

## МНОЖЕСТВЕННЫЕ СРАВНЕНИЯ В БИМЕДИЦИНСКИХ ИССЛЕДОВАНИЯХ: ПРОБЛЕМА И СПОСОБЫ РЕШЕНИЯ

© 2020 г. <sup>1</sup>А. Н. Наркевич, <sup>1</sup>К. А. Виноградов, <sup>2,3,4,5</sup>А. М. Гржибовский

<sup>1</sup>ФГБОУ ВО «Красноярский государственный медицинский университет им. проф. В. Ф. Войно-Ясенецкого», г. Красноярск; <sup>2</sup>ФГБОУ ВО «Северный государственный медицинский университет», г. Архангельск;  
<sup>3</sup>Западно-Казахстанский медицинский университет им. Марата Оспанова, г. Актобе, Казахстан;  
<sup>4</sup>Казахский Национальный Университет им. аль-Фараби, г. Алматы, Казахстан;  
<sup>5</sup>ФГАУ ВО «Северо-Восточный федеральный университет им. М. К. Аммосова», г. Якутск

В русскоязычной биомедицинской научной литературе одной из широко распространенных, но редко обсуждаемых проблем статистического анализа является проблема множественных сравнений. Она заключается в том, что увеличение числа проверяемых в процессе исследования статистических гипотез приводит к росту вероятности возникновения ошибок первого типа и ложных выводов о наличии различий там, где их нет, что связано с риском проведения необоснованных медицинских вмешательств, профилактических мероприятий и сопряжено с необоснованными расходами. В зарубежной научной литературе этой проблеме посвящено множество работ, как предлагающих новые подходы к решению проблемы множественных сравнений, так и описывающих применение уже общепризнанных методов. Однако в русскоязычной литературе такие работы встречаются редко. Целью данной статьи является восполнение имеющегося пробела путём представления способов решения проблемы множественных сравнений в медико-биологических исследованиях. Представлены методы решения проблемы на этапах планирования, статистического анализа и интерпретации результатов. Рассмотрены одношаговые методы, такие как поправки Бонферрони и Сидака (Шидака), и многошаговые: Холма – Бонферрони, Холма – Сидака (Шидака), Бенджамини – Хохберга и другие. Приведены примеры их использования, а также алгоритм их реализации как вручную, так и с помощью пакета статистических программ SPSS.

**Ключевые слова:** множественные сравнения, поправка Бонферрони, поправка Сидака (Шидака), метод Холма, метод Бенджамини – Хохберга

## MULTIPLE COMPARISONS IN BIOMEDICAL RESEARCH: THE PROBLEM AND ITS SOLUTIONS

<sup>1</sup>A. N. Narkevich, <sup>1</sup>K. A. Vinogradov, <sup>2,3,4,5</sup>A. M. Grjibovski

<sup>1</sup>Voino-Yasenetsky Krasnoyarsk State Medical University, Krasnoyarsk, Russia;  
<sup>2</sup>Northern State Medical University, Arkhangelsk, Russia; <sup>3</sup>West Kazakhstan Marat Ospanov Medical University, Aktobe, Kazakhstan; <sup>4</sup>Al-Farabi Kazakh National University, Almaty, Kazakhstan;  
<sup>5</sup>M. K. Ammosov North-Eastern Federal University, Yakutsk, Russia

One of the most common but rarely discussed problems in Russian biomedical research is a problem of multiple comparisons. When a researcher performs pairwise comparisons of means in several groups the number of tested statistical hypotheses increases leading to inflation of the alpha-error. In international scientific literature this issue is well-described and several solutions are offered. The aim of this article is to describe the problem of alpha error inflation and present methods for solving the problem of multiple comparisons. The methods suggested in this paper can be applied at the stages of research planning, data analysis and interpretation of the results. Bonferroni, Sidak, Holm-Bonferroni, Holm-Sidak and the Benjamin-Hochberg methods are described in details. We also present user-friendly examples for manual calculations as well as a description of implementation of the suggested solutions using SPSS software.

**Key word:** multiple comparisons, Bonferroni correction, Sidak correction, Holm correction, Benjamini-Hochberg correction

### Библиографическая ссылка:

Наркевич А. Н., Виноградов К. А., Гржибовский А. М. Множественные сравнения в биомедицинских исследованиях: проблема и способы решения // Экология человека. 2020. № 10. С. 55–64.

### For citing:

Narkevich A. N., Vinogradov K. A., Grjibovski A. M. Multiple Comparisons in Biomedical Research: the Problem and its Solutions. *Ekologiya cheloveka* [Human Ecology]. 2020, 10, pp. 55-64.

В настоящее время одной из широко распространенных, но редко обсуждаемых проблем статистического анализа является проблема множественных сравнений. Эта проблема заключается в том, что увеличение числа проверяемых в процессе исследования

статистических гипотез приводит к увеличению вероятности возникновения ошибок первого типа [2, 3, 21]. Увеличение вероятности возникновения ошибок первого типа приводит к тому, что исследователи могут констатировать наличие случайных, но статистически

значимых различий или связей, что в свою очередь приводит к ошибочным заключениям и выводам. Особенно актуальна эта проблема для генетических исследований, при проведении которых современная техника позволяет измерять тысячи, десятки тысяч различных параметров, в последующем подвергаемых статистическому анализу и сравнению [9, 12, 15]. В то же время в русскоязычной биомедицинской литературе часто встречается сравнение нескольких групп без проведения коррекции на инфляцию ошибки первого типа, что делает результаты таких исследований как минимум сомнительными, особенно при дихотомизации результатов на « $p < 0,05$ » и « $p > 0,05$ », что является малоинформативным способом представления результатов проверки статистических гипотез и затрудняет понимание результатов читателями в случае множественных сравнений.

В зарубежной научной литературе данной проблеме посвящена масса работ, как предлагающих новые подходы к решению проблемы множественных сравнений, так и описывающих применение уже общепризнанных методов [11, 13, 14, 19]. Однако в отечественной литературе такие работы встречаются довольно редко [1, 4, 5].

В чем же заключается проблема множественных сравнений? В биомедицинской литературе общепринятым считается использовать уровень альфа-ошибки, равный 5 %, т. е. за критическое значение уровня значимости берется 0,05. Это означает, что мы в 5 % случаев готовы отклонить верную нулевую гипотезу, то есть принять решение о наличии различий там, где их на самом деле нет. Это называется ошибкой первого рода или типа. Если допустить, что истинных различий между сравниваемыми группами нет, то величина достигнутого уровня значимости ( $p$ ) будет соответствовать вероятности обнаружения выявленных или еще более выраженных отличий между сравниваемыми группами в исследованиях с аналогичными объемами выборки. Если мы берем 0,05 за критический уровень значимости, то вероятность ошибки первого типа будет 5 %. Вероятность отсутствия этой ошибки составит 0,95, или 95 %. При проведении, например, шести сравнений (сравниваем попарно четыре группы – 1–2, 1–3, 1–4, 2–4, 2–3, 3–4) вероятность отсутствия ошибки первого типа в любом из сравнений составит  $0,95^6$ , то есть  $0,95^6 = 0,815$ , или 81,5 %, а значит, вероятность сделать хотя бы одну ошибку первого типа будет равна  $1 - 0,95^6 = 1 - 0,815 = 0,185$ , или 18,5 % вместо объявленных 5 %! Таким образом становится понятным, что в публикациях, где результаты представлены наподобие « $p < 0,05_{1-2}$ ,  $p > 0,05_{2-3}$ ,  $p > 0,05_{1-3}$ ,  $p < 0,05_{1-4}$ ,  $p < 0,05_{2-4}$ ,  $p > 0,05_{2-3}$ ,  $p > 0,05_{3-4}$ » (чаще, конечно, мы видим такие результаты в таблицах), абсолютно невозможно сделать вывод о статистической значимости различий между группами 1 и 2, 1 и 4 и 2 и 4, а потому результаты должны интерпретироваться в лучшем случае как сомнительные.

Максимальное количество возможных сравнений в зависимости от количества групп в исследовании рассчитывается с помощью формулы:  $n = 0,5 \times N \times (N - 1)$ , где  $N$  – количество изучаемых групп. Нам очень нравится пример, который был описан ранее, но который по-прежнему актуален. «При попарных сравнениях среднемесячных значений тех или иных показателей максимальное количество возможных сравнений составит  $n = 0,5 \times 12 \times (12 - 1) = 66$ . Если оставить критический уровень значимости без изменений (0,05), то вероятность случайного обнаружения статистически значимых различий составит  $1 - 0,95^{66} = 0,966$ , или 96,6 %» [2]. То есть, если не обращать внимание на проблему множественных сравнений, риск ошибочных выводов крайне велик!

В данной статье мы рассматриваем способы решения проблемы множественных сравнений, которые могут применяться в медико-биологических исследованиях. Для удобства изложения информации мы разделили их на три группы по этапу исследования, на котором они могут применяться: 1) на этапе планирования исследования; 2) этапе статистического анализа; 3) этапе интерпретации результатов статистического анализа.

#### Решение проблемы множественных сравнений на этапе планирования исследования

Рассмотрим решение проблемы множественных сравнений на этапе планирования исследования на примере. Исследователем планируется сравнить число госпитализаций за 10 лет у пациентов с сахарным диабетом, которые лечились двумя разными методами, и пациентов, которые не лечились. В данном случае I группа (опытная группа № 1) – пациенты с сахарным диабетом, которые лечились методом № 1, II группа (опытная группа № 2) – пациенты с сахарным диабетом, которые лечились методом № 2, III группа (контрольная группа) – пациенты с сахарным диабетом, которые не принимали лечения, а признак, который сравнивается в этих трех группах, – число госпитализаций за 10 лет (рис. 1).

Группа пациентов без лечения (контрольная)	Группа пациентов с лечением методом № 1 (опытная № 1)	Группа пациентов с лечением методом № 2 (опытная № 2)
--------------------------------------------	-------------------------------------------------------	-------------------------------------------------------

Рис. 1. Три группы пациентов, страдающих сахарным диабетом

Классически при проведении таких экспериментов исследователи прибегают к попарному сравнению имеющихся групп. То есть если исследователь имеет три группы, то попарная оценка различий осуществляется между I и II, I и III, II и III группами. Таким образом, исследователь осуществляет проверку трех гипотез о наличии различий между группами (рис. 2).

Опыт статистического анализа данных показывает, что даже при наличии трех и более групп от попарного сравнения данных групп в целом можно отказаться, но при этом решить поставленные задачи [6]. На самом



Рис. 2. Этапы сравнения трех групп пациентов, страдающих сахарным диабетом

деле при планирования такого эксперимента можно ограничиться двумя сравнениями. Если принять во внимание исходную задачу и группы, которые имеются у исследователя, то скорее всего перед ним стоит задача доказать, что число госпитализаций у пациентов без лечения отличается от числа госпитализаций у пациентов с лечением. Следующая задача — доказать, что число госпитализаций у пациентов с лечением методом № 1 отличается от числа госпитализаций у пациентов с лечением методом № 2.

Если перед исследователем стоят действительно две описанные выше на примере задачи, то эти задачи могут быть решены путем всего двух сравнений (рис. 3). Для этого можно объединить пациентов с лечением методами № 1 и № 2 в одну группу, а затем сравнить число госпитализаций в группах пациентов без лечения и пациентов с лечением методом № 1 и методом № 2. И на следующем этапе сравнить число госпитализаций в группах пациентов с лечением методом № 1 и пациентов с лечением методом № 2.

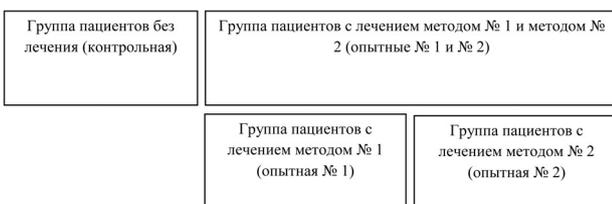


Рис. 3. Сравнение трех групп в два этапа

Таким образом, если имеются три и более группы сравнения, практически всегда необходимо будет перейти к попарным сравнениям, но при этом еще на этапе планирования эксперимента можно снизить число данных сравнений и сформировать несколько иные группы пациентов. Надо отметить, что способ, описанный в настоящем разделе, подойдет далеко не всегда, но в первую очередь следует оценить, возможно ли с его помощью уменьшить число сравнений при планировании эксперимента.

С помощью программного обеспечения SPSS данный подход может быть реализован путем проведения плановых сравнений (contrasting) в ходе дисперсионного анализа, если соблюдаются условия для его применения. Детали применения дисперсионного анализа в SPSS, Statistica и Stata описаны в [1,

4, 5]. Для проведения плановых сравнений средних арифметических в нескольких группах необходимо в выпадающем меню «Анализ (Analyze)» выбрать «Сравнение средних (Compare Means)», в котором, в свою очередь, выбрать «Дисперсионный анализ» (One-Way ANOVA). В появившемся диалоговом окне (рис. 4) выбираем зависимую переменную и группировочную переменную, как описано в [2], поэтому во избежание повторного представления материала остановимся только на плановых сравнениях, которые выбираем посредством нажатия на кнопку «Contrasts» в нижнем ряду, после чего появится диалоговое окно как на рис. 5

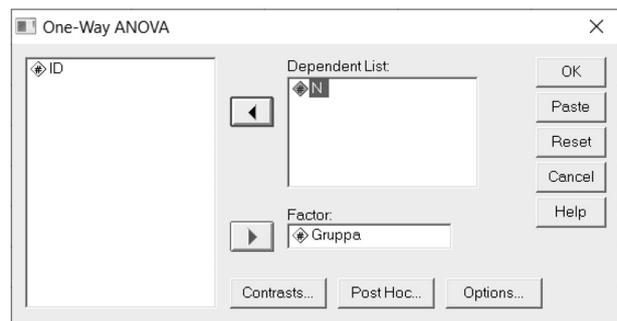


Рис. 4. Диалоговое окно для дисперсионного анализа с выбранными переменными для анализа в SPSS

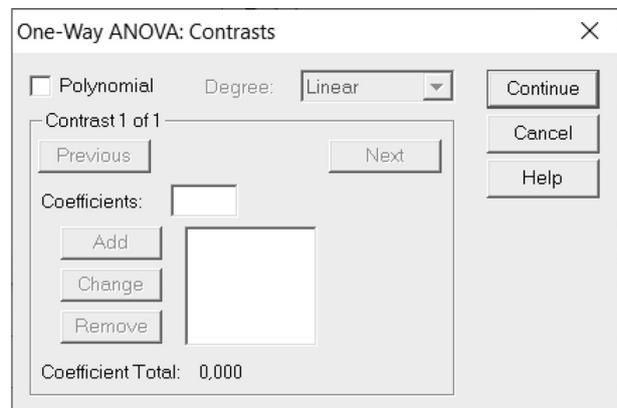


Рис. 5. Диалоговое окно для плановых сравнений (Contrasts) в SPSS

Для этого в соответствующем меню необходимо дать обозначения группам, чтобы программа «поняла», какие сравнения надо проводить. Особенностью кодировки групп при проведении плановых сравнений в SPSS при наличии нескольких групп является необходимость обеспечения суммы номеров групп, равной нулю. Группа, которая не участвует в сравнениях, кодируется как 0. На практическом примере эта особенность выглядит понятнее. Если имеются три группы (как на рис. 1–3), то для сравнения контрольной и объединенной опытной группы необходимо задать коды для групп в виде –2, 1 и 1. Таким образом, значение первой группы и сумма значений второй будут равны нулю, то есть SPSS «поймет», что мы хотим сделать. Для сравнения двух опытных групп надо прежде всего исключить

из анализа контрольную группу, а двум опытным группам присвоить значения  $-1$  и  $1$ , что в сумме также составляет  $0$ . Для первого сравнения вводим вышеупомянутые коэффициенты, как показано на рис. 6. После введения коэффициентов (после каждого нажимаем кнопку «Add»), следующее плановое сравнение вводим после нажатия на кнопку «Next».

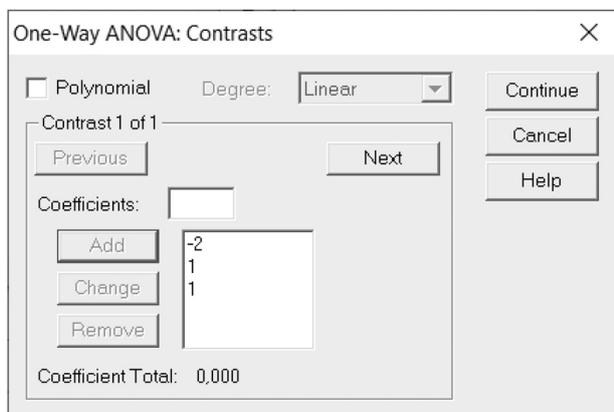


Рис. 6. Коэффициенты для осуществления планового сравнения контрольной группы и объединенной группы из двух экспериментальных

Для второго сравнения вводим коэффициенты, как показано на рис. 7.

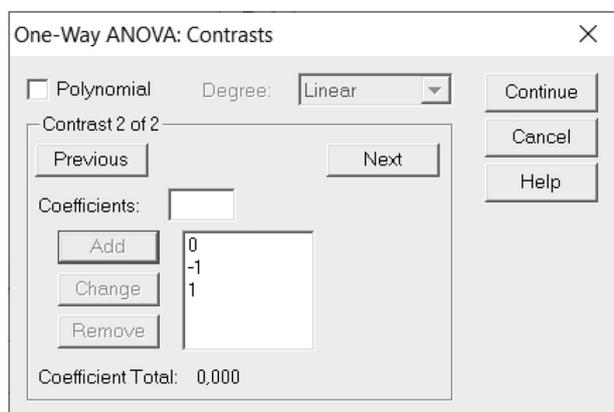


Рис. 7. Коэффициенты для осуществления планового сравнения двух экспериментальных групп

Запускаем анализ кнопкой «Continue», затем «OK». В результате программа выдает три таблицы с ответами, первая из которых содержит результат дисперсионного анализа (для особенностей интер-

претации — см. [2]). Следующие две таблицы посвящены плановым сравнениям. Таблица «Contrasts coefficients» повторяет введенные коэффициенты для первого и второго сравнений (рис. 8).

**Contrast Coefficients**

Contrast	Gruppa		
	1	2	3
1	-2	1	1
2	0	-1	1

Рис. 8. Таблица «Contrasts Coefficients» в SPSS. Объяснения в тексте

Результаты проверки статистических гипотез в двух сравнениях представлены в таблице «Contrast Tests» (рис. 9). Номера сравнений соответствуют тем, что указаны в предыдущей таблице и пояснены выше.

В табл. 9 показаны средние значения разности между средними для запланированных сравнений (Value of Contrast), стандартные ошибки для этих разностей, значение  $t$ -критерия с указанием числа степеней свободы ( $df$ ) и достигнутого уровня значимости (Sig (2-tailed)). Из таблицы видно, что есть статистически значимые различия между контрольной группой и комбинированной экспериментальной, но нет между двумя экспериментальными группами независимо от того, допускаем ли мы соблюдение условия равенства дисперсий.

**Решение проблемы множественных сравнений на этапе статистического анализа**

Если избежать попарных сравнений путем перегруппировки изучаемых пациентов на этапе планирования не удастся, можно воспользоваться способами уменьшения числа сравнений групп, которые применяются в апостериорных тестах Ньюмена — Кейлса и Даннета при проведении дисперсионного анализа [10, 16, 18].

Процедура, используемая в апостериорном тесте Ньюмена — Кейлса заключается в следующем. На первом этапе для каждой исследуемой группы рассчитывается среднее арифметическое, затем группы упорядочиваются по мере увеличения или уменьшения среднего арифметического. На следующих этапах производится сравнение последней группы с I, II, III и так далее до предпоследней. Далее сравнивается предпоследняя группа с I, II, III и так далее

**Contrast Tests**

			Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
Contrast							
N	Assume equal variances	1	-5,40	1,289	-4,190	26	,000
		2	-,80	,718	-1,114	26	,275
	Does not assume equal variances	1	-5,40	1,472	-3,669	11,630	,003
		2	-,80	,624	-1,283	17,611	,216

Рис. 9. Таблица «Contrast Tests» в SPSS. Объяснения в тексте

до предпоследней. Суть процедуры уменьшения числа сравнений заключается в том, что если между какими-то парами групп нет статистически значимых различий, то дальнейшее сближение не происходит, так как при последующем сближении априори нет статистически значимых различий.

Процедура, используемая в апостериорном тесте Даннета, похожа на предыдущую, только она применяется в том случае, когда среди групп есть контрольная группа, с которой необходимо сравнить все остальные. На первом этапе для каждой исследуемой группы также рассчитывается среднее арифметическое, затем группы упорядочиваются по мере увеличения или уменьшения среднего арифметического, далее контрольная группа сравнивается с наиболее от нее отличающейся по значению среднего арифметического. Суть процедуры заключается в том, что опять же если между контрольной группой и какой-то из групп различий нет, то дальнейшее сближение не осуществляется.

Таким образом, исследователем может быть уменьшено число гипотез, которые проверяются в процессе эксперимента. Следует отметить, что описанные в этом разделе процедуры подойдут только в случае использования параметрических тестов, т. е. в случае подчинения сравниваемых данных закону нормального распределения и соблюдения прочих условий их применения. Для попарного сравнения непараметрических тестов данные процедуры непригодны.

В SPSS решение проблемы множественных сравнений на этапе статистического анализа найдено посредством включения 18 статистических критериев, призванных держать под контролем ошибку первого типа, при проведении попарных сравнений (Post-hoc Tests), из которых наилучшее сочетание мощности и контроля за ошибкой первого типа предлагает критерий REGWQ (Ryan, Einot, Gabriel, Welsh Q-критерий), который можно рекомендовать как критерий выбора при сравнении любого количества групп приблизительно равного объема. Если же объемы групп различаются, мы рекомендуем GT-2 критерий Хохберга (Hochberg's GT-2 test). Для обоих вышеперечисленных критериев необходимо соблюдение условия равенства дисперсий. При несоблюдении этого условия целесообразно применять критерий Games-Howell. Для сравнения нескольких групп с контрольной в SPSS применяется критерий Даннета (Dunnnett's test). Алгоритм применения вышеупомянутых критериев детально описан в наших более ранних публикациях [2, 4].

#### **Решение проблемы множественных сравнений на этапе интерпретации результатов статистического анализа**

Если ни один из предыдущих способов и процедур не позволяет в процессе планирования и статистического анализа уменьшить число проверяемых гипотез, то уже после проведенного статистического анализа также возможно скорректировать результаты для

решения проблемы множественных сравнений. Для этого используются различные методы корректировки значений  $p$  и  $\alpha$ .

Значение  $p$  (уровень значимости) — значение, которое получает исследователь по результатам применения какого-либо статистического теста (например,  $p = 0,015$  или  $p = 0,896$ ). Значение  $\alpha$  — граничное значение уровня значимости, которое используется при интерпретации результатов применения статистических тестов. Как правило, в медико-биологических исследованиях в качестве  $\alpha$  применяется значение 0,050 — если  $p \leq 0,050$ , то нулевая гипотеза отклоняется, т. е. различия статистически значимы, а если  $p > 0,050$ , то нулевая гипотеза принимается, т. е. различия статистически незначимы.

Все методы корректировки значений  $p$  и  $\alpha$  также можно разделить на две группы: одношаговые и многошаговые. К одношаговым методам корректировки относятся поправки Бонферрони и Сидака (Шидака) [8, 17, 20, 22]. К многошаговым методам относятся методы Холма — Бонферрони, Холма — Сидака (Шидака) и метод Бенджамини — Хохберга [7, 14].

#### **Поправка Бонферрони**

Рассмотрим наиболее известный из всех представленных методов — поправку Бонферрони. При проведении множественных сравнений данная поправка позволяет скорректировать значение  $\alpha$ , которое должно использоваться для принятия решения о принятии или отклонении нулевой гипотезы. Корректировка производится по формуле:

$$\alpha_B = \frac{\alpha}{m},$$

где  $\alpha$  — первоначальный уровень альфа (0,050);  $\alpha_B$  — скорректированный уровень  $\alpha$  с помощью поправки Бонферрони;  $m$  — число сравнений (гипотез).

Таким образом, если исследователь первоначально использовал в качестве  $\alpha$  значение 0,050 и произвел 10 попарных сравнений, то в качестве граничного значения  $\alpha$  необходимо использовать значение 0,005. Если при каких-то сравнениях  $p \leq 0,005$ , то только в них можно считать наличие статистически значимых различий. В остальных случаях ( $p > 0,005$ ) необходимо признать отсутствие статистически значимых различий.

Корректировка значения  $\alpha$  не всегда удобна при представлении результатов исследований в статье или диссертации, так как при представлении результатов проверки гипотез принято указывать значение  $p$ . Помимо этого на разных этапах исследования  $\alpha$  может быть различным, что затрудняет как представление результатов, так и их чтение. Проще всего скорректировать полученные значения  $p$ , а граничное значение  $\alpha$  так и останется равным 0,050. Корректировка значений  $p$  производится по формуле:

$$p_{Adjusted} = p \cdot m,$$

где  $p$  — полученное по результатам сравнения зна-

чение  $p$ ;  $p_{Adjusted}$  – скорректированное значение  $p$ ;  $m$  – число сравнений (гипотез).

Таким образом, если исследователь по результатам сравнения между какими-либо группами получил  $p = 0,024$ , но при этом произвел десять попарных сравнений, то в качестве значения  $p$  необходимо использовать  $p = 0,240$ . В свою очередь скорректированное значение  $p$  сравнивается с привычным значением  $\alpha$ , равным 0,050.

В связи с тем, что корректировка с помощью поправки Бонферрони значений  $\alpha$  и  $p$  осуществляется в одно действие, данный метод относится к одношаговым.

**Поправка Сидака (Шидака)**

Поправка Сидака (Шидака) является менее известной, но менее консервативной при решении проблемы множественных сравнений большого числа групп. Данная поправка применяется так же, как и поправка Бонферрони, за исключением того, что скорректированные значения  $\alpha$  и  $p$  рассчитываются с применением иных несколько более сложных формул. Скорректированное значение  $\alpha$  рассчитывается по формуле:

$$\alpha_s = 1 - (1 - \alpha)^{\frac{1}{m}}$$

где  $\alpha$  – первоначальный уровень альфа (0,050);  $\alpha_s$  – скорректированный уровень  $\alpha$  с помощью поправки Сидака (Шидака);  $m$  – число сравнений (гипотез).

Скорректированное значение  $p$ :

$$p_{Adjusted} = 1 - (1 - p)^m,$$

где  $p$  – полученное по результатам сравнения значение  $p$ ;  $p_{Adjusted}$  – скорректированное значение  $p$ ;  $m$  – число сравнений (гипотез).

**Метод Холма – Бонферрони**

Метод Холма – Бонферрони относится уже к многошаговым процедурам. На первом шаге применения метода реальные, полученные в ходе статистического анализа, значения  $p$  сортируются по возрастанию:

$$p_1 \leq \dots \leq p_n,$$

где  $p_1$  – наименьшее значение  $p$  из всех полученных;  $p_n$  – наибольшее значение  $p$  из всех полученных.

Затем для каждого  $i$ -того значения  $p$  рассчитывается свое собственное скорректированное граничное значение  $\alpha$  по формуле:

$$\alpha_{H-B}^i = \frac{\alpha}{m - i + 1},$$

где  $\alpha$  – первоначальный уровень альфа (0,050);  $\alpha_{H-B}^i$  – скорректированный уровень  $\alpha$  по методу Холма – Бонферрони;  $m$  – число сравнений (гипотез);  $i$  – порядковый номер значения  $p$  из упорядоченного по возрастанию списка.

На каждом шаге после расчета скорректированного значения  $\alpha$  осуществляется проверка условия:

$$p_1 \leq \alpha_{H-B}^i$$

Если на первом шаге нулевая гипотеза отвергнута, т. е. условие выполнено и различия статистически значимы, то осуществляется переход к следующему шагу ( $i = i + 1$ ). Алгоритм останавливается, когда принимается первая нулевая гипотеза. В таком случае все остальные нулевые гипотезы тоже принимаются.

Коррекция значения  $p$  для приведения в статью или диссертации также осуществляется пошагово с применением формулы:

$$p_{Adjusted}^i = p \cdot (m - i + 1),$$

где  $p$  – полученное по результатам сравнения значение  $p$ ;  $p_{Adjusted}^i$  – скорректированное значение  $p$ ;  $m$  – число сравнений (гипотез);  $i$  – порядковый номер значения  $p$  из упорядоченного по возрастанию списка.

Если на каком-либо шаге скорректированное значение  $p$  меньше, чем предыдущее, то скорректированное  $p$  значение приравнивается к предыдущему значению  $p$ .

**Метод Холма – Сидака (Шидака)**

Метод Холма – Сидака (Шидака) по своей этапности похож на метод Холма – Бонферрони. На этапе применения метода полученные в результате статистического анализа значения  $p$  также сортируются по возрастанию.

Собственное скорректированное граничное значение  $\alpha$  для каждого  $i$ -того значения  $p$  рассчитывается по формуле:

$$\alpha_{H-S}^i = 1 - (1 - \alpha)^{\frac{1}{m-i+1}},$$

где  $\alpha$  – первоначальный уровень альфа (0,050);  $\alpha_{H-S}^i$  – скорректированный уровень  $\alpha$  по методу Холма – Сидака (Шидака);  $m$  – число сравнений (гипотез);  $i$  – порядковый номер значения  $p$  из упорядоченного по возрастанию списка.

После каждого расчета скорректированного граничного значения  $\alpha$  для принятия или отклонения нулевой гипотезы также проверяется условие:

$$p_1 \leq \alpha_{H-S}^i$$

Если на каком-либо шаге нулевая гипотеза не отклоняется, то дальнейший расчет скорректированных значений  $\alpha$  не осуществляется, а все последующие нулевые гипотезы принимаются.

Коррекция значения  $p$  с применением метода Холма – Сидака (Шидака) для приведения в статью или диссертации осуществляется пошагово по формуле:

$$p_{Adjusted}^i = 1 - (1 - p)^{m-i+1},$$

где  $p$  – полученное по результатам сравнения значение  $p$ ;  $p_{Adjusted}^i$  – скорректированное значение  $p$ ;  $m$  – число сравнений (гипотез);  $i$  – порядковый номер значения  $p$  из упорядоченного по возрастанию списка.

В отношении метода Холма – Сидака (Шидака) также применяется условие: если на каком-либо шаге скорректированное значение  $p$  меньше, чем предыдущее, то скорректированное  $p$  значение приравнивается предыдущему значению  $p$ .

**Метод Бенджамини – Хохберга**

Метод Бенджамини – Хохберга по этапности абсолютно идентичен двум предыдущим. Однако расчет скорректированных значений  $\alpha$  и  $p$  осуществляется по иным формулам, за счет чего данный метод является наименее консервативным из рассматриваемых в статье методов.

Собственное скорректированное граничное значение  $\alpha$  для каждого  $i$ -того значения  $p$  рассчитывается по формуле:

$$\alpha_{B-H}^i = i \cdot \frac{\alpha}{m},$$

где  $\alpha$  – первоначальный уровень альфа (0,050);  $\alpha_{B-H}^i$  – скорректированный уровень  $\alpha$  по методу Бенджамини – Хохберга;  $m$  – число сравнений (гипотез);  $i$  – порядковый номер значения  $p$  из упорядоченного по возрастанию списка.

Коррекция значения  $p$  с применением метода Бенджамини – Хохберга для приведения в статье или диссертации также осуществляется пошагово по формуле:

$$p_{Adjusted}^i = \frac{p_i \cdot m}{i},$$

где  $p$  – полученное по результатам сравнения значение  $p$ ;  $p_{Adjusted}^i$  – скорректированное значение  $p$ ;  $m$  – число сравнений (гипотез);  $i$  – порядковый номер

значения  $p$  из упорядоченного по возрастанию списка.

В отношении метода Бенджамини – Хохберга также следует отметить, что если на каком-либо шаге скорректированное значение  $p$  меньше, чем предыдущее, то скорректированное  $p$  значение приравнивается предыдущему значению  $p$ .

**Ручная реализация методов коррекции значений  $\alpha$  и  $p$**

Предположим, что в ходе попарного сравнения пяти групп между собой исследователем были получены следующие значения  $p$  – 0,000034, 0,001, 0,010, 0,015, 0,028, 0,037, 0,043, 0,050, 0,087, 0,128. При игнорировании проблемы множественных сравнений исследователь может сделать заключение об отклонении восьми гипотез (т. е. сделать заключение о наличии между восемью парами групп статистически значимых отличий), так как первые восемь значений  $p \leq 0,050$ .

Решая проблему множественных сравнений с помощью приведенных формул, используя поправку Бонферрони (табл. 1), исследователь отклонит только две гипотезы. Это можно установить двумя путями. Во-первых, сравнивая имеющиеся значения  $p$  со скорректированными значениями  $\alpha_B$ . Во-вторых, сравнивая скорректированные значения  $p_{Adjusted}$  с классической границей 0,050. В любом случае заключения будут одинаковыми – отклоняются только две нулевые гипотезы с двумя наименьшими значениями  $p$ .

Таблица 1

**Значения  $p$ , скорректированные значения  $\alpha$  и  $p$ , принятие или отклонение нулевых гипотез при применении поправки Бонферрони**

Параметр	Значение									
$p$	0,000034	0,001	0,010	0,020	0,028	0,037	0,043	0,050	0,087	0,128
$\alpha_B$	0,005	0,005	0,005	0,005	0,005	0,005	0,005	0,005	0,005	0,005
$p_{Adjusted}$	0,000340	0,010	0,100	0,200	0,280	0,370	0,430	0,500	0,870	1,000
$H_0$	–	–	+	+	+	+	+	+	+	+

Таблица 2

**Значения  $p$ , скорректированные значения  $\alpha$  и  $p$ , принятие или отклонение нулевых гипотез при применении поправки Сидака (Шидака)**

Параметр	Значение									
$p$	0,000034	0,001	0,010	0,020	0,028	0,037	0,043	0,050	0,087	0,128
$\alpha_S$	0,005	0,005	0,005	0,005	0,005	0,005	0,005	0,005	0,005	0,005
$p_{Adjusted}$	0,000339	0,010	0,096	0,183	0,247	0,314	0,356	0,401	0,598	0,746
$H_0$	–	–	+	+	+	+	+	+	+	+

Таблица 3

**Значения  $p$ , скорректированные значения  $\alpha$  и  $p$ , принятие или отклонение нулевых гипотез при применении метода Холма – Бонферрони**

Параметр	Значение									
$p$	0,000034	0,001	0,010	0,020	0,028	0,037	0,043	0,050	0,087	0,128
$i$	1	2	3	4	5	6	7	8	9	10
$m-i+1$	10	9	8	7	6	5	4	3	2	1
$\alpha_{H-B}$	0,005	0,006	0,006	0,007	0,008	0,010	0,013	0,017	0,025	0,050
$p_{Adjusted}$	0,000340	0,009	0,080	0,140	0,168	0,185	0,185	0,185	0,185	0,185
$H_0$	–	–	+	+	+	+	+	+	+	+

Таблица 4

Значения  $p$ , скорректированные значения  $\alpha$  и  $p$ , принятие или отклонение нулевых гипотез при применении метода Холма – Сидака (Шидака)

Параметр	Значение									
	0,000034	0,001	0,010	0,020	0,028	0,037	0,043	0,050	0,087	0,128
$p$	0,000034	0,001	0,010	0,020	0,028	0,037	0,043	0,050	0,087	0,128
$i$	1	2	3	4	5	6	7	8	9	10
$m-i+1$	10	9	8	7	6	5	4	3	2	1
$\alpha_{H-S}$	0,005	0,006	0,006	0,007	0,009	0,010	0,013	0,017	0,025	0,050
$p_{Adjusted}$	0,000340	0,009	0,077	0,132	0,157	0,172	0,172	0,172	0,172	0,172
$H_0$	–	–	+	+	+	+	+	+	+	+

Таблица 5

Значения  $p$ , скорректированные значения  $\alpha$  и  $p$ , принятие или отклонение нулевых гипотез при применении метода Бенджамини – Хохберга

Параметр	Значение									
	0,000034	0,001	0,010	0,020	0,028	0,037	0,043	0,050	0,087	0,128
$p$	0,000034	0,001	0,010	0,020	0,028	0,037	0,043	0,050	0,087	0,128
$\alpha/m$	0,005	0,005	0,005	0,005	0,005	0,005	0,005	0,005	0,005	0,005
$i$	1	2	3	4	5	6	7	8	9	10
$\alpha_{B-H}$	0,005	0,010	0,015	0,020	0,025	0,030	0,035	0,040	0,045	0,050
$p_{Adjusted}$	0,000340	0,005	0,033	0,050	0,056	0,061	0,061	0,063	0,097	0,128
$H_0$	–	–	–	–	+	+	+	+	+	+

При использовании поправки Сидака (Шидака) для решения проблемы множественных сравнения в приведенном примере получены несколько меньшие значения скорректированных значений  $p$  (табл. 2). Однако число отклоняемых нулевых гипотез осталось тем же – отклоняются только две нулевые гипотезы.

Результаты принятия или отклонения нулевых гипотез с применением методов Холма – Бонферрони (табл. 3) и Холма – Сидака (Шидака) (табл. 4) позволяют получить те же самые результаты – отклоняются лишь две нулевые гипотезы. То есть только между двумя парами групп можно констатировать статистически значимые отличия.

Как было указано выше, метод Бенджамини – Хохберга менее консервативен, чем предыдущие. В связи с этим результаты его применения существенно отличаются (табл. 5). Так, по результатам применения данного метода отклоняются четыре нулевые гипотезы, т. е. исследователь может констатировать различия между четырьмя парами исследуемых групп. Такая консервативность на текущий момент является практически единственным решением при множественных многотысячных проверках статистических гипотез.

### Программная реализация методов коррекции значений $\alpha$ и $p$

Вне всяких сомнений, при большом числе проверяемых гипотез (сотня, тысяча или десятки тысяч), несмотря на простоту описанных методов, ручной расчет скорректированных значений  $\alpha$  и  $p$  для решения проблемы множественных сравнений весьма проблематичен. Однако существующие программные средства позволяют упростить, автоматизировать, этот процесс. Один из наиболее простых способов

это осуществить – использовать возможности языка программирования python. Для этого в любой среде программирования языка python необходимо подключить модули statsmodels и pandas (<https://www.statsmodels.org/stable/index.html>, <https://github.com/pandas-dev/pandas/releases/tag/v1.1.0>). На вопросах установки данных модулей останавливаться в данной статье не будем, но их подключение осуществляется путем написания следующих строк кода:

```
import statsmodels.api as statsmodels
import pandas
```

Затем задается список полученных в результате множественных сравнений значений  $p$ :

```
p_value = [0.000034, 0.001, 0.010, 0.020, 0.028, 0.037, 0.043, 0.050, 0.087, 0.128]
```

На следующем этапе осуществляется непосредственная коррекция значений  $p$  с использованием заданного метода (в коде – `method = <bonferroni>`) и проверка статистических гипотез в соответствии с заданным значением  $\alpha$  (в коде – `alpha = 0.05`):

```
adjusted_p_values = statsmodels.stats.multipletests(p_value, alpha = 0.05, method = 'bonferroni', is_sorted = False)
```

На заключительном этапе необходимо осуществить вывод результатов на экран с помощью следующей конструкции:

```
print(str(adjusted_p_values))
```

Полный код реализации коррекции значений  $p$  с применением описанных методов выглядит следующим образом:

```
import statsmodels.api as statsmodels
import pandas
p_value = [0.000034, 0.001, 0.010, 0.020, 0.028, 0.037, 0.043, 0.050, 0.087, 0.128]
adjusted_p_values = statsmodels.stats
```

```

multipletests(p_value, alpha=0.05, method =
'bonferroni', is_sorted = False)
print(str(adjusted_p_values))

```

Для изменения метода коррекции следует изменить значение параметра «method». В примере нашего кода он равен «bonferroni». В таком случае для коррекции будет использоваться поправка Бонферрони. Для применения поправки Сидака (Шидака) следует заменить значение параметра «method» на «sidak», для применения метода Холма – Бонферрони – на «holm», метода Холма – Сидака (Шидака) – на «holm-sidak», метода Бенджамини – Хохберга – на «fdr\_bh».

Рассмотрим выводимые результаты при применении различных методов решения проблемы множественных сравнений путем программной реализации на языке python. При применении поправки Бонферрони (method='bonferroni') выводимые результаты выглядят следующим образом:

```

(array([True, True, False, False, False, False,
False, False, False, False]), array([0.00034,
0.010, 0.100, 0.200, 0.280, 0.370, 0.430,
0.500, 0.870, 1.000]), 0.005116196891823743,
0.005)

```

Как видно из приведенного текста, результаты представляются в виде двух списков и двух значений. Первый список включает в себя информацию об отклонении нулевых гипотез при заданном значении  $\alpha$  (в коде – alpha = 0.05): True означает, что нулевая гипотеза отклоняется, а False – принимается.

Второй список включает в себя скорректированные значения  $p$  с применением заданного метода (в коде – method = «bonferroni»). Два значения, которые приводятся в конце, – скорректированные значения  $\alpha$  поправками Бонферрони и Сидака (Шидака). Они приводятся независимо от того, какой метод коррекции значений  $p$  выбран.

При применении поправки Сидака (Шидака) (method = «sidak») выводимые результаты выглядят следующим образом:

```

(array([True, True, False, False, False, False,
False, False, False, False]), array([0.000339,
0.009955, 0.095617, 0.182927, 0.247229,
0.314096, 0.355653, 0.401263, 0.597553,
0.745806]), 0.005116196891823743, 0.005)

```

При применении метода Холма – Бонферрони (method = «holm») выводимые результаты выглядят следующим образом:

```

(array([True, True, False, False, False, False,
False, False, False, False]), array([0.00034,
0.009, 0.080, 0.140, 0.168, 0.185, 0.185,
0.185, 0.185, 0.185]), 0.005116196891823743,
0.005)

```

При применении метода Холма – Сидака (Шидака) (method = «holm-sidak»):

```

(array([True, True, False, False, False, False,
False, False, False, False]), array([0.00033995,
0.00896408, 0.07725531, 0.13187447,
0.15666992, 0.17180723, 0.17180723,

```

```

0.17180723, 0.17180723, 0.17180723]),
0.005116196891823743, 0.005)

```

И в заключение приведем результаты решения проблемы множественных сравнений с применением метода Бенджамини – Хохберга (method = «fdr\_bh»):

```

(array([True, True, True, True, False, False,
False, False, False, False]), array([0.00034,
0.005, 0.03333333, 0.05, 0.056, 0.06142857,
0.06142857, 0.0625, 0.09666667, 0.128]),
0.005116196891823743, 0.005)

```

Как видно из представленных результатов программной реализации описанных в статье методов, приводимые скорректированные значения  $p$  и число отклоненных нулевых гипотез соответствуют полученным ручным применением данных методов.

Таким образом, рассмотренная нами проблема множественных сравнений может быть решена различными способами на разных этапах научного исследования. Учитывая широкую распространенность проблемы и серьезность последствий, связанных с обнаружением не существующих в реальности различий и связей в результате инфляции ошибки первого типа, сама проблема, равно как и способы её решения, должны рассматриваться при подготовке аспирантов и докторантов. Использование представленных путей решения проблемы позволит повысить уровень качества расчетов и представления результатов, снизить число найденных случайным образом различий и связей и, в конечном итоге, повысить уровень исследований и публикаций, а также уменьшить количество решений, принятых на основании ошибочных выводов.

#### Авторство

Наркевич А. Н. внес существенный вклад в концепцию и дизайн исследования, получение, анализ и интерпретацию данных, подготовил первый вариант статьи, окончательно утвердил присланную в редакцию рукопись; Виноградов К. А. внес существенный вклад в концепцию и дизайн исследования, получение, анализ и интерпретацию данных, существенно переработал первый вариант статьи, на предмет важного интеллектуального содержания, окончательно утвердил присланную в редакцию рукопись; Гржибовский А. М. внес существенный вклад в концепцию и дизайн исследования, получение, анализ и интерпретацию данных, существенно переработал первый вариант статьи на предмет важного интеллектуального содержания, окончательно утвердил присланную в редакцию рукопись.

Наркевич Артем Николаевич – ORCID 0000-0002-1489-5058; SPIN: 9030-1493

Виноградов Константин Анатольевич – ORCID 0000-0001-6224-5618; SPIN: 6924-0110

Гржибовский Андрей Мечиславович – ORCID 0000-0002-5464-0498; SPIN 5118-0081

#### Список литературы / References

1. Буненков Н. С., Буненкова Г. Ф., Комков В. В., Грищенко О. А., Немков А. С. SAS Enterprise Guide 6.1 для врачей: сравнение групп // Медицинский академический журнал. 2019. Т. 19, № 4. С. 33–40. DOI: 10/17816/MAJ17736

Bunenkov N. S., Bunenkova G. F., Komok V. V., Grinenko O. A., Nemkov A. S. SAS Enterprise Guide 6.1 for doctors: comparison of groups. *Meditsinskiy akademicheskij zhurnal* [Medical academic journal]. 2019, 19 (4), pp. 33-40. DOI: 10/17816/MAJ17736 [In Russian]

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

Grijbovski A. M. Analysis of three or more independent groups of quantitative data. *Ekologiya cheloveka* [Human Ecology]. 2008, 3, pp. 50-58. [In Russian]

3. Гржибовский А. М. Сравнение трех и более независимых групп с использованием непараметрического критерия Краскела – Уоллиса в программе Stata // Экология человека. 2014. № 6. С. 55–58.

Grijbovski A. M. Comparison of three or more independent groups using the nonparametric criterion Kruskal - Wallis in Stata software. *Ekologiya cheloveka* [Human Ecology]. 2014, 6, pp. 55-58. [In Russian]

4. Гржибовский А. М., Иванов С. В., Горбатова М. А. Сравнение количественных данных трех и более независимых выборок с использованием программного обеспечения Statistica и SPSS: параметрические и непараметрические критерии // Наука и Здравоохранение. 2016. № 4. С. 5–37.

Grijbovski A. M., Ivanov S. V., Gorbatova M. A. Comparison of quantitative data of three or more independent samples using software Statistica and SPSS: parametric and nonparametric criteria. *Nauka i Zdravookhranenie* [Science and Health]. 2016, 4, pp. 5-37. [In Russian]

5. Гржибовский А. М., Иванов С. В., Горбатова М. А. Сравнение количественных данных трех и более парных выборок с использованием программного обеспечения Statistica и SPSS: параметрические и непараметрические критерии // Наука и Здравоохранение. 2016. № 5. С. 5–29.

Grijbovski A. M., Ivanov S. V., Gorbatova M. A. Comparison of quantitative data of three or more paired samples using software Statistica and SPSS: parametric and nonparametric criteria. *Nauka i Zdravookhranenie* [Science and Health]. 2016, 5, pp. 5-29. [In Russian]

6. Наркевич А. Н., Виноградов К. А. Настольная книга автора медицинской диссертации: пособие. М.: Инфра-М, 2019. 454 с.

Narkevich A. N., Vinogradov K. A. *Nastol'naya kniga avtora meditsinskoy dissertatsii: posobie* [A reference book of the author of a medical dissertation: manual]. Moscow, Infra-M Publ., 2019, 454 p.

7. Benjamini Y., Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: series B (Methodological)*. 1995, 57 (1), pp. 289-300. DOI: 10.1111/j.2517-6161.1995.tb02031.x

8. Bonferroni C. E., Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*. 1936, 62 p.

9. Che R., Jack J. R., Motsinger-Reif A. A., Brown C. C. An adaptive permutation approach for genome-wide association study: evaluation and recommendations for use. *BioData Mining*. 2014, 7, pp. 9. DOI: 10.1186/1756-0381-7-9

10. Dunnett C. W. A multiple comparison procedure for comparing several treatments with a control. *Journal of the*

*American Statistical Association*. 1955, 50, pp. 1096-1121. DOI: 10.1080/01621459.1955.10501294

11. Foulkes A. C., Watson D. S., Griffiths C. E. M., Warren R. B., Huber W., Barnes M. R. Research Techniques Made Simple: Bioinformatics for Genome-Scale Biology. *Journal of Investigative Dermatology*. 2017, 137, pp. e163-e168. DOI: 10.1016/j.jid.2017.07.095

12. Gao X., Starmer J., Martin E. R. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genetic Epidemiology*. 2008, 32 (4), pp. 361-369. DOI: 10.1002/gepi.20310

13. Holland B. S., Copenhaver M. D. Improved Bonferroni-type multiple testing procedures. *Psychological Bulletin*. 1988, 104 (1), pp. 145-149. DOI: 10.1037//0033-2909.104.1.145

14. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*. 1979, 6 (2), pp. 65-70.

15. Johnson R. C., Nelson G. W., Troyer J. L., Lautenberger J. A., Kessing B. D. Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics*. 2010, 11, pp. 724. DOI: 10.1186/1471-2164-11-724

16. Keuls M. The use of the «studentized range» in connection with an analysis of variance. *Euphytica*. 1952, 1 (2), pp. 112-122. DOI: 10.1007/bf01908269

17. Moran M. Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos*. 2003, 100 (2), pp. 403-405. DOI: 10.1034/j.1600-0706.2003.12010.x

18. Newman D. The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. *Biometrika*. 1939, 31 (1), pp. 20-30. DOI: 10.1093/biomet/31.1-2.20

19. Rothman K. J. No Adjustments Are Needed for Multiple Comparisons. *Epidemiology*. 1990, 1 (1), pp. 43-46. DOI: 10.1097/00001648-199001000-00010

20. Seidler J., Vondráček J. Í., Saxl I. The life and work of Zbyněk Šidák (1933-1999). *Applications of Mathematics*. 2000, 45 (5), pp. 321. DOI: 10.1023/A:1022238410461. hdl:10338.dmlcz/134443

21. Shaffer J. P. Multiple Hypothesis Testing. *Annual Review of Psychology*. 1995, 46 (1), pp. 561-584. DOI: 10.1146/annurev.ps.46.020195.003021

22. Šidák Z. K. Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *Journal of the American Statistical Association*. 1967, 62 (318), pp. 626-633. DOI: 10.1080/01621459.1967.10482935

#### Контактная информация:

Наркевич Артём Николаевич — кандидат медицинских наук, доцент, декан медико-психолого-фармацевтического факультета, заведующий кафедрой медицинской кибернетики и информатики, заведующий лабораторией медицинской кибернетики и управления в здравоохранении ФГБОУ ВО «Красноярский государственный медицинский университет им. проф. В. Ф. Войно-Ясенецкого» Минздрава России  
Адрес: 660022, г. Красноярск, ул. Партизана Железняка, д. 1

E-mail: narkevichart@gmail.com