

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ НЕПАРАМЕТРИЧЕСКОГО АНАЛИЗА МНОГОФАКТОРНЫХ БИОЛОГИЧЕСКИХ КЛАССИФИКАЦИЙ

Т. Г. ОГАНЯН

Предлагаются методы проверки существенности взаимодействий и эффектов уровней качественных факторов, которые позволяют получать дополнительную информацию и повышают эффективность непараметрического анализа. С помощью предложенных методов решается задача по определению факторов, имеющих существенное значение при титровании вирусов гриппа А2 и В.

Ключевые слова: непараметрические методы, вирусы гриппа.

По сравнению с традиционными параметрическими методами методы непараметрической статистики более универсальны. Есть все основания утверждать, что применение этих методов в биологии и медицине более оправдано, чем традиционных. Основными причинами, по которым исследователи обращаются к ним, являются ненормальное распределение признака и то, что измерения проводятся не в интервальной шкале. Этим объясняется широкое распространение непараметрических методов в медико-биологических исследованиях в последние годы [4, 6, 7, 9].

Рассмотрим предложенную в работе Лисенкова и др. [5] непараметрическую процедуру комплексного анализа ортогональных многоуровневых планов с количественными факторами, которая является дальнейшим развитием методов, использующих предварительное «выравнивание» наблюдений. Эта процедура была разработана для анализа главных эффектов в многоуровневых и неполных классификациях типа факторных планов и их дробных реплик (например, латинских квадратов). Процедура комплексного анализа, использующая «выравнивание» наблюдений, более эффективна, чем критерии Фридмана и Дарбина [8], основанные на независимом ранжировании в каждом блоке, так как здесь кроме сравнений внутри блоков используется информация, извлеченная из межблоковых сравнений.

При этом анализ проводится в несколько этапов. Вначале определяются оценки эффектов уровней каждого фактора:

$$b_i^j = \frac{u-1}{N} \frac{\sum_{u=1}^N y_u x_{iu}^j}{\sum_{u=1}^N x_{iu}^j} - b_0, \quad (1)$$

где b_0 — среднее значение выходного показателя,

$$j = 1, 2, \dots, p_i,$$

$\sum_{u=1}^N y_u$ — сумма откликов, при которых i -тый фактор был на j -том уровне,

$$x_{iu}^j = \begin{cases} 1, & \text{если } j\text{-тый уровень } i\text{-того фактора встречается в } u\text{-том опыте,} \\ 0, & \text{в других случаях.} \end{cases}$$

Далее вычисляются «остатки»:

$$d_u = y_u - \sum_{i=1}^n \sum_{j=1}^{p_i} b_i^{j(u)} x_{iu}^j,$$

и затем для каждого фактора находится значение выходного показателя, скорректированного на эффекты остальных факторов:

$$y_{i(u)}^* = d_u + b_i^{j(u)} x_{iu}^j.$$

Заключительный этап сводится к ранжированию скорректированных значений выходного показателя в едином ряду в порядке возрастания. Затем значение $y_{i(u)}^*$, стоящего в ранжированном ряду на R -том месте, заменяется некоторой меткой $a_N(R_i)$, которая является нормальной или приближенной [6]. Далее используется статистика типа H , как в случае анализа однофакторных классификаций [1, 6]:

$$H_i = \frac{(N-1)}{N} \frac{\sum_{j=1}^{p_i} \frac{1}{n_j} \sum_{R \in S_j} a_N^2(R_i)}{\sum_{i=1}^n a_N^2(R_i)},$$

где S_j — множество индексов, соответствующих наблюдениям для j -того уровня.

Нулевая гипотеза о незначимости i -того фактора отвергается в пользу односторонней альтернативы о наличии сдвига (влияние фактора признается существенным), если расчетное значение H_i больше критического значения χ^2 — распределения с числом степеней свободы $p_i - 1$ для заданного уровня значимости.

Для факторов, признанных существенными после проверки по H -критерию, предлагается проводить более детальный анализ с помощью свободных от распределения L_k -статистик [7]:

$$L_k = \sum_{j=1}^{p_i} W_j^{ks} a_N(R_j) / n_j,$$

где $a_N(R_j) = \sum_{s=1}^{n_j} a_N(R_{j,s})$ — сумма меток при j -том уровне фактора,

n_j — число наблюдений для j -того уровня фактора,

$$N = \sum_{j=1}^{p_i} n_j.$$

Статистика L_k имеет асимптотически нормальное распределение. При анализе количественных факторов задаемся $W_j^{ks} = \Theta_{kj}$, где Θ_{kj} — являются ортогональными полиномами k -того порядка [7]:

$$\Theta_{kj} = \alpha + \beta x_1 + \gamma x_j^2 + \dots + \eta x_j^k.$$

При этом имеем

$$M \{L_k\} = 0,$$

$$D \{L_k\} = \frac{p_i}{N(N-1)} \sum_{j=1}^{p_i} \Theta_{kj}^2 \sum_{v=1}^{n_j} a_N^2(R_{jv}).$$

При анализе качественных факторов задаемся $W_j^{ks} = T_j^s$, где

$$T_j^s = \begin{cases} 1, & \text{если рассматривается } s\text{-тый уровень фактора,} \\ 0, & \text{в других случаях.} \end{cases}$$

Статистика L_k^* в этом случае имеет следующие параметры распределения:

$$M \{L_k^*\} = 0,$$

$$D \{L_k^*\} = \frac{(N - n_j)}{N n_j (N - 1)} \sum_{j=1}^{p_i} \sum_{v=1}^{n_j} a_N^2(R_{jv}).$$

Проверка гипотез сводится к вычислению значений

$$z_k = L_k / [D \{L_k\}]^{1/2}$$

и сравнению их с критическим значением стандартного нормального распределения для выбранного уровня значимости.

Использование L_k^* -статистики позволяет для количественных факторов проверять наличие линейных, квадратичных и т. д., составляющих (тренда). Статистики L_k^* позволяют проверять гипотезы о существенности уровней факторов, как это принято в параметрической теории.

Помимо проверки существенности главных эффектов, предлагается проверять значимость эффектов парных взаимодействий факторов. Для этого после вычисления «остатков» d_u находятся значения отклика, скорректированного на эффекты остальных факторов следующим образом:

$$y_{ir}^*(u) = d_u + b_i^{j(u)} x_{iu}^j + b_r^{s(u)} x_{iu}^s,$$

где $b_i^{j(u)}$ — эффект j -того уровня i -того фактора.

$b_r^{s(u)}$ — эффект s -того уровня r -того фактора.

$b_i^{j(u)}$ и $b_r^{s(u)}$ — определяются из выражения (1),

u — опыты, в которых встречаются j -тый уровень i -того фактора с s -тым уровнем r -того фактора.

Заключительный этап сводится к ранжированию значений скорректированного таким образом выходного показателя и применению H -критерия. В случае существенности эффекта взаимодействия для про-

верки сочетаний уровней i -того и $г$ -того факторов используется средний ранг суммы столбца взаимодействий j -того уровня i -того фактора с s -тым уровнем $г$ -того фактора. Общий вид такой статистики следующий:

$$a_N(\bar{R}_{js.}) = \sum_{v=1}^{n_{js}} a_N(R_v) / n_{js},$$

где n_{js} — число элементов столбца взаимодействий.

Проверка гипотезы существенности сочетаний уровней двух факторов сводится к вычислению величины

$$z_{js} = \frac{a_N(\bar{R}_{js.}) - M\{a_N(\bar{R}_{js.})\}}{[D\{a_N(\bar{R}_{js.})\}]^{1/2}}$$

и сравнению ее с критическим значением стандартного нормального распределения Z_q .

Рассмотрим применение разработанных методов при определении возможности титрования вирусов гриппа В и А2 с использованием эритроцитов, обработанных формалином и акроленом [2, 3].

При этом был реализован факторный эксперимент $2 \times 3 \times 3 \times 3$ из 54 опытов с 4-мя параллельными титрованиями. Выходным показателем являлся титр вируса в реакции прямой гемагглютинации. Изучалось влияние факторов, обозначенных в табл. 1, 2 следующим образом: X_1 —

Таблица 1
Результаты эксперимента $2 \times 3 \times 3 \times 3$ по титрованию вирусов гриппа А2 и В

$X_1 - X_2$	1-1	1-2	1-3	2-1	2-2	2-3
1-1	7,50	6,50	8,00	8,00	7,75	8,75
1-2	8,25	8,50	7,00	8,50	11,00	9,00
1-3	7,75	6,50	8,25	8,00	8,00	9,20
2-1	7,25	7,00	7,50	7,75	7,25	8,00
2-2	8,25	10,50	7,25	8,75	10,25	8,25
2-3	7,00	6,00	8,75	7,00	8,00	8,70
3-1	7,00	6,25	7,00	7,25	6,75	8,00
3-2	8,00	7,25	6,75	8,50	10,50	8,25
3-3	6,50	6,00	7,00	7,00	7,75	8,50

тип вируса гриппа на 2-х уровнях: А2-1, В-2, X_2 —способ стабилизации эритроцитов на 3-х уровнях: нативные (НЭ)—1, формализированные (ФЭ)—2, акроленнизированные (АЭ)—3, X_3 —концентрация эритроцитов на 3-х уровнях: 0,75%—1, 1,00%—2, 1,25%—3, X_4 —происхождение эритроцитов на 3-х уровнях: человеческие—1, куриные—2, морской свинки—3.

Проверка условия нормальности распределения результирующего признака в этом эксперименте проводилась с помощью W -критерия.

Причем для ряда ячеек рассматриваемого плана нормальность распределения отклика не подтвердилась.

В этих условиях применение параметрических методов может исказить результаты анализа, поэтому проводился непараметрический анализ с помощью разработанных методов (табл. 2).

Таблица 2

Результаты непараметрического анализа эксперимента $2 \times 3 \times 3 \times 3$

X_1		X_2		X_3		X_4	
уровень	z_1	уровень	z_2	уровень	z_3	уровень	z_4
A2	-3,72	HЭ	-2,11	0,75%	3,06	куриные*	5,04
B	3,72	ФЭ	0	1,00%	0	морской свинки	-2,52
—	—	AЭ	2,11	1,25%	-3,06	человека	-2,52

$X_1 X_2$		$X_2 X_4$	
сочетание уровней	z_{12}	сочетание уровней	z_{24}
A2 — HЭ	2,41	ФЭ — куриные	5,02
A2 — ФЭ	-2,17	ФЭ — морской свинки	-3,40
B — HЭ	2,41	ФЭ — человека	-3,52
B — ФЭ	-2,17	AЭ — куриные	2,25
—	—	AЭ — человека	3,11

В результате анализа установлено, что наиболее существенное значение имеет тип вируса гриппа. Установлена определенная динамика в значениях титров при применении различных концентраций эритроцитов—с ее увеличением титры вирусов снижаются. Оптимальной является концентрация 0,75%. Использование куриных эритроцитов также повышает чувствительность титрования.

Из 6-ти парных взаимодействий значимыми оказались только два: происхождение эритроцитов—стабилизация, тип вируса—стабилизация.

Анализ первого взаимодействия подтверждает логическое заключение, что модифицированные рецепторы стабилизированных эритроцитов имеют различную чувствительность в реакции гемагглютинации с различными типами вирусов. Влияние второго парного взаимодействия показывает, что эритроциты человека, курицы и морской свинки по-разному реагируют на стабилизацию. Это проявляется в различном изменении величины титров обоих антигенов при использовании всех видов стабилизированных эритроцитов в сравнении с нативными.

Незначимые двойные взаимодействия связаны с влиянием концентрации эритроцитов. Из этого следует, что все виды эритроцитов, разные методы стабилизации и все типы вирусов при изменении концен-

трации эритроцитов дают одинаковые равнозначные изменения титров вирусов гриппа.

Институт эпидемиологии, вирусологии
и медицинской паразитологии МЗ АрмССР
им. А. Б. Алексаняна

Поступило 11.I 1980 г.

**ԲԱԶՄԱԳՈՐԾՈՆ ԿԵՆՍԱՐԱՆԱԿԱՆ ԴԱՍԱԿԱՐԳՈՒՄՆԵՐԻ
ՈՂ-ՊԱՐԱՄԵՏՐԻԿ ԷՖԵԿՏԻՎՈՒԹՅԱՆ ԲԱՐՁՐԱՑՈՒՄԸ**

Տ. Գ. ՕՂԱՆՅԱՆ

Առաջարկվում են կենսաբանական դասակարգումների ոչ-պարամետրիկ անալիզի մեթոդներ, որոնք ունեն բարձր էֆեկտիվություն և հնարավորություն են տալիս ստանալ լրացուցիչ ինֆորմացիա:

Դիտարկված մեթոդների օգնությամբ լուծվում է էրիտրոցիտներով կայունացված գրիպի վիրուսների տիրաբացիայի զգայունության բաղադրանքային խնդիրը:

**INCREASE OF THE EFFECTIVENESS OF NON-PARAMETRIC
ANALYSIS FOR MULTIFACTOROUS BIOLOGICAL CLASSIFICATIONS**

T. G. OGANIAN

Methods of verification of interaction validity and qualitative factor level effects which allow to obtain additional information and increase the effectiveness of non-parametric analysis are proposed. The problems of determination of factors which play an important role for titration of influenza A₂ and B viruses by the use of proposed methods have been solved.

Л И Т Е Р А Т У Р А

1. Гаек Я., Шидак Э. Теория ранговых критериев. М., 1971.
2. Грабовский П. М., Блюмин С. Л., Козловская Г. И., Кодкин Г. Х., Оганян Т. Г. Сб.: *Вопр. медицины. вирусологии.* 534, М., 1975.
3. Грабовский П. М., Лисенков А. Н., Кодкин Г. Х., Оганян Т. Г., Козловская Г. И., Таранцева Л. П. Сб.: *Вопросы медицины. вирусологии.* 535, М., 1975.
4. Гублер Е. В., Генкин А. А. *Применение параметрических критериев статистики в медико-биологических исследованиях.* Л., 1973.
5. Лисенков А. Н., Кодкин Г. Х., Оганян Т. Г. Сб.: *Статистические методы планирования и анализа медико-биологических экспериментов.* 13, Киев, 1975.
6. Лисенков А. Н., Никитина Е. П., Оганян Т. Г. *Непараметрические методы анализа многофакторных экспериментов.* М., 1976.
7. Оганян Т. Г., Кодкин Г. Х., *Биолог. ж. Армении.* 32, 11, 1979, 1135.
8. Hodges J L., Lehman E. L. *Ann. Math. Stat.* 33, 482, 1962,
9. Lienert G. A. *Verteilungsfreie Methoden in der Biostatistik.* Meisenheim, 1975.