**Лабораторная работа № 1**

«Проверка статистических гипотез»

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 статистических гипотез.

1. Исходные данные

Характеристики наблюдаемых случайных величин:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| СВ | Распределение | Параметры | Математическое ожидание, *mi* | Дисперсия, |
| *X*1 | N(5, 2) | μ=5, σ=2 | 5 | 4 |
| *X*2 | N(5, 2) | μ=5, σ =2 | 5 | 4 |

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

Выборочные характеристики:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| СВ | Среднее, | Оценка дисперсии, | Оценка с.к.о., | Объем выборки, *ni* |
| *X*1 | 4.92795 | 5.28676 | 2.29929 | 100 |
| *X*2 | 4.93663 | 4.69015 | 2.16568 | 100 |
| *Pooled* | 4.93229 | 4.98845 | 2.23348 | 200 |

*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  α = 0,05 | Ошибка стат. решения |
| z-test | m = 5 | -0.36026 | 0.71865 | Не отвергаем H0 | нет |
| t-test | m = 5 | -0.31337 | 0.75466 | Не отвергаем H0 | нет |
| χ2-test  (m – изв) | σ2 = 4 | 130.97704 | 0.04095 | Отвергаем H0 | да (ложное отклонение) |
| χ2-test  (m – не изв) | σ2 = 4 | 130.84725 | 0.03536 | Отвергаем H0 | да (ложное отклонение) |

*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  α = 0,05 | Ошибка стат. решения |
| 2-sample  t-test | m1 = m2 | -0.02747 | 0.97811 | Не отвергаем H0 | нет |
| 2-sample  F-test  (m – изв) | σ12 = σ22 | 1.12735 | 0.55007 | Не отвергаем H0 | нет |

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

4. Исследование распределений статистик критерия

Статистическая гипотеза: *H*0: σ1 = σ2

Формула расчёта статистики критерия *Z*: Z ∼ F(99, 99) при H0

Формула расчёта статистики *P-value*: P-value = 2(1 – F(Z))

Число серий экспериментов *N* = 10000

Теоретические характеристики:

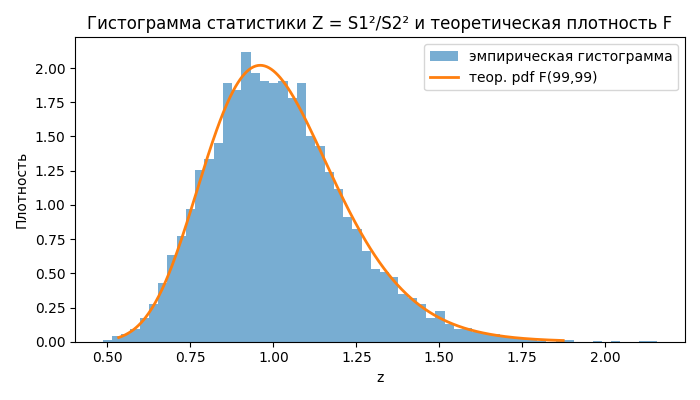
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| СВ | Распределение в условиях *H*0 | Параметры | Математическое ожидание | Дисперсия | С.к.о. |
| *Z* | F(99, 99) | Df1 = 99, df2 = 99 | 1.02062 | 0.04342 | 0.20837 |
| *P-value* | U(0, 1) | A = 0, b = 1 | 0.50000 | 0.08333 | 0.28868 |

Выборочные характеристики:

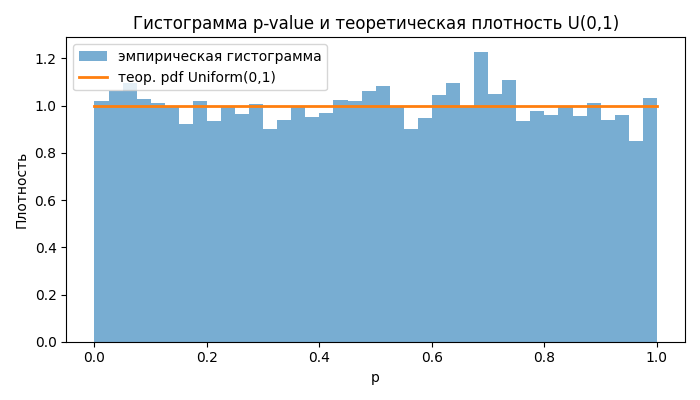
|  |  |  |  |
| --- | --- | --- | --- |
| СВ | Среднее | Оценка дисперсии | Оценка с.к.о. |
| *Z* | 1.01696 | 0.04334 | 0.20818 |
| *P-value* | 0.49837 | 0.08288 | 0.28788 |

*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)**