**Лабораторная работа № 2**

«Критерии согласия и однородности выборок»

студента группы . Дата сдачи:\_\_\_\_\_\_\_\_\_\_

Ведущий преподаватель: оценка: подпись:\_\_\_\_\_\_\_

Вариант №\_\_\_\_\_\_\_

*Цель работы*: изучение функций Statistics and Machine Learning Toolbox™ 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, σ2 |
|  |  |  |  |

Объём выборки *n*1 = \_\_\_\_\_\_\_\_\_

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

Выборочные характеристики:

|  |  |  |
| --- | --- | --- |
| Среднее, | Оценка дисперсии, s2 | Оценка с.к.о., s |
|  |  |  |

2. Визуальное представление выборки

Гистограммы частот:



*NBins* = 5 *NBins* = 10



*NBins* = 15 *NBins* = \_\_\_\_

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

3. Критерий хи-квадрат

*а*) Статистическая гипотеза: *H*0: *X* ~ *N*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Число интервалов группировки | Выборочное значение статистики критерия | *p-value* | Статистическое решение при  α = \_\_\_\_\_\_ | Ошибка стат. решения |
| 5 |  |  |  |  |
| 10 |  |  |  |  |
| 15 |  |  |  |  |
|  |  |  |  |  |

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Число интервалов группировки | Выборочное значение статистики критерия | *p-value* | Статистическое решение при  α = \_\_\_\_\_\_ | Ошибка стат. решения |
| 5 |  |  |  |  |
| 10 |  |  |  |  |
| 15 |  |  |  |  |
|  |  |  |  |  |

*в*) Статистическая гипотеза: *H*0: *X* ~ χ2(5)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Число интервалов группировки | Выборочное значение статистики критерия | *p-value* | Статистическое решение при  α = \_\_\_\_\_\_ | Ошибка стат. решения |
| 5 |  |  |  |  |
| 10 |  |  |  |  |
| 15 |  |  |  |  |
|  |  |  |  |  |

*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  α = \_\_\_\_\_\_ | Ошибка стат. решения |
| *X* ~ *N*(*m*, σ) |  |  |  |  |
| *X* ~ *R* |  |  |  |  |
| *X* ~ χ2(5) |  |  |  |  |

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



Эмпирическая и теоретические функции распределения

*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 = \_\_\_\_\_\_\_\_\_

HIST: X, Y ECDF: X, Y

 

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Критерий | Стат. гипотеза, *H*0 | Выборочное значение статистики критерия | *p-value* | Стат. решение при  α = \_\_\_\_\_\_ | Ошибка стат. решения |
| Chi-squared |  |  |  |  |  |
| KS-test |  |  |  |  |  |
| Sign test |  |  |  |  |  |
| U-test |  |  |  |  |  |

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