



Министерство науки и высшего образования Российской Федерации
Калужский филиал
федерального государственного бюджетного
образовательного учреждения высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(КФ МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИУК «Информатика и управление»

КАФЕДРА ИУК4 «Программное обеспечение ЭВМ,

информационные технологии»

Практическое занятие №7

«Критерии согласия»

ДИСЦИПЛИНА: «Методы обработки информации»

Выполнил: студент гр. ИУК4-72Б _____ (_____)
(подпись) (Ф.И.О.)

Проверил: _____ (_____)
(подпись) (Ф.И.О.)

Дата сдачи (защиты):

Результаты сдачи (защиты):

- Балльная оценка:

- Оценка:

Калуга, 2023

Постановка задачи

Пусть проверяется простая гипотеза относительно закона распределения для выборки из ПЗ№2 $H_0: F(x) = F_{\text{теор}}(x)$, при различных уровнях значимости α и для различных объемов выборки.

1. объем $k_1 < 20$ (любые k_1 значений из заданной выборки);
2. объем $k_2 = N$ (полный объем исходной выборки) Используя критерии согласия Колмогорова, ω^2 (Крамера – фон Мизеса), χ^2 Пирсона принять или опровергнуть основную гипотезу.

Результаты выполнения работы

Используемые критерии значимости: 0.05,0.01,0.001.

Используется полная выборка и 11 случайных значений из выборки.

Статистика критерия согласия Колмогорова рассчитывается по формуле:

$$D_n = \sup_x |F_n(x) - F(x)|$$

Статистика критерия согласия Крамера-фон-Мизеса рассчитывается по формуле:

$$S_\omega = n\omega^2 = \frac{1}{12n} + \sum_{i=1}^n \left(F(x_i, \theta) - \frac{2i-1}{2n} \right)^2$$

Статистика критерия согласия Пирсона рассчитывается по формуле:

$$\chi^2 = n \sum_{i=1}^k \frac{\left(\frac{n_i}{n} - P_i(\theta) \right)^2}{P_i(\theta)}$$

Для расчета критических значений используется библиотека `scipy.stats`

Листинг программы:

```
import argparse
import csv
import random
import typing

import numpy as np
import prettytable as pt
import scipy.stats as st
from scipy.stats import norm
```

```

def ft(x):
    mean_value = np.mean(x)
    std_deviation = np.std(x, ddof=1)
    cdf_value = norm.cdf(x, loc=mean_value, scale=std_deviation)
    return cdf_value

def chi_squared_test(points: np.array, expected_frequencies=None,
    bins=None):
    if bins is None:
        bins = int(np.sqrt(len(points)))

    observed_frequencies, bin_edges = np.histogram(points, bins=bins)

    if expected_frequencies is None:
        expected_frequencies = (np.ones_like(observed_frequencies) *
            len(points) / bins)

    chi2 = np.sum(
        (observed_frequencies - expected_frequencies) ** 2
        / expected_frequencies
    )
    dof = bins - 1

    return chi2, dof

def test_hypothesis(sample: np.array, alpha: float) -> pt.PrettyTable:
    sample = np.sort(sample)
    F_emp = np.arange(1, len(sample) + 1) / len(sample)
    F_theor = ft(sample)
    D = np.max(np.abs(F_emp - F_theor))
    W2 = np.sum((F_emp - F_theor) ** 2) + 1 / (
        12 * len(sample))

    chi2, dof = chi_squared_test(sample)

    D_crit = st.kstwobign.ppf(1 - alpha) / np.sqrt(len(sample))
    W2_crit = st.chi2.ppf(1 - alpha, df=1)
    chi2_crit = st.chi2.ppf(1 - alpha, df=len(sample) - 1)
    result = pt.PrettyTable()
    result.field_names = ["Критерий", "Статистика", "Критическое значение",
        "Вывод"]
    result.add_row(
        ["Колмогоров", round(D, 4), round(D_crit, 4),
            "Принять" if D < D_crit else "Отклонить"]
    )
    result.add_row(
        ["Крамер-фон Мизес", round(W2, 4), round(W2_crit, 4),
            "Принять" if W2 < W2_crit else "Отклонить"]
    )
    result.add_row(
        ["Пирсон", round(chi2, 4), round(chi2_crit, 4),
            "Принять" if chi2 < chi2_crit else "Отклонить"]
    )
    return result

```

```

if __name__ == "__main__":
    parser = argparse.ArgumentParser()
    parser.add_argument("-file")

    args = parser.parse_args()
    file = args.file or "../data/Test14.csv"

    points = []
    with open(file, newline='') as csvfile:
        reader = csv.reader(csvfile, delimiter=' ', quotechar='|')
        for row in reader:
            points.append(float("".join(row)))
    points = np.array(points)

    alpha = np.array([0.01, 0.05, 0.001])

    sample_k = 11
    sample = random.sample(list(points), 11)
    sample_result = []
    for a in alpha:
        sample_result.append(test_hypothesis(sample, a))

    all_points_k = len(points)
    all_points = points
    all_points_result = []
    for a in alpha:
        all_points_result.append(test_hypothesis(all_points, a))

    for i in range(len(alpha)):
        print(f"Уровень значимости alpha = {alpha[i]}")
        print(f"Объем выборки k_1 = {sample_k}")
        print(sample_result[i])
        print(f"Объем выборки k_2 = {all_points_k}")
        print(all_points_result[i])

```

Уровень значимости alpha = 0.01				
Объем выборки k_1 = 11				
Критерий	Статистика	Критическое значение	Вывод	
Колмогоров	0.1975	0.4907	Принять	
Краммер-фон Мизес	0.0878	6.6349	Принять	
Пирсон	0.1818	23.2093	Принять	
Объем выборки k_2 = 135				
Критерий	Статистика	Критическое значение	Вывод	
Колмогоров	0.0696	0.1401	Принять	
Краммер-фон Мизес	0.0844	6.6349	Принять	
Пирсон	15.9852	174.9963	Принять	

Рисунок 1 – Значения статистики для уровня значимости $\alpha = 0.01$

Уровень значимости $\alpha = 0.05$

Объем выборки $k_1 = 11$

Критерий	Статистика	Критическое значение	Вывод
Колмогоров	0.1975	0.4095	Принять
Краммер-фон Мизес	0.0878	3.8415	Принять
Пирсон	0.1818	18.307	Принять

Объем выборки $k_2 = 135$

Критерий	Статистика	Критическое значение	Вывод
Колмогоров	0.0696	0.1169	Принять
Краммер-фон Мизес	0.0844	3.8415	Принять
Пирсон	15.9852	162.0156	Принять

Рисунок 2 – Значения статистики для уровня значимости $\alpha = 0.05$

Уровень значимости $\alpha = 0.001$				
Объем выборки $k_1 = 11$				
	Критерий	Статистика	Критическое значение	Вывод
	Колмогоров	0.1975	0.5878	Принять
	Крамeр-фон Мизес	0.0878	10.8276	Принять
	Пирсон	0.1818	29.5883	Принять
Объем выборки $k_2 = 135$				
	Критерий	Статистика	Критическое значение	Вывод
	Колмогоров	0.0696	0.1678	Принять
	Крамeр-фон Мизес	0.0844	10.8276	Принять
	Пирсон	15.9852	190.3313	Принять

Рисунок 3 – Значения статистики для уровня значимости $\alpha = 0.001$