

Лабораторная работа № 2 «Критерии согласия и однородности выборок»

студента Шамаева Сергея группы Б21-514. Дата сдачи: _____
Ведущий преподаватель: Трофимов Александр Геннадьевич оценка: _____
подпись: _____

Вариант № 4 (19)

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
$X \sim R(5, 15)$	5, 15	10	8.333

Объём выборки $n_1 = 300$

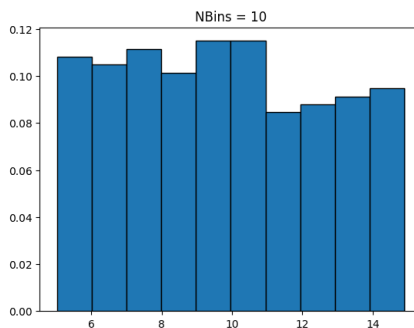
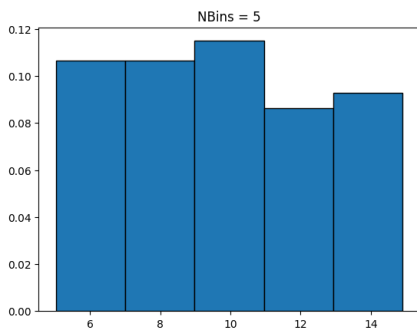
Примечание: для генерации случайных чисел использовать функции **rand**, **randn**, **chi2rnd** (scipy.stats: **uniform.rvs**, **norm.rvs**, **chi2.rvs**)

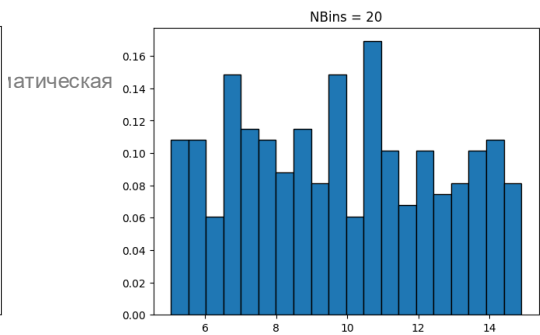
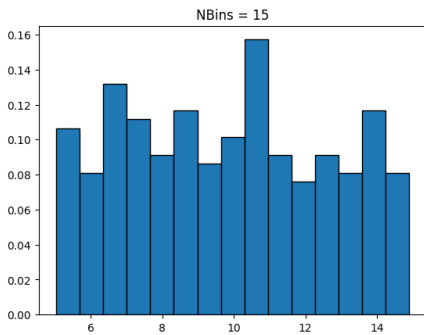
Выборочные характеристики:

Среднее, \bar{x}	Оценка дисперсии, s^2	Оценка с.к.о., s
9.808	7.916	2.813

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

Число интервалов группировки	Выборочное значение статистики критерия	p -value	Статистическое решение при $\alpha = 0.05$	Ошибка стат. решения
5	14.963	0.0005	Не принимается	нет
10	18.574	0.0096	Не принимается	нет
15	40.842	5.205e-05	Не принимается	нет
20	57.998	2.239e-06	Не принимается	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	2.543	0.280	Принимается	нет
10	3.912	0.790	Принимается	нет
15	13.339	0.345	Принимается	нет
20	24.372	0.110	Принимается	нет

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

Число интервалов группировки	Выборочное значение статистики критерия	<i>p-value</i>	Статистическое решение при $\alpha = 0.05$	Ошибка стат. решения
5	830.638	1.772 e-178	Не принимается	нет
10	929.224	3.119 e-194	Не принимается	нет
15	1151.061	5.726 e-237	Не принимается	нет
20	1201.439	4.562 e-243	Не принимается	нет

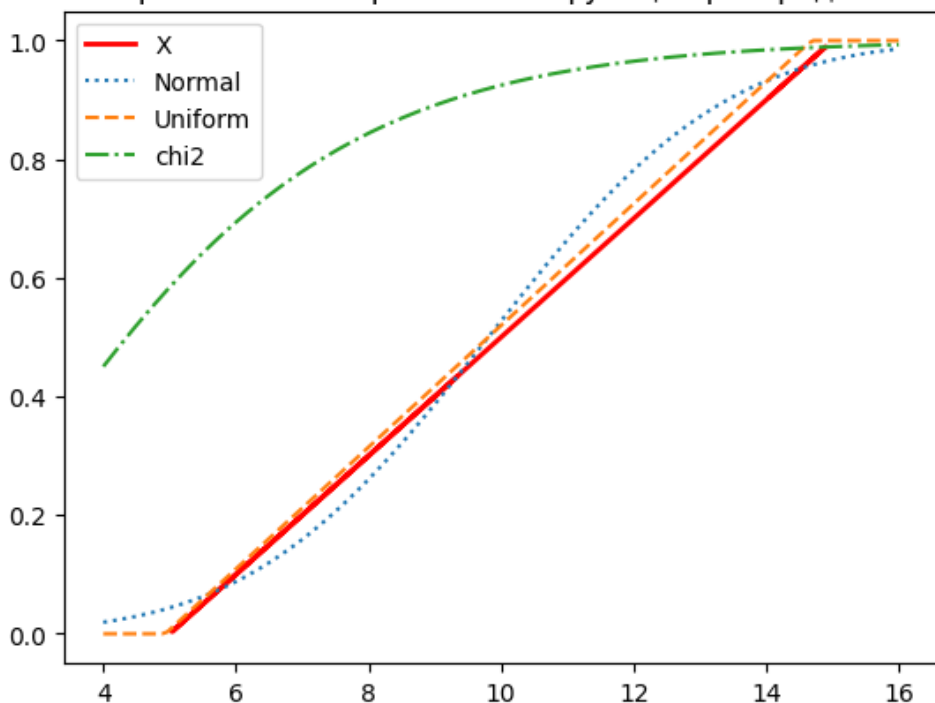
Примечание: при расчетах использовать функции `chi2gof`, `fitdist` (`scipy.stats: histogram`, `chisquare`)

4. Критерий Колмогорова

Статистическая гипотеза, H_0	Выборочное значение статистики критерия	<i>p-value</i>	Статистическое решение при $\alpha = 0.05$	Ошибка стат. решения
$X \sim N(m, \sigma)$	0.067	0.127	Принимается	да
$X \sim R$	0.036	0.808	Принимается	нет
$X \sim \chi^2(5)$	0.607	2.356 e-106	Не принимается	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**)

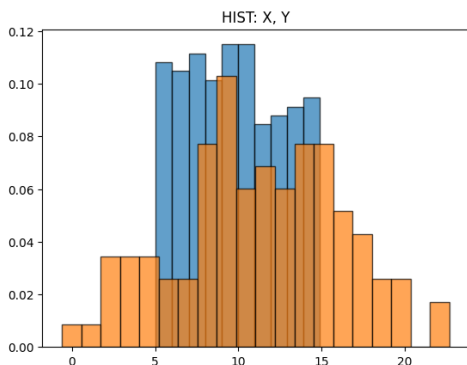
5. Двухвыборочные критерии

Характеристики наблюдаемой случайной величины Y :

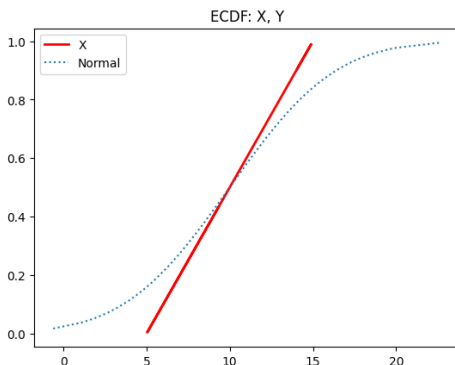
Распределение	Параметры	Математическое ожидание	Дисперсия
$Y \sim N(10, 5)$	10, 5	10	25

Объем выборки $n_2 = 100$

HIST: X, Y



ECDF: X, Y



Критерий	Стат. гипотеза, H_0	Выборочное значение статистики критерия	p -value	Стат. решение при $\alpha = 0.05$	Ошибка стат. решения
Chi-squared	$F(x)=F(y)$	18.606	0.058	Принимается	Да
KS-test	$F(x)=F(y)$	0.19	0.008	Не принимается	Нет
Sign test	$F(x)=F(y)$	-4.0	0.484	Принимается	Да
U-test	$F(x)=F(y)$	-0.25	0.799	Принимается	Да

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