Лабораторная работа 7
Тестирование гипотез о распределениях

Гузовская Александра Чеславовна
Б9123-01.03.02сп

19.05.2025

Критерий Колмогорова

Имеем:
Случайная выборка $X_1, X_2, \dots X_n$ из X
Проверяем нулевую гипотезу $H_0 : F_X = F$ против $H_1 : F_X \neq F$

$$K = \sqrt{n} \cdot \sup_{x \in R} \left| \hat{F}_n(x) - F(x) \right|$$

p-значение: $p = 1 - F_K(k)$

Функция распределения

$$F_K(x) = \sum_{k=-\infty}^{+\infty} (-1)^k \cdot e^{-2k^2x^2} \cdot I(x > 0)$$

Критерий χ^2

Имеем:
Случайная выборка $X_1, X_2, \dots X_n$ из X
Проверяем нулевую гипотезу $H_0 : F_X = F$ против $H_1 : F_X \neq F$

$$\chi^2_{k-1} = \sum_{i=1}^{n} \frac{(n_i - np_i)^2}{np_i} \sim \chi^2(k-1)$$

где
$n_i = \sum_{j=1}^{n} I(X_j \in T_i)$ — количество попаданий в i-ый интервал,
$p_i = P(X \in T_i) = F(\sup T_i) - F(\inf T_i)$ — вероятность попадания в i-ый интервал,
$R = \bigcup_{i=1}^{k} T_i$ — разбиение,
k — количество непересекающихся промежутков разбиения

p-значение: $p = 1 - F_{\chi^2_{k-1}}(\chi^2)$

Критерий Колмогорова-Смирнова

Имеем:
Две случайные выборки
$X_1, X_2, \dots X_n$ из X и $Y_1, Y_2, \dots Y_n$ из Y
Проверяем нулевую гипотезу $H_0 : F_X = F_Y$ против $H_1 : F_X \neq F_Y$

$$K = \sqrt{\frac{n+m}{n \cdot m}} \cdot \sup_{x \in R} \left| \hat{F}_n(x) - \hat{F}_m(x) \right|$$

p-значение: $p = 1 - F_K(k)$

Код программы

```
import numpy as np
import scipy.stats as sps
from statsmodels.distributions.empirical_distribution import ECDF

def kolmogorov_cdf(x):
    """Функция распределения статистики Колмогорова"""
    if x <= 0:
        return 0.0
    k = np.arange(1, 200)
    terms = (-1)**(k-1) * np.exp(-2 * (k**2) * (x**2))
    return 1 - 2 * np.sum(terms)

def kolmogorov_test(sample, cdf):
    """Критерий Колмогорова для проверки соответствия распределению"""
```

```
    n = len(sample)
    Dn = np.max(np.abs(np.arange(1, n+1)/n - cdf(np.sort(sample))))
    p_value = 1 - kolmogorov_cdf(np.sqrt(n) * Dn)
    return p_value

def chi2_test(sample, bins, cdf):
    """Критерий хи-квадрат для проверки соответствия распределению"""
    observed, _ = np.histogram(sample, bins=bins)
    n = len(sample)
    expected = n * np.diff(cdf(bins))
    chi2_stat = np.sum((observed - expected)**2 / expected)
    p_value = 1 - sps.chi2.cdf(chi2_stat, df=len(bins)-1)
    return p_value

def homogenity_test(sample1, sample2):
    """Критерий Колмогорова-Смирнова для проверки однородности"""
    n = len(sample1)
    m = len(sample2)

    combo_sorted = np.sort(np.concatenate([sample1, sample2]))

    ecdf1 = ECDF(sample1)(combo_sorted)
    ecdf2 = ECDF(sample2)(combo_sorted)

    Dn = np.max(np.abs(ecdf1 - ecdf2))
    K = np.sqrt((n * m) / (n + m)) * Dn
    p_value = 1 - kolmogorov_cdf(K)

    p_lib = sps.ks_2samp(sample1, sample2).pvalue
    return p_value, p_lib

def generate_samples(th_in_group):
    """Генерация выборок по номеру в группе"""
    n = 100
    dist_type = th_in_group % 3

    if dist_type == 0:
        samples = [sps.norm.rvs(loc=mu, scale=sigma, size=n) for mu, sigma in [(4,
```

```
    elif dist_type == 1:
        samples = [sps.uniform.rvs(loc=mu, scale=sigma, size=n) for mu, sigma in |
    else:
        samples = [sps.binom.rvs(n=k, p=p, size=n) for k, p in [(13, 0.7), (13, 0.

    return samples

def test_hypotheses(sample1, sample2, sample3, dist_type):
    results = {}

    if dist_type == 'norm':
        mu, sigma = np.mean(sample1), np.std(sample1)
        p_custom = kolmogorov_test(sample1, lambda x: sps.norm.cdf(x, loc=mu, scal
        p_lib = sps.kstest(sample1, 'norm', args=(mu, sigma))[1]
    elif dist_type == 'uniform':
        a = np.min(sample1)
        b = np.max(sample1)
        scale = b - a
        p_custom = kolmogorov_test(sample1, lambda x: sps.uniform.cdf(x, loc=a, sc
        p_lib = sps.kstest(sample1, 'uniform', args=(a, scale))[1]
    elif dist_type == 'binom':
        p_hat = np.mean(sample1)/len(sample1)
        bins = np.arange(0, 14) - 0.5
        observed = np.histogram(sample1, bins=bins)[0]
        expected = len(sample1) * np.diff(sps.binom.cdf(bins, n=13, p=p_hat))

        bins = np.arange(0, max(sample1)+2) - 0.5
        observed, _ = np.histogram(sample1, bins=bins)
        expected = len(sample1) * np.diff(sps.binom.cdf(bins, n=13, p=0.7))
        p_custom = chi2_test(sample1, bins, lambda x: sps.binom.cdf(x, n=13, p=0.7
        p_lib = sps.chisquare(observed, f_exp=expected).pvalue
    results['same_dist'] = {'custom': p_custom, 'library': p_lib}

    p_custom_other = kolmogorov_test(sample1, lambda x: sps.uniform.cdf(x, -3, 6))
    p_lib_other = sps.kstest(sample1, 'uniform', args=(-3, 6))[1]
    results['other_dist'] = {'custom': p_custom_other, 'library': p_lib_other}

    results['homogenity'] = {
```

```
        'sample1 vs sample2': homogenity_test(sample1, sample2),
        'sample1 vs sample3': homogenity_test(sample1, sample3),
        'sample2 vs sample3': homogenity_test(sample2, sample3)
    }

    p_shapiro = sps.shapiro(sample1)[1]
    results['normality'] = p_shapiro

    return results

if __name__ == "__main__":
    np.random.seed(42)
    th_in_group = 9

    sample1, sample2, sample3 = generate_samples(th_in_group)
    results = test_hypotheses(sample1, sample2, sample3, 'norm')

    # test_values = [0.5, 1.0, 1.5]
    # for x in test_values:
    #     print(f"kolmogorov_cdf({x}) = {kolmogorov_cdf(x)} vs scipy: {sps.kstwobi

    print("\nРезультаты проверки гипотез:\n")
    print(f"1. Соответствие своему распределению: custom={results['same_dist']['cu
    print(f"2. Соответствие другому распределению: custom={results['other_dist']['
    print("3. Проверка однородности:")
    [print(f"{pair}: кастом={res[0]:.4f} | lib={res[1]:.4f} | ={abs(res[0]-res[1])
        for pair, res in results['homogenity'].items()]
    print(f"4. Проверка нормальности (Шапиро-Вилк): p={results['normality']:.4f}\n
```

Вывод программы

```
Результаты проверки гипотез:

1. Соответствие своему распределению: custom=0.9762, scipy=0.9467
2. Соответствие другому распределению: custom=0.0000, scipy=0.0000
3. Проверка однородности:
sample1 vs sample2: кастом=0.6994 | lib=0.7021 | Δ=2.7e-03
sample1 vs sample3: кастом=0.0243 | lib=0.0241 | Δ=2.5e-04
sample2 vs sample3: кастом=0.0101 | lib=0.0099 | Δ=2.1e-04
4. Проверка нормальности (Шапиро-Вилк): p=0.6552
```