УДК 519.233.4

СРАВНЕНИЕ ТРЕХ И БОЛЕЕ НЕЗАВИСМЫХ ГРУПП С ИСПОЛЬЗОВАНИЕМ НЕПАРАМЕТРИЧЕСКОГО КРИТЕРИЯ КРАСКЕЛА – УОЛЛИСА В ПРОГРАММЕ STATA

© 2014 г. **Т. Н. Унгуряну, *А. М. Гржибовский**

Северный государственный медицинский университет, г. Архангельск
*Норвежский институт общественного здравоохранения,
г. Осло, Норвегия

В статье рассматриваются теоретические основы применения критерия Краскела – Уоллиса для сравнения трех и более независимых групп количественных или порядковых данных. Приводится пример расчета критерия «вручную» с помощью формул, а также пошаговый алгоритм использования его в пакете статистических программ STATA. Особое внимание уделяется необходимым условиям, которые должны соблюдаться для применения критерия Краскела – Уоллиса. Представлены рекомендации для оформления результатов в научных публикациях.

Ключевые слова:
непараметрические тесты, независимые группы, критерий Краскела – Уоллиса

Критерий Краскела – Уоллиса (Kruskal-Wallis H-test) является непараметрическим аналогом однофакторного дисперсионного анализа для сравнения трех и более независимых групп [2, 6]. Данный критерий используется, если распределение в группах не подчиняется закону нормального распределения, что нередко встречается в медицинских исследованиях, особенно в выборках малого объема. В таких ситуациях следует либо трансформировать имеющиеся данные с помощью различных арифметических преобразований до достижения нормальности распределения [4], после чего можно будет применять дисперсионный анализ, либо применять критерий Краскела – Уоллиса, иногда еще называемый непараметрическим дисперсионным анализом. Критерий Краскела – Уоллиса рассчитывается с использованием не фактических значений переменных, а их рангов, поэтому является методом выбора при сильно скошенных распределениях. С его помощью проверяют нулевую гипотезу о том, что медианные значения признака в популяциях, из которых были извлечены исследуемые выборки, не различаются [1].

Расчет критерия Краскела – Уоллиса. Сначала все значения, независимо от того, какой выборке они принадлежат, упорядочивают по возрастанию, как если бы это была одна объединенная выборка. Каждому значению присваивается ранг от наименьшего к наибольшему – номер его места в упорядоченном ряду. Совпадающим значениям присваивают одинаковый ранг, равный среднему тех мест, которые эти величины делят между собой в общем упорядоченном ряду. После ранжирования следует проверить, чтобы общее количество рангов было равно количеству наблюдений в объединенной выборке. Затем вычисляют суммы рангов, относящихся к каждой группе (R_i). Далее подсчитывают тестовую статистику критерия Краскела – Уоллиса (H) по формуле:

$$H = \frac{12}{N \times (N+1)} \times \Sigma \frac{R_i^2}{n_i} - 3 \times (N+1),$$

где R_i – сумма рангов для каждой группы; n_i – количество наблюдений в каждой группе; N – общее количество наблюдений в объединенной выборке.

Если число сравниваемых групп 3, а количество наблюдений в каждой группе не менее 5 (для четырех групп – общее число наблюдений не менее 10), то расчетное значение тестовой статистики H сравнивают с критическим значением хи-квадрат Пирсона (χ^2), так как распределение H близко распределению χ^2 с числом степеней свободы df = k – 1, где k – число групп. Если расчетное значение H равно или превышает критическое значение χ^2, то H_0 отвергается.

Если число наблюдений в группах менее 5, то в качестве критического значения используют табличные значения распределения Краскела –

Уоллиса. В этом случае, если расчетное значение H равно или превышает критическое значение $H_{0,05}$, H_0 отвергается. При использовании таблицы критических значений критерий Краскела — Уоллиса установить различия между тремя группами можно, если минимальное число наблюдений в одной группе составляет 3, а в двух других группах — по 2 наблюдения. При сопоставлении четырех или пяти групп минимальное число наблюдений в каждой группе должно быть равно 2.

Пример. Для оценки дозовой нагрузки химическими веществами, загрязняющими питьевую воду, изучалось количество потребляемой для питья водопроводной воды среди разных возрастных групп населения. Поскольку исследуемые группы являются независимыми, а данные имеют ненормальное распределение, то для проверки нулевой гипотезы о равенстве медианных значений количества потребляемой для питья воды в популяциях детей, подростков и взрослого населения использовался критерий Краскела — Уоллиса (табл. 1).

Таблица 1

Ранжирование значений водопотребления в трех возрастных группах для расчета тестовой статистики критерия Краскела — Уоллиса

Группа 1, дети (n=8)		Группа 2, подростки (n=7)		Группа 3, взрослые (n=9)	
Вода, л/день	Ранг	Вода, л/день	Ранг	Вода, л/день	Ранг
1,22	1	1,47	6	1,56	10
1,24	2	1,52	7,5	1,58	11
1,31	3,5	1,55	9	1,81	13
1,31	3,5	1,70	12	1,89	15
1,45	5	1,93	16	2,00	18
1,52	7,5	2,00	18	2,00	18
1,84	14	3,00	23	2,55	21
2,52	20			2,58	22
				4,00	24
Сумма рангов	56,5		91,5		152

$$H = \frac{12}{24 \times (24+1)} \times \left(\frac{56{,}5^2}{8} + \frac{91{,}5^2}{7} + \frac{152^2}{9} \right) - 3 \times (24+1) =$$

$$= 0{,}02 \times (399{,}0 + 1196{,}0 + 2567{,}1) - 75 = 8{,}24$$

Расчетное значение тестовой статистики критерия Краскела — Уоллиса (H) в данном примере оказалось равным 8,24. Для оценки нулевой гипотезы необходимо расчетное значение критерия (H) сравнить с табличным значением критерия χ^2. Из таблицы критических значений критерия χ^2 для числа степеней свободы df = k — 1 = 3 — 1 = 2 и уровня статистической значимости 0,05 критическое значение χ^2 составляет 5,99. Так как расчетное значение больше критического, то имеют место статистически значимые ($p < 0{,}05$) различия в количестве потребляемой питьевой воды в разных возрастных группах населения.

Как и дисперсионный анализ, критерий Краскела — Уоллиса поможет выяснить, имеются ли различия между группами, но не сможет показать, между какими из групп эти различия существуют. При обнаружении статистически значимых различий между группами с помощью критерия Краскела — Уоллиса далее следует проводить апостериорные сравнения с помощью критерия Манна — Уитни или двухвыборочного критерия Вилкоксона [3, 7]. Следует помнить, что, поскольку STATA не дает возможности автоматически проводить апостериорные сравнения с помощью непараметрических методов статистики, исследователям самим необходимо рассчитывать новые критические уровни значимости исходя из представленных выше формул или как показано в табл. 2. Во всех приведенных примерах исследователи должны принимать во внимание проблему множественных сравнений и рассчитывать новые критические уровни значимости [5].

Если мы принимаем традиционные 0,05 за критический уровень значимости, то вероятность ошибки первого типа составляет 5 %, значит, вероятность отсутствия этой ошибки составит 0,95, или 95 %. Если мы проводим три сравнения (сравниваем попарно три группы, проверяем три статистические гипотезы), то вероятность отсутствия ошибки первого типа в любом из сравнений составит $0{,}95^n$, то есть $0{,}95^3 = 0{,}857$, или 85,7 %, а значит, вероятность сделать хотя бы одну ошибку первого типа будет равна $1 - 0{,}95^n = 1 - 0{,}857 = 0{,}142$, или 14,2 % вместо декларируемых 5 %. В такой ситуации необходимо использовать меньший критический уровень значимости, который рассчитывается по формуле: $p^* = 1 - 0{,}95^{1/n}$, где n — количество производимых сравнений. Для данного примера $p^* = 1 - 0{,}95^{1/3} = 0{,}0170$, то есть различия между группами можно считать статистически значимыми только если $p < 0{,}0170$. Из этого следует, что в публикациях, где встречается «$p < 0{,}05_{1\text{-}2}$, $p > 0{,}05_{2\text{-}3}$, $p > 0{,}05_{1\text{-}3}$», совершенно невозможно сделать вывод о статистической значимости различий между группами 1 и 2, а потому результаты должны интерпретироваться читателем минимум как сомнительные.

Для ситуации с тремя сравниваемыми группами количество возможных попарных сравнений равно количеству изучаемых групп (табл. 2). Если групп больше, то количество возможных попарных сравнений можно рассчитать по формуле: $n = 0{,}5N\,(N - 1)$, где N — количество изучаемых групп. Например, если имеется 12 групп (например, при попарных сравнениях среднемесячных значений тех или иных показателей), то максимальное количество возможных сравнений составит $n = 0{,}5 \times 12 \times (12 - 1) = 66$. Если оставить критический уровень значимости без изменений (0,05), то верятность случайного обнаружения статистически значимых различий составит $1 - 0{,}95^{66} = 0{,}966$, или 96,6 %. Критический уровень значимости для данного примера при проведении всех 66 сравнений должен быть установлен на уровне $1 - 0{,}95^{1/66} = 0{,}00078$, то есть статистически значимыми могут считаться только те различия, для которых $p < 0{,}00078$,

Таблица 2

Количество возможных сравнений, вероятность ошибки первого типа и уровни значимости для наиболее часто встречающегося в литературе количества сравниваемых групп

Количество сравниваемых групп	2	3	4	5
Количество попарных сравнений	1	3	6	10
Вероятность случайного выявления статистически значимых различий (ошибка 1 типа) для множественных попарных сравнений, %	5	14	26	40
Критический уровень значимости	0,0500	0,0170	0,0085	0,0051
Количество сравнений с контрольной группой	1	2	3	4
Вероятность случайного выявления статистически значимых различий (ошибка 1 типа) для множественных сравнений с контрольной группой, %	5	10	14	19
Критический уровень значимости	0,0500	0,0253	0,0170	0,0127

Расчет критерия Краскела — Уоллиса в STATA. Для использования критерия Краскела — Уоллиса в STATA [9] необходимо открыть диалоговое окно kwallis-Kruskal-Wallis equality-of-populations rank, которое открывается при помощи меню Statistics ® Summaries, tables, and tests ® Nonparametric tests of hypotheses ® Kruskal-Wallis rank test (рис. 1). В поле Outcome variable помещается изучаемая переменная (Water). В поле Variable defining groups помещается группировочная переменная (Group).

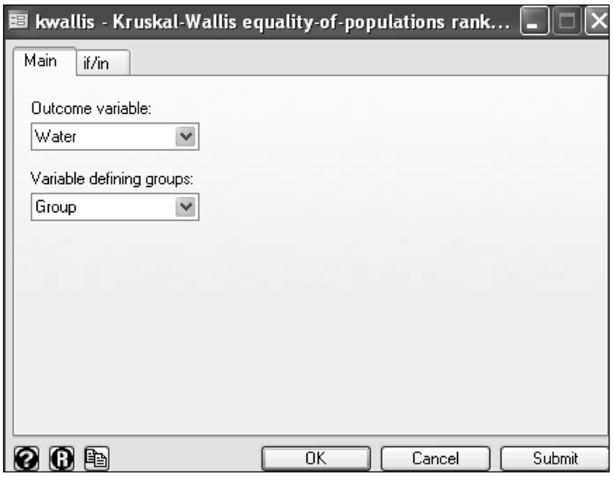

Рис. 1. Диалоговое окно для расчета критерия Краскела — Уоллиса

Результаты сравнения групп с помощью критерия Краскела — Уоллиса представлены на рис. 2, где в таблице указаны номер группы, число наблюдений в каждой группе (Obs) и сумма рангов для каждой группы (Rank Sum). Под таблицей даны значения критерия, обозначенные как chi-squared, количество степеней свободы (df) и достигнутый уровень значимости различий (probability). Результаты показывают, что есть статистически значимые различия в количестве потребляемой для питья воды в трех возрастных группах населения. Для того чтобы узнать, какие группы различаются между собой, следует провести попарные сравнения групп при помощи критерия Манна — Уитни с новым критическим уровнем значимости: 0,05 / 3 = 0,017.

```
. kwallis Water, by(Group)
Kruskal-Wallis equality-of-populations rank test
------------------------
Group  Obs  Rank Sum
----------------------
   1     8    56.50
   2     7    91.50
   3     9   152.00
------------------------
chi-squared = 8.244 with 2 d.f. probability = 0.0162
chi-squared with ties =  8.265 with 2 d.f. probability = 0.0160
```

Рис. 2. Результаты сравнения водопотребления в группах детей, подростков и взрослых с помощью критерия Краскела — Уоллиса

Результаты попарных сравнений, выполненных с помощью критерия Манна — Уитни (или двухвыборочного критерия Вилкоксона, Two-sample Wilcoxon rank-sum test), показали, что достигнутый уровень статистической значимости (p) составил между 1-й и 2-й группами 0,055; 2-й и 3-й группами — 0,184; 1-й и 3-й группами — 0,009. Таким образом, среднее количество потребляемой для питья воды статистически значимо различается только между 1-й группой (детьми) и 3-й группой (взрослыми).

При представлении результатов сравнения групп с помощью критерия Краскела — Уоллиса следует указывать значение тестовой статистики (H) и достигнутый уровень статистической значимости (p), например, H = 8,24; p = 0,016. В случае апостериорных сравнений с помощью критерия Манна — Уитни необходимо указать z-значение и величину достигнутого уровня статистической значимости (p).

Можно ли применять непараметрический критерий Краскела — Уоллиса, если данные подчиняются закону нормального распределения? Критерий имеет несколько меньшую статистическую мощность, чем дисперсионный анализ, поэтому при нормальном распределении и выполнении прочих условий дисперсионный анализ является методом выбора. Некоторые исследователи не рекомендуют применять параметрические методы (в том числе и дисперсионный анализ), если объем каждой из групп составляет менее 30 наблюдений, даже если выборочные данные имеют нормальное распределение [8]. Можно ли использовать дисперсионный анализ при отклонении распределения от нормального? При наличии больших выборок с равными дисперсиями дисперсионный анализ достаточно устойчив к небольшим отклонениям распределения от нормального, особенно при равных объемах выборок. При малых выборках применение дисперсионного анализа для скошенных распределений может привести к сильно искаженным результатам, поэтому рекомендуется в такой ситуации применять критерий Краскела — Уоллиса.

Список литературы

1. *Банержи А.* Медицинская статистика понятным языком: вводный курс. М. : Практическая медицина, 2007. 287 с.

2. *Гланц С.* Медико-биологическая статистика. М. : Практика, 1998. 460 с.

3. *Гржибовский А. М.* Анализ количественных данных для двух независимых групп // Экология человека. 2008. № 2. С. 54—61.

4. *Гржибовский А. М.* Типы данных, проверка распределения и описательная статистика // Экология человека. 2008. № 1. С. 52—58.

5. *Гржибовский А. М.* Анализ трех и более независимых групп количественных данных // Экология человека. 2008. № 3. С. 50—58.

6. *Медик В. А., Токмачев М. С.* Математическая статистика в медицине : учеб. пособие. М. : Финансы и статистика, 2007. 800 с.

7. *Петри А., Сэбин К.* Наглядная медицинская статистика / пер. с англ. под ред. В. П. Леонова. М. : ГЭОТАР-Медиа, 2009. 168 с.

8. *Chang Y. H.* Biostatistics 101: Data presentation // Singapore Medical Journal. 2003. N 6. P. 280—285.

9. *Hamilton L. C.* Statistics with STATA: Updated for Version 10. Brooks/Cole, Cengage Learning, 2009. 491 p.

References

1. Banerjee A. *Meditsinskaya statistika ponyatnym yazykom: vvodnyi kurs* [Medical Statistics Made Clear: Introduction]. Moscow, 2007, 287 p.

2. Glantz S. Basic biostatistics (translated from English into Russian). Moscow, Praktika Publ., 1998. 459 p.

3. Grjibovski A. M. Analysis of quantitative data for two independent groups. *Ekologiya cheloveka* [Human Ecology]. 2008, 2, pp. 54-61. [in Russian]

4. Grjibovski A. M. Data types, control of distribution and descriptive statistics. *Ekologiya cheloveka* [Human Ecology]. 2008, 1, pp. 52-58. [in Russian]

5. Grjibovski A. M. Analysis of three and more independent groups of quantitative data. *Ekologiya cheloveka* [Human Ecology]. 2008, 3, pp.50-58. [in Russian]

6. Medik V. A., Tokmachev M. S. *Matematicheskaya statistika v meditsine* [Mathematical Statistics in Medicine]. Moscow, Finansy i statistika Publ., 2007, 800 p.

7. Petrie A., Sabin K. *Naglyadnaya statistika v meditsine* [Medical Statistics at Glance]. Moscow, GEOTAR-Media Publ., 2003, 144 p.

8. Chang Y. H. Biostatistics 101: Data presentation. *Singapore Medical Journal.* 2003, 6, pp. 280-285.

9. Hamilton L. C. *Statistics with STATA: Updated for Version 10.* Brooks/Cole, Cengage Learning, 2009, 491 p.

ANALYSIS OF THREE INDEPENDENT GROUPS USING NON-PARAMETRIC KRUSKAL-WALLIS TEST IN STATA SOFTWARE

T. N. Unguryanu, *A. M. Grjibovski

International School of Public Health, Northern State Medical University Arkhangelsk, Russia
**Department of International Public Health, Norwegian Institute of Public Health, Oslo, Norway*

In this paper, we have presented theoretical principles of analysis of three or more independent groups using the Kruskal-Wallis test. We have presented calculations using formulas as well as a step-by-step algorithm for use of this test in STATA software. Moreover, we have given practical examples with special emphasis on assumptions of this test. We have also presented recommendations for presentation of results in scientific publications.

Keywords: non-parametric tests, independent groups, Kruskal-Wallis test

Контактная информация:

Гржибовский Андрей Мечиславович — профессор, доктор медицины, старший советник Норвежского института общественного здоровья, г. Осло, Норвегия; директор Архангельской международной школы общественного здоровья ГБОУ ВПО «Северный государственный медицинский университет» Министерства здравоохранения Российской Федерации, г. Архангельск

Адрес: Nasjonalt folkehelseinstitutt, Pb 4404 Nydalen, 0403 Oslo, Norway

Тел.: +47 22048319,

E-mail: angr@fhi.no