**Лабораторная работа № 2**

«Критерии согласия и однородности выборок»

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 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 |
|  |  |  |  |

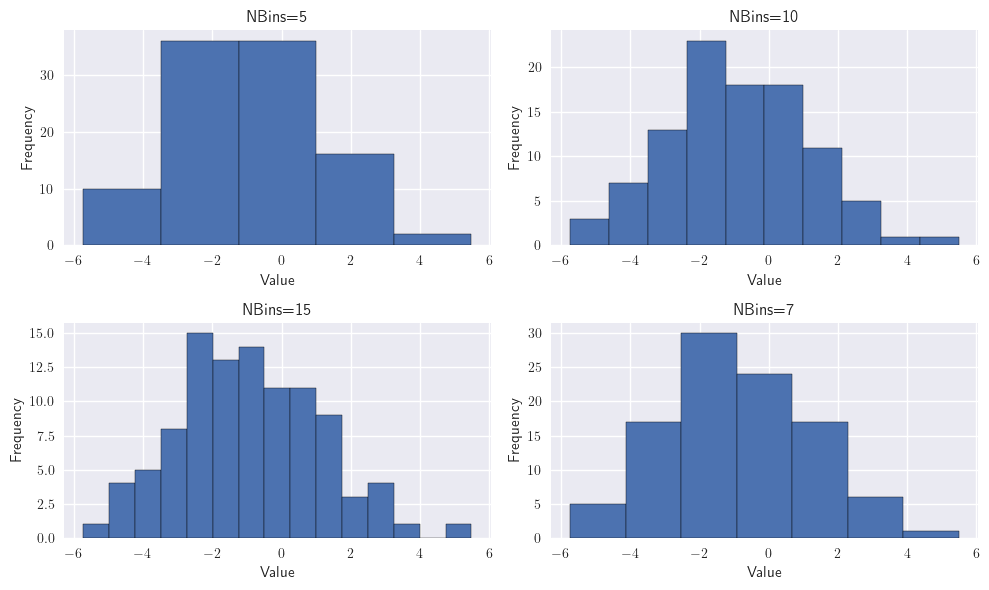
Объём выборки *n*1 = **100**

*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 |
| -0.91 | 4.37 | 2.09 |

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 | 0.48 | 0.79 | принимается | Нет |
| 10 | 2.13 | 0.95 | принимается | Нет |
| 15 | 6.94 | 0.86 | принимается | Нет |
| 7 | 0.99 | 0.91 | принимается | Нет |

*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 | 47.60 | 0.00 | отклоняется | Нет |
| 10 | 55.20 | 0.00 | отклоняется | Нет |
| 15 | 56.90 | 0.00 | отклоняется | Нет |
| 7 | 48.12 | 0.00 | отклоняется | Нет |

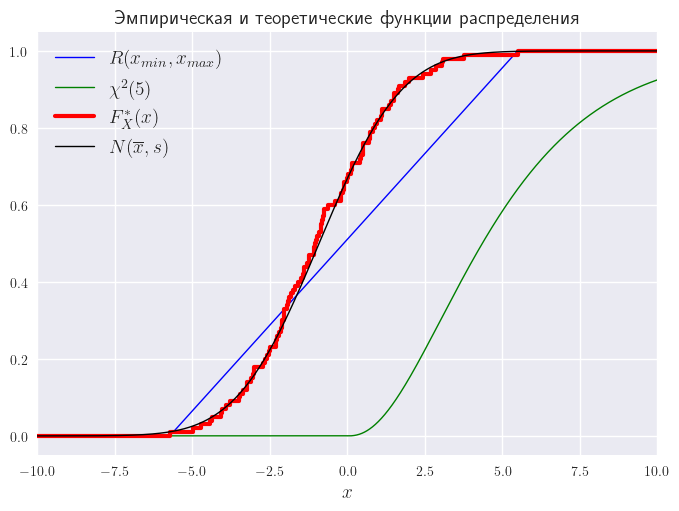
*в*) Статистическая гипотеза: *H*0: *X* ~ c2(5)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Число интервалов группировки | Выборочное значение статистики критерия | *p-value* | Статистическое решение при  a = 0.05 | Ошибка стат. решения |
| 5 |  | 0.00 | отклоняется | Нет |
| 10 |  | 0.00 | отклоняется | Нет |
| 15 |  | 0.00 | отклоняется | Нет |
| 7 |  | 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) | 0.06 | 0.86 | принимается | Нет |
| *X* ~ *R* | 0.25 | 0.00 | отклоняется | Нет |
| *X* ~ c2(5) | 0.80 | 0.00 | отклоняется | Нет |

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

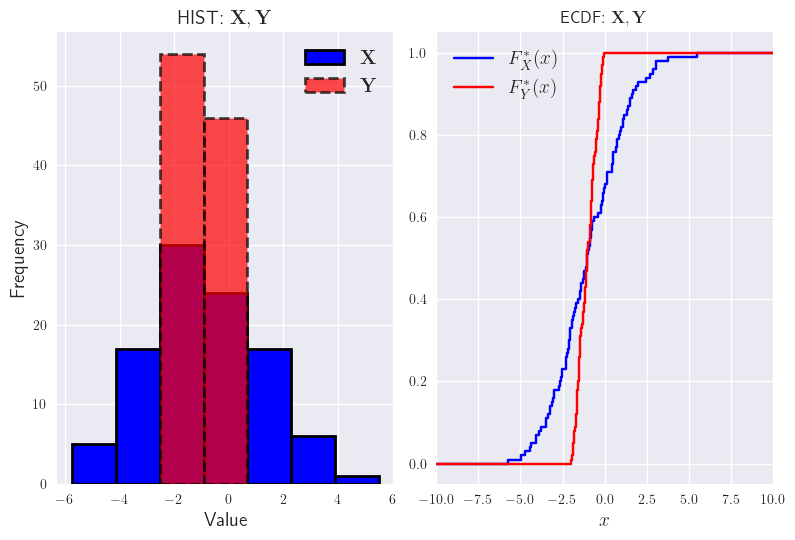
*Примечание*: для построения графиков использовать функции **ecdf, cdf (scipy.stats: uniform.cdf, norm.cdf, chi2.cdf; statsmodels.distributions. empirical\_distribution: ECDF)**

5. Двухвыборочные критерии

Характеристики наблюдаемой случайной величины *Y*:

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

Объём выборки *n*2 = **100**



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Критерий | Статистическая гипотеза, *H*0 | Выборочное значение статистики критерия | *p-value* | Статистическое решение при  a = 0.05 | Ошибка стат. решения |
| Chi-squared |  | 59.77 | 0.00 | отклоняется | Нет |
| KS-test |  | 0.34 | 0.00 | отклоняется | Нет |
| Sign test |  | 1 | 0.92 | принимается | Да |
| U-test |  | 4989 | 0.98 | принимается | Да |

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