

## УПРАВЛЕНИЕ ДИЗАЙНОМ БИМЕДИЦИНСКОГО ИССЛЕДОВАНИЯ НА ОСНОВЕ ПЕРЕГРУППИРОВКИ ПАР ИССЛЕДУЕМЫХ ФАКТОРОВ<sup>1</sup>

**Туровский Я.А.**

*Воронежский государственный университет,  
Россия, г. Воронеж Университетская пл. д.1  
Институт проблем управления им. В.А. Трапезникова РАН,  
Россия, г. Москва ул. Профсоюзная д.65  
yaroslav\_turovsk@mail.ru*

**Борзунов С.В., Вахтин А.А.**

*Воронежский государственный университет,  
Россия, г. Воронеж Университетская пл. д.1  
sborzunov@gmail.com, alvahtin@gmail.com*

*Аннотация: В работе представлена модификация алгоритма статистического оценивания результатов попарных сравнений факторов на основе группировки результатов тестов, использующая перегруппировку пар факторов. Предложенный подход в управлении может быть полезен в задачах биомедицинской природы, содержащих большое количество факторов.*

Ключевые слова: множественные сравнения, оценка значимости, задача кластеризации.

### Введение

Одной из серьёзных проблем, возникающих при планировании и обработке результатов медицинских и биологических исследований является учёт эффекта множественных сравнений, который может существенно повлиять на оценку полученных результатов. В свою очередь для устранения и/или снижения влияния этого эффекта предложен целый ряд алгоритмов и подходов, которые, в свою очередь, как, например, поправка Бонферрони, могут также оказать существенное влияние на интерпретацию полученных результатов [1-3]. Таким образом, представляет несомненный интерес направление работ, связанных с совершенствованием методов управления дизайном биомедицинских исследований в задачах минимизации влияния эффекта множественных сравнений.

В ранних работах [4-5] авторами предложен алгоритм оценки распределения положительных решений о принятии альтернативной гипотезы в пространстве факторов исследования. Иными словами, наличие ложно-положительных решений должно быть случайным, тем временем, как группировка положительных решений в определённых участках пространства факторов исследования свидетельствует о неслучайности появления таких результатов. Развивая подобный подход можно предложить классифицировать факторы пространства эксперимента на основе близости результатов, полученных с их участием. Суть подхода заключается в перемещении результатов, связанных с какими либо значениями исследуемого фактора в попытке получить новую группу статистически-значимых положительных результатов. При этом порядок значений на оси фактора естественно может меняться, что и служит основой классификации (см. рис. 1, на котором представлен результат перестановки строк и столбцов для матрицы, содержащей 13 положительных элементов).

---

<sup>1</sup> Работа выполнена при финансовой поддержке РФФИ, проект № 19-29-01156 мк

$i \setminus j$	$j_1$	$j_2$	$j_3$	$j_4$	$j_5$	$j_6$
$i_1$		+	+	+		+
$i_2$						
$i_3$		+		+		+
$i_4$						+
$i_5$		+	+	+		+
$i_6$						+



$(i_5 \leftrightarrow i_6), (i_2 \leftrightarrow i_6), (j_1 \leftrightarrow j_6), (j_3 \leftrightarrow j_4)$

$i \setminus j$	$j_1$	$j_2$	$j_3$	$j_4$	$j_5$	$j_6$
$i_1$	+	+	+	+		
$i_2$	+	+	+	+		
$i_3$	+	+	+			
$i_4$	+					
$i_5$	+					
$i_6$						

Рис. 1. Пример перестановки строк и столбцов

## 1 Перегруппировка пар факторов как основа алгоритма статистического оценивания

Из множества всевозможных результатов эксперимента сформируем подмножество  $Q$  факторов, достоверно демонстрирующих значимую попарную зависимость между ними. Такое подмножество универсального множества  $U$  (содержащего результаты полного набора опытов проведенного эксперимента) будем называть «кластером» испытаний. Его элементы и мощность (т.е. размер множества) подлежат определению. Для управления дизайном исследования наибольший интерес представляет ситуация, при которой число выявленных различий (при некотором фиксированном значении  $p$ ) без учёта эффекта множественных сравнений сравнимо с числом, к которому стремится их количество при только ложноположительных значениях попарного статистического критерия. В качестве начального шага предлагаемого алгоритма вычислим вероятность положительного исхода одиночного опыта  $p_+$  в предположении равномерного распределения положительных исходов в множестве согласно  $p_+ = s/|U|$ , где  $s$  — количество выявленных в эксперименте положительных исходов.

Предлагается следующий алгоритм для выявления подмножества факторов, продемонстрировавших значимые различия. Ранжируем строки и столбцы по числу достоверных различий, т.е. выполняем сортировку по убыванию числа различий. Далее записываем данные в булевскую матрицу  $M$  так, чтобы слева направо, а также сверху вниз, число различий в столбцах уменьшалось. Формируем ограничивающий прямоугольник  $P$  с одной из вершин в левом верхнем углу, за пределами прямоугольника достоверных различий нет. Далее применяем критерий  $\chi^2$  для определения факта значимости серии различий в опытах, попавших в ограничивающий прямоугольник  $P$ . В случае отсутствия достоверной значимости уменьшаем прямоугольник на один столбец и на одну строку, и повторно применяем критерий  $\chi^2$  до тех пор пока не получим статистически значимые доказательства концентрации отклонений  $H_0$ . В итоге, кластер испытаний составляют факторы, попадающие в прямоугольник  $P$  на заключительном шаге алгоритма.

Блок-схема алгоритма приведена на рис. 2.

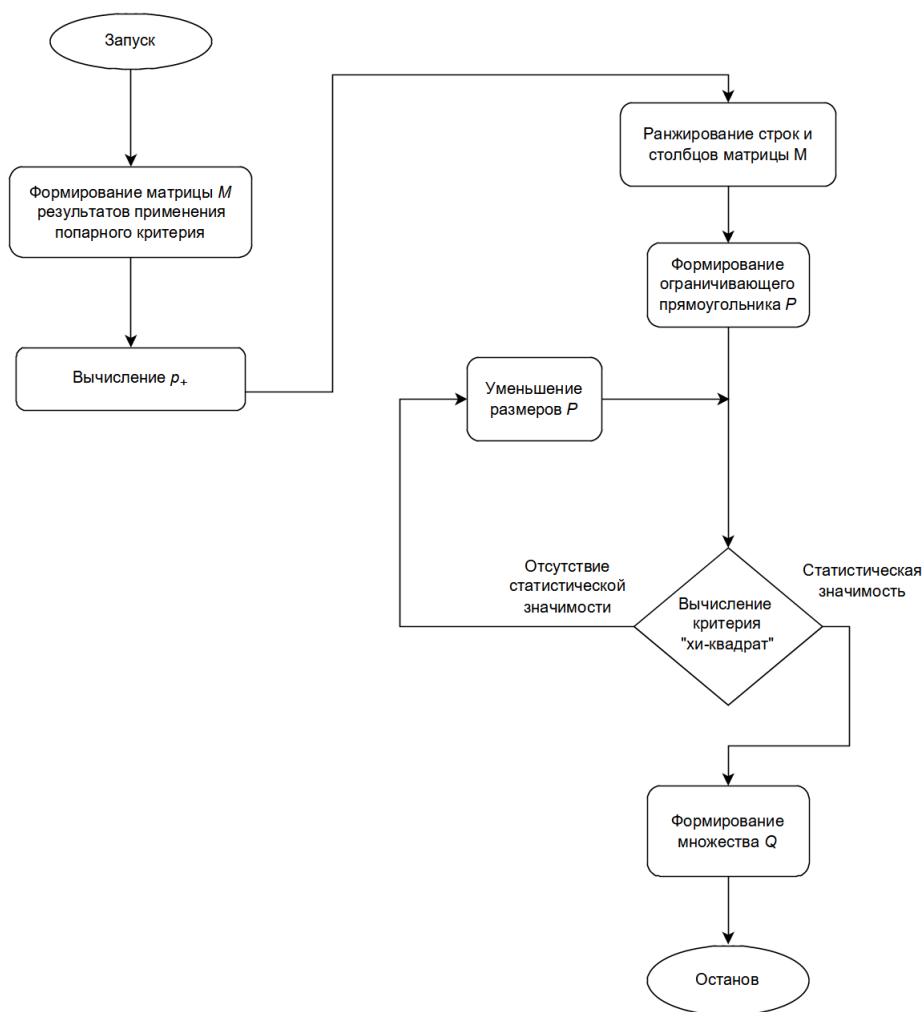


Рис. 2. Блок-схема алгоритма статистического оценивания на основе группировки результатов тестов

Работу представленного алгоритма продемонстрируем на модельном примере, предположив, что исследуется ряд факторов  $a_1, a_2, a_3, a_4, b_1, b_2$ , показавших по итогам серии экспериментов следующие результаты попарных сравнений (см. табл. 1, где представлены элементы матрицы  $M$ ). Серым цветом выделен ограничивающий прямоугольник  $P$  на первой итерации.

Таблица 1. Пример результатов применения попарного критерия – уровни значимости сравнений

Варианты постановок и серии опытов	Фактор $a_1$ – фактор $b_1$	Фактор $a_2$ – фактор $b_1$	Фактор $a_3$ – фактор $b_1$	Фактор $a_4$ – фактор $b_1$	Фактор $a_1$ – фактор $b_2$	Фактор $a_2$ – фактор $b_2$	Фактор $a_3$ – фактор $b_2$	Фактор $a_4$ – фактор $b_2$
1	+	+	+	+	+			
2	+	+	+	+	+		+	
3	+	+	+					
4	+					+		
5								
6								

Из табл. 1 видно, что ограничивающий прямоугольник  $P$  содержит не все попарные комбинации факторов:  $\{(a_1, b_1), (a_2, b_1), (a_3, b_1), (a_4, b_1), (a_1, b_2), (a_2, b_2), (a_3, b_2)\}$ . На первую итерацию алгоритма подаются такие данные:  $4 \times 7 = 28$  результатов опыта в ограничивающем прямоугольнике, из них различия достоверны в  $4+3+3+2+2+1+1 = 16$  результатах. Плотность различий равна  $16/28 = 0,571\dots$

После уменьшения размера  $P$ , как показано в табл. 2, плотность различий составит  $(3+3+3+2+2)/15 = 0,866\dots$ , что согласно статистическому критерию «хи-квадрат» следует признать значимым. Итак, значимые различия сформировали кластер  $Q = \{(a_1, a_2, a_3, a_4, b_1)\}$ .

Таблица 2. Ограничивающий прямоугольник на заключительном шаге работы алгоритма

Фактор $a_1$ – фактор $b_1$	Фактор $a_2$ – фактор $b_1$	Фактор $a_3$ – фактор $b_1$	Фактор $a_4$ – фактор $b_1$	Фактор $a_1$ – фактор $b_2$
+	+	+	+	+
+	+	+	+	+
+	+	+		

## Заключение

В работе представлено развитие подхода на основе разработанной авторами методики принятия решения об отнесении альтернативной статистической гипотезы к истинно- или ложноположительным. Используя оценку распределения положительных исходов применения статистических критериев осуществлён переход к классификации значений составляющих пространство для каждого фактора. В работе представлен принципиальный подход и разработанные алгоритмы, продемонстрирован на модельных примерах вариант работы алгоритма, обеспечивающий изменение порядка следования значений факторов, обеспечивающих их группировку в статистически-значимые кластеры. Таким образом, осуществлена задача классификации значений факторов (переменных), на основе схожести структуры решений о принятии или отклонении статистических гипотез. Предложенный подход в управлении может быть полезен в задачах биомедицинской природы, содержащих большое количество факторов и возможных значений этих факторов.

## Литература

1. Рунион Р. Справочник по непараметрической статистике: Современный подход. – Москва: Финансы и статистика, 1982. – 198 с.
2. Hochberg Y., Tamhane A.C. Multiple comparison procedures. — New York: Wiley, 1987. – 482 p.
3. Ludbrook J. Multiple comparison procedures updated // Clinical and Experimental Pharmacology and Physiology. V. 25. 1998. – P. 1032-1037.
4. Туровский Я.А. Борзунов С.В., Вахтин А.А. Алгоритм оценки результатов статистического анализа данных биомедицинской природы в условиях эффекта множественных сравнений // Программная инженерия. Т. 12. 2021, № 9. – С. 470-474. – DOI:10.17587/prin.12.470-474.
5. Туровский Я.А. Борзунов С.В., Вахтин А.А. Алгоритм коррекции статистического оценивания с учетом эффекта множественных сравнений на основе группировки результатов тестов // Программная инженерия. Т. 13. 2022, № 3. – С. 148-152. – DOI:10.17587/prin.13.148-152.