Санкт-Петербургский Государственный Университет

Поляков Иван Михайлович

Отчёт по Лабораторной Работе № 8_2

Однофакторный непараметрический анализ, критерий Краскела-Уоллиса

Направление 01.04.02: «Прикладная математика и информатика»
Образовательная программа ВМ.5505.2021: «Математическое и информационное обеспечение экономической деятельности»

Преподаватель:
доктор технических наук,
профессор Буре Владимир Мансурович

Санкт-Петербург
2022 г.

1 Постановка задачи

Необходимо проверить гипотезу однородности данных ниже с помощью критерия Краскела-Уоллиса на уровне значимости $\alpha = 0.05$

```
Исходные данные
1: [10.1, 7.3, 5.6, 6.2, 8.4, 8.1, 8.0, 7.6, 5.3, 7.2]
2: [11.7, 12.2, 11.8, 7.8, 8.9, 9.9, 12.4, 11.0, 10.3, 13.8, 10.5, 9.8, 9.1]
3: [10.2, 12.0, 8.8, 8.7, 10.5, 11.0, 9.1]
```

2 Ход работы

Критерий Краскела-Уоллиса применяется, когда нельзя сказать ничего определённого об альтернативах к нулевой гипотезе, и неизвестно, их какого распределения были взяты выборки.

Результаты данных составляют k независимых выборок, взятых из распределений с непрерывными функциями распределения.

$$\begin{matrix} X_{11}, \ldots, X_{1n_1} \\ X_{21}, \ldots, X_{2n_2} \\ \ldots \\ X_{k1}, \ldots, X_{kn_k} \end{matrix} \tag{1}$$

Предлагается упорядочить все величины X_{ij} по возрастанию и обозначить ранг r_{ij} у каждого числа X_{ij} во всей совокупности.

Далее формулируется нулевая гипотеза H_0 : *все k выборок однородны, то есть являются выборками из одного и того же закона распределения*, а также формулируется альтернативная гипотеза H_1 : *отрицание нулевой гипотезы.*

Пусть $\overline{R} = \frac{1}{n_i} \cdot \sum\limits_{j=1}^{n_i} r_{ij}$ - средний ранг по выборке. Предполагается, что если между выборками нет систематических различий, то введённые средние ранги не должны отличаться от общего среднего ранга, рассчитанного по формуле $\frac{n+1}{2}$, где $n = \sum\limits_{i=1}^{k} n_i$.

Статистика данного критерия имеет вид

$$H = \frac{12}{n(n+1)} \cdot \sum_{i=1}^{k} n_i \cdot \left(\overline{R}_i - \frac{n+1}{2}\right)^2. \tag{2}$$

Критическое значение вычисляется из распределения χ^2. Гипотеза H_0 отвергается на уровне значимости α, если статистика критерия $H > \chi^2_{1=\alpha}$, где $\chi^2_{1=\alpha}$ – квантиль уровня $1 - \alpha$ распределения хи-квадрат с $k - 1$ степенями свободы, в противном случае нет основания отвергнуть её на уровне значимости α.

Из полученных результатов можно заключить, что все выборки были взяты из разных генеральных совокупностей, распределённых по разным законам распределения.

Результаты рабоыт критерия приведены ниже.

```
Гипотеза H0 об однородности выборок на уровне значимости 0.05 отклоняется,
так как статистика H больше критического значения распределения хи-квадрата
(15.6101 > 5.991)
```