

Лабораторная работа № 2

«Критерии согласия и однородности выборок»

студента Грачева Игоря группы Б23-504. Дата сдачи: _____
 Ведущий преподаватель: _____ оценка: _____ подпись: _____

Вариант №5

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, σ^2 |
|---------------|-------------|------------------------------|-----------------------|
| $N(5, 2)$ | m, σ | 5 | 4 |

Объём выборки $n_1 = 100$

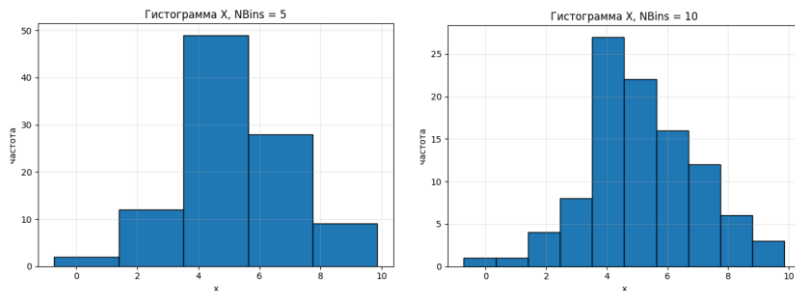
Примечание: для генерации случайных чисел использовать функции **rand**, **randn**, **chi2rnd** (scipy.stats: **uniform.rvs**, **norm.rvs**, **chi2.rvs**)

Выборочные характеристики:

| Среднее, \bar{x} | Оценка дисперсии, s^2 | Оценка с.к.о., s |
|--------------------|-------------------------|--------------------|
| 5.1831 | 3.5019 | 1.8713 |

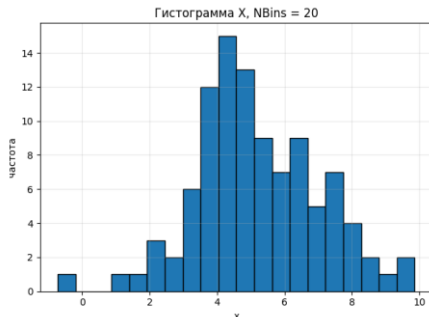
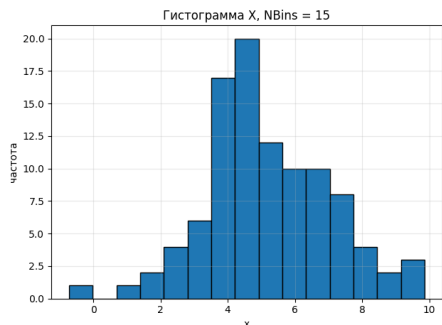
2. Визуальное представление выборки

Гистограммы частот:



$NBins = 5$

$NBins = 10$



$NBins = 15$

$NBins = 20$

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

3. Критерий хи-квадрат

а) Статистическая гипотеза: $H_0: X \sim N(m, \sigma)$

| Число интервалов группировки | Выборочное значение статистики критерия | p -value | Статистическое решение при $\alpha = 0.05$ | Ошибка стат. решения |
|------------------------------|---|------------|--|----------------------|
| 5 | 5.8470 | 0.2109 | Не отклоняется | Нет |
| 10 | 8.1885 | 0.5153 | Не отклоняется | Нет |
| 15 | 14.2908 | 0.4283 | Не отклоняется | Нет |
| 20 | 16.3465 | 0.6340 | Не отклоняется | Нет |

б) Статистическая гипотеза: $H_0: X \sim R$

| Число интервалов группировки | Выборочное значение статистики критерия | p -value | Статистическое решение при $\alpha = 0.05$ | Ошибка стат. решения |
|------------------------------|---|------------|--|----------------------|
| 5 | 70.7 | 4.441e-16 | Отклоняем | Нет |
| 10 | 74 | 2.288e-13 | Отклоняем | Нет |
| 15 | 77.6 | 1.182e-11 | Отклоняем | Нет |
| 20 | 80 | 3.835e-10 | Отклоняем | Нет |

в) Статистическая гипотеза: $H_0: X \sim \chi^2(5)$

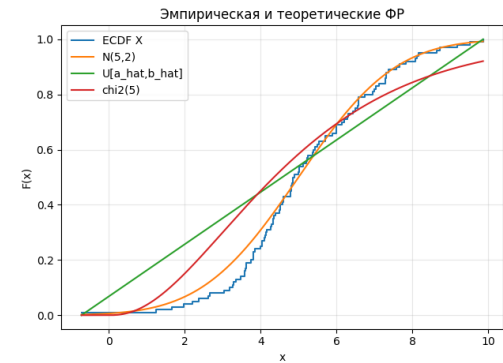
| Число интервалов группировки | Выборочное значение статистики критерия | p -value | Статистическое решение при $\alpha = 0.05$ | Ошибка стат. решения |
|------------------------------|---|------------|--|----------------------|
| 5 | 34.0359 | 7.237e-07 | Отклоняем | Нет |
| 10 | 40.9944 | 5.014e-06 | Отклоняем | Нет |
| 15 | 42.0767 | 1.202e-04 | Отклоняем | Нет |
| 20 | 49.1220 | 1.763e-04 | Отклоняем | 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 $\alpha = 0.05$ | Ошибка стат. решения |
|--------------------------------|---|------------|--|----------------------|
| $X \sim N(m, \sigma)$ | 0.0985 | 0.2685 | Не отклоняем | Нет |
| $X \sim R$ | 0.2787 | 2.329e-07 | Отклоняем | Нет |
| $X \sim \chi^2(5)$ | 0.2483 | 6.423e-06 | Отклоняем | Нет |

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



Эмпирическая и теоретические функции распределения

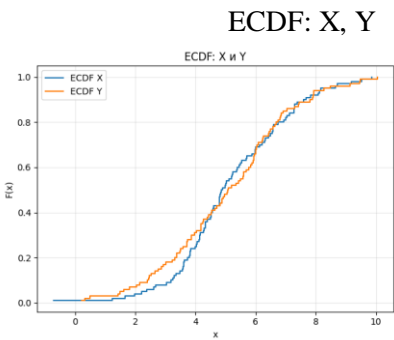
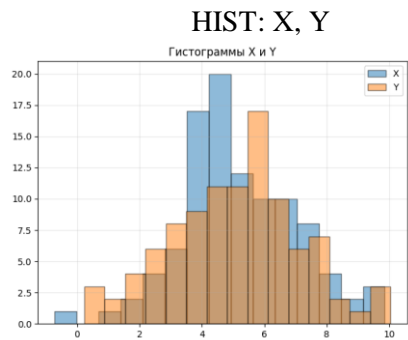
Примечание: для построения графиков использовать функции `ecdf`, `cdf` (`scipy.stats: uniform.cdf`, `norm.cdf`, `chi2.cdf`; `statsmodels.distributions.empirical_distribution: ECDF`)

5. Двухвыборочные критерии

Характеристики наблюдаемой случайной величины Y:

| Распределение | Параметры | Математическое ожидание | Дисперсия |
|---------------|-----------|-------------------------|-----------|
| N(5, 2) | m, σ | 5 | 4 |

Объём выборки $n_2 = 100$



| Критерий | Стат. гипотеза, H_0 | Выборочное значение статистики критерия | p -value | Стат. решение при $\alpha = 0.05$ | Ошибка стат. решения |
|----------|-----------------------|---|------------|-----------------------------------|----------------------|
|----------|-----------------------|---|------------|-----------------------------------|----------------------|

| | | | | | |
|-------------|---------------------|-------------------|--------|----------------|-----|
| Chi-squared | $F_x = F_y$ | $\chi^2 = 4.4159$ | 0.3526 | Не отклоняется | Нет |
| KS-test | $F_x = F_y$ | $D = 0.1$ | 0.7021 | Не отклоняется | Нет |
| Sign test | $\text{Med}(X-Y)=0$ | $N_+=53, N_-=47$ | 0.6173 | Не отклоняется | Нет |
| U-test | $F_x = F_y$ | $U = 5115$ | 0.7797 | Не отклоняется | Нет |

Примечание: при расчетах использовать функции **chi2gof**, **kstest2**, **signtest**, **ranksum** (**scipy.stats: chisquare**, **ks_2samp**; **statsmodels.stats.descriptivestats. sign_test**, **ranksums**)