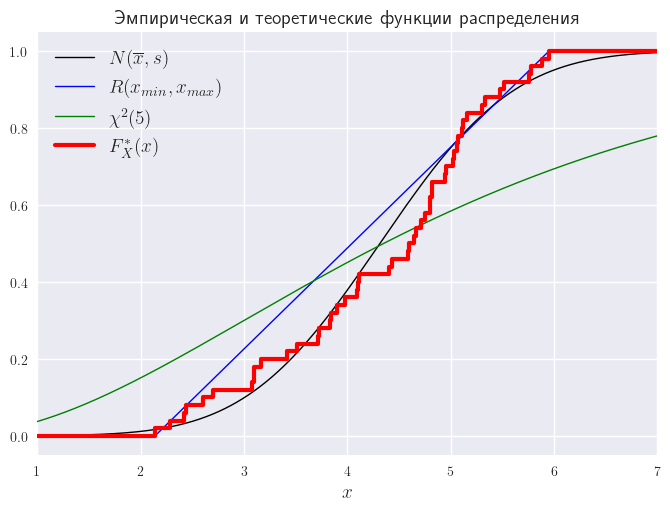
*в*) Статистическая гипотеза: *H*0: *X* ~ c2(5)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Число интервалов группировки | Выборочное значение статистики критерия | *p-value* | Статистическое решение при  a = 0.05 | Ошибка стат. решения |
| 5 | 55.16 | 0.00 | отклоняется | Нет |
| 10 | 58.71 | 0.00 | отклоняется | Нет |
| 15 | 71.30 | 0.00 | отклоняется | Нет |
| 9 | 51.65 | 0.00 | отклоняется | Нет |

*Примечание*: при расчетах использовать функции **chi2gof**, **fitdist (scipy.stats: histogram, chisquare)**

4. Критерий Колмогорова

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Статистическая гипотеза, *H*0 | Выборочное значение статистики критерия | *p-value* | Статистическое решение при  a = 0.05 | Ошибка стат. решения |
| *X* ~ *N*(*m*, s) | 0.15 | 0.21 | принимается | Да |
| *X* ~ *R* | 0.18 | 0.06 | принимается | Нет |
| *X* ~ c2(5) | 0.31 | 0.00 | отклоняется | Нет |

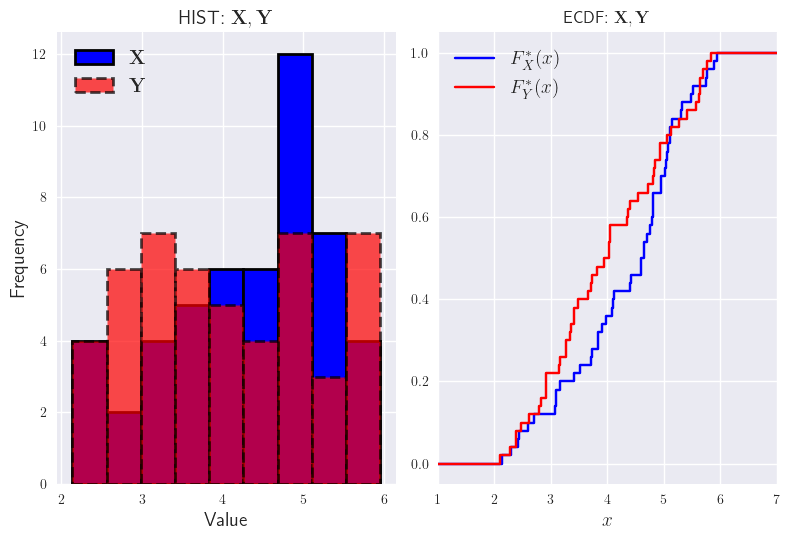
*Примечание*: при расчетах использовать функции **kstest, lillietest, fitdist (scipy.stats: kstest)**

*Примечание*: для построения графиков использовать функции **ecdf, cdf (scipy.stats: uniform.cdf, norm.cdf, chi2.cdf; statsmodels.distributions. empirical\_distribution: ECDF)**

5. Двухвыборочные критерии

Характеристики наблюдаемой случайной величины *Y*:

|  |  |  |  |
| --- | --- | --- | --- |
| Распределение | Параметры | Математическое ожидание | Дисперсия |
|  |  |  |  |

Объём выборки *n*2 = **50**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Критерий | Статистическая гипотеза, *H*0 | Выборочное значение статистики критерия | *p-value* | Статистическое решение при  a = 0.05 | Ошибка стат. решения |
| Chi-squared |  | 7.12 | 0.52 | принимается | Нет |
| KS-test |  | 0.22 | 0.18 | принимается | Нет |
| Sign test |  | 5 | 0.20 | принимается | Нет |
| U-test |  | 1459 | 0.15 | принимается | Нет |

*Примечание*: при расчетах использовать функции **chi2gof, kstest2, signtest, ranksum (scipy.stats: chisquare, ks\_2samp; statsmodels.stats.descriptivestats. sign\_test, ranksums)**