

Лабораторная работа № 1
«Проверка статистических гипотез»

студента Шамаева Сергея группы Б21-514. Дата
сдачи: 16.11.2023

Ведущий преподаватель: Трофимов А. Г. оценка:

подпись: _____

Вариант № 4(19)

Цель работы: изучение функций Statistics and Machine Learning Toolbox™
MATLAB / Python SciPy.stats для проверки статистических гипотез.

1. Исходные данные

Характеристики наблюдаемых случайных величин:

СВ	Распределение	Параметры	Математическое ожидание, m_i	Дисперсия, σ_i^2
X_1	$R(5, 15)$	5, 15	10	8.333
X_2	$N(10, 5)$	10, 5	10	25

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

Выборочные характеристики:

СВ	Среднее, \bar{x}_i	Оценка дисперсии, s_i^2	Оценка с.к.о., s_i	Объем выборки, n_i
X_1	9.808	7.890	2.809	300
X_2	9.503	22.746	4.769	100
<i>Pooled</i>	9.732	11.651	3.413	400

Указание: для расчета использовать функции **mean**, **var**, **std** (**scipy.stats: describe**)

2. Однопараметрические критерии

Для случайной величины X_1 :

Тест	Стат. гипотеза, H_0	Выборочное значение статистики критерия	p -value	Стат. решение при $\alpha = 0.05$	Ошибка стат. решения
z-test	$m = 10$	-1.184	0.236	H_0 принимается	нет
t-test	$m = 10$	-1.184	0.237	H_0 принимается	нет
χ^2 -test (m – изв)	std = 2.887	285.374	0.562	H_0 принимается	нет
χ^2 -test (m – не изв)	std = 2.887	284.423	0.563	H_0 принимается	нет

Указание: для проверки гипотез использовать функции **ztest**, **ttest**, **vartest** (**scipy.stats: ttest_1samp**, **chisquare**)

3. Двухвыборочные критерии

Для случайных величин X_1, X_2 :

Тест	Стат. гипотеза, H_0	Выборочное значение статистики критерия	p -value	Стат. решение при $\alpha = 0.05$	Ошибка стат. решения
2-sample t-test	$m_1 = m_2$	0.601	0.549	H_0 принимается	нет
2-sample F-test (m – изв)	$\sigma_1 = \sigma_2$	0.344	4.152e-09	H_0 не принимается	нет
2-sample F-test (m – не изв)	$\sigma_1 = \sigma_2$	2.901	2.699e-12	H_0 не принимается	нет

Указание: для проверки гипотез использовать функции **ttest2**, **vartest2** (**scipy.stats: ttest_ind, chisquare**)

4. Исследование распределений статистик критерия

Статистическая гипотеза: $H_0: m_1 = m_2 (\sigma_1, \sigma_2 - \text{не изв.})$

Формула расчёта статистики критерия
$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$$

Формула расчёта статистики *P-value*: $2 * \min(1 - F(Z), F(Z))$

Число серий экспериментов $N = 1500$

Теоретические характеристики:

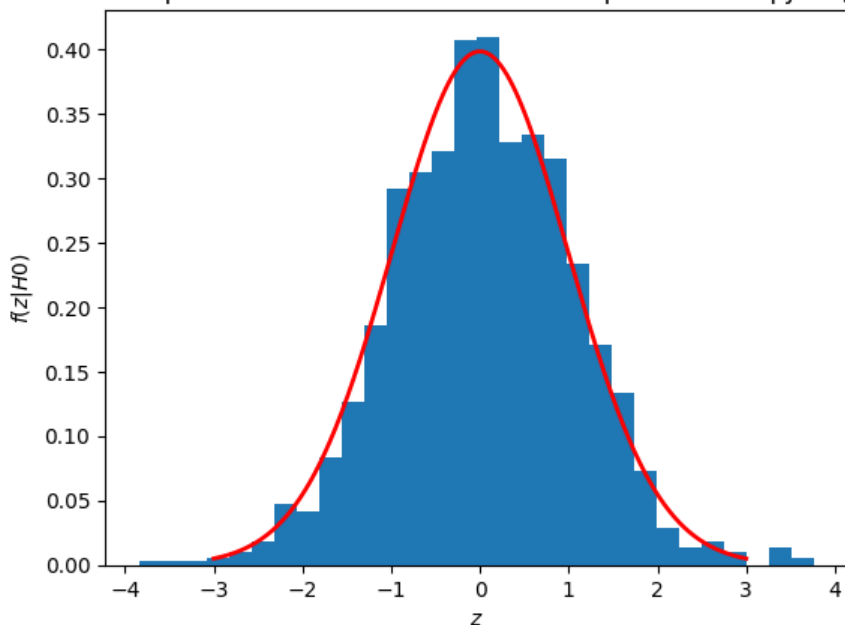
СВ	Распределение в условиях H_0	Параметры	Математическое ожидание	Дисперсия	С.к.о.
Z	$T(n)$	$n = 121$	0	$121/119 = 1.017$	1.009
$P\text{-value}$	$R(a, b)$	$a = 0, b = 1$	0.5	0.083	0.29

Выборочные характеристики:

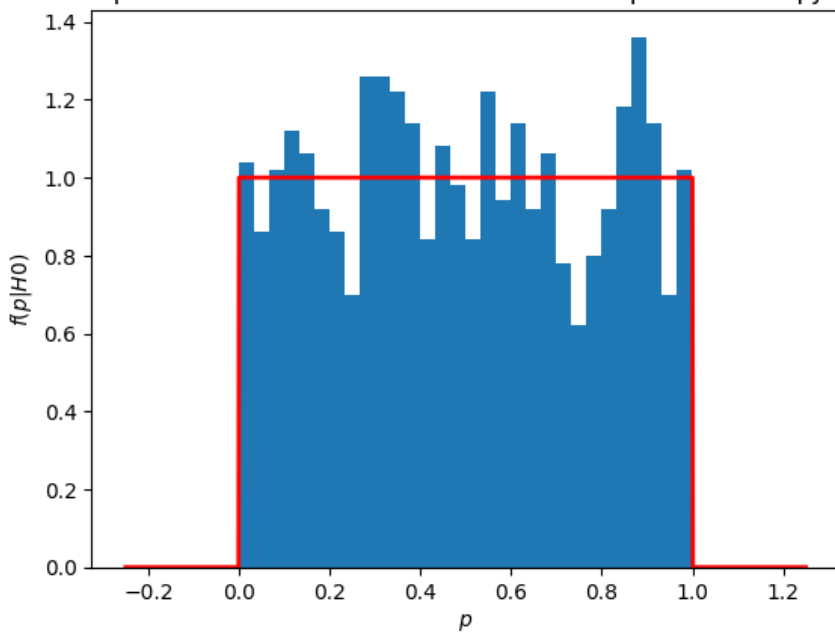
СВ	Среднее	Оценка дисперсии	Оценка с.к.о.
Z	0.043	1.050	1.023
$P\text{-value}$	0.497	0.083	0.288

Указание: при расчете выборочных значений статистики критерия использовать функции **norminv**, **tinv**, **chi2inf**, **finv** (**scipy.stats: norm.ppf, t.ppf, chi2.ppf, f.ppf**)

Гистограмма частот статистики Z и теоретическая функция



Гистограмма частот статистики P-value и теоретическая функция



Указание: для построения гистограмм и теоретических функций плотности использовать функции **hist**, **normpdf**, **tpdf**, **chi2pdf**, **fpdf** (**scipy.stats: norm.pdf, t.pdf, chi2.pdf, f.pdf, histogram; matplotlib.pyplot: hist**)