

АНАЛИЗ МЕДИКО-БИОЛОГИЧЕСКИХ ЭКСПЕРИМЕНТОВ
С НЕПОЛНЫМ ЧИСЛОМ НАБЛЮДЕНИЙ

Т. Г. ОГАНЯН

Предлагается метод, позволяющий проводить анализ результатов многофакторных медико-биологических экспериментов, когда имеется неполное число наблюдений, т. е. часть данных потеряна или пропущена.

С помощью предложенного метода решается задача по изучению влияния комплекса факторов на чувствительность методики титрования.

Ключевые слова: многофакторный эксперимент, дисперсионный анализ, стабилизированные эритроциты, титрование.

При постановке многофакторного эксперимента следует заботиться об обеспечении равного числа наблюдений в каждом опыте, но даже при большой осторожности бывают случаи потери наблюдений. Особенно трудно иметь равное число наблюдений в каждом опыте при постановке медико-биологического эксперимента. Причины этого могут быть самыми различными; выход из строя экспериментальной установки, гибель части животных, нарушение нормального хода реакции и т. д. При этом поскольку дисперсионный анализ нелегко приспособить к неполным множествам данных (неортогональным комплексам), математическая обработка результатов экспериментов сильно затрудняется или же становится невозможной [4].

В настоящей работе получены результаты, позволяющие применять для анализа многофакторных экспериментов при возникновении нарушения условия ортогональности упрощенную процедуру вместо сложного анализа в случаях, когда потеряно не более 20% опытных данных. Неполноту отдельных наблюдений удается компенсировать, опираясь на учет внутренней связи между несколькими варьирующими признаками. Метод заключается во введении оценок пропущенных наблюдений, которые подбираются таким образом, чтобы минимизировать остаточную дисперсию или дисперсию «внутри ячеек», и затем в завершении анализа, как и в случае полного набора данных.

Следует подчеркнуть, что замена пропущенных наблюдений их оценками никоим образом не восполняет информацию, потерянную из-за недостатка данных, но дает возможность легкими методами вычислений, разработанные для анализа ортогональных комплексов, применять в тех случаях, когда имеются неполные данные. При подсчете соответствующей

ших степеней свободы полученные оценки не включаются. Рассмотрим модель следующего вида:

$$y_{ij\dots} = \mu + \alpha_i + \beta_j + \dots + \tau_\nu + \varepsilon_{ij\dots}, \quad (1)$$

где $y_{ij\dots}$ — экспериментальный результат, полученный с i -тым уровнем I-го фактора, j -тым уровнем II-го фактора, ..., ν -тым уровнем n -го фактора; μ — общий эффект во всех опытах (истинное среднее совокупности, из которой получена выборка); α_i — эффект, обусловленный i -тым уровнем I-го фактора; β_j — эффект, обусловленный j -тым уровнем II-го фактора и т. д.; $\varepsilon_{ij\dots}$ — ошибка результатов, обусловленная действием случайных причин, имеющая распределение $N(0, \sigma^2)$.

При нахождении оценок пропущенных наблюдений за основу берется модель (1), т. е. делается предположение, что эффекты взаимодействия факторов незначимы и поэтому можно ограничиться только линейной моделью. Основное уравнение дисперсионного анализа для сумм квадратов в случае n факторов запишется следующим образом [3]:

$$SS_0 = SS - \sum_{\tau=1}^n SS_\tau = \sum_i^{k_1} \sum_j^{k_2} \dots \sum_\nu^{k_n} y_{ij\dots}^2 - \sum_i^{k_1} T_{1\dots}^2 / \Pi_1 - \sum_j^{k_2} T_{j\dots}^2 / \Pi_2 - \dots - \sum_\nu^{k_n} T_{\nu\dots}^2 / \Pi_n + (n-1) T_{\text{общ}}^2 / N, \quad (2)$$

здесь k_τ ($\tau = \overline{1, n}$) — число уровней I-го, II-го, ..., n -го факторов $T_{1\dots}$, $T_{j\dots}$ и т. д. — суммы вариантов по градациям комплекса (для удобства в дальнейшем будем обозначать их просто T_1, T_2 и т. д.); SS_0 — остаточная сумма квадратов; SS — общая сумма квадратов отклонений всех результатов от их общего среднего; SS_1, SS_2, \dots, SS_n — суммы квадратов, характеризующие отклонения k_τ средних для независимых групп от общего среднего;

$$\Pi_1 = k_2 k_3 \dots k_n,$$

$$\Pi_2 = k_1 k_3 \dots k_n,$$

$$\vdots$$

$$\Pi_n = k_1 k_2 \dots k_{n-1}.$$

Вначале рассмотрим случай, когда имеется одно пропущенное наблюдение (обозначим пропущенное значение z_1). Найдем наилучшее значение z_1 , т. е. такое, которое минимизирует выражение (2). Для этого возьмем из выражения (2) производную по z_1 и приравняем ее нулю. Учитывая, что производные постоянных равны нулю, можем записать:

$$\frac{d(SS_0)}{dz_1} = 2z_1 - \frac{2T_1(z_1)}{\Pi_1} - \frac{2T_2(z_1)}{\Pi_2} - \dots - \frac{2T_n(z_1)}{\Pi_n} + \frac{2(n-1)T_{\text{общ}}(z_1)}{N} = 0, \quad (3)$$

где $T_1(z_1)$, $T_2(z_2)$ и т. д.—суммы вариант, в которые входит z_1 .

Запишем выражение (3) в виде:

$$\frac{d(SS_0)}{dz_1} = 2z_1 - \frac{2[z_1 + T'_1(z_1)]}{\Pi_1} - \frac{2[z_1 + T'_2(z_1)]}{\Pi_2} - \dots - \frac{2[z_1 + T'_n(z_1)]}{\Pi_n} + \frac{2(n-1)[z_1 + T'_{общ}(z_1)]}{N} = 0, \quad (4)$$

где $T'_i(z_1)$ —суммы вариант без z_1 .

Приведем (4) к общему знаменателю, подставим значения Π_1 , Π_2 , ..., Π_n и после преобразований будем иметь:

$$(N + n - 1 - \sum_{\tau=1}^n k_{\tau}) z_1 - \sum_{\tau=1}^n k_{\tau} T'_{\tau}(z_1) + (n - 1) T'_{общ}(z_1) = 0.$$

Введем обозначения

$$A_{11} = N + n - 1 - \sum_{\tau=1}^n k_{\tau},$$

$$C_1 = \sum_{\tau=1}^n k_{\tau} T'_{\tau}(z_1) - (n - 1) T'_{общ}(z_1).$$

Оценка пропущенного наблюдения определится следующим образом:

$$z_1 = \frac{C_1}{A_{11}}.$$

Рассмотрим случай, когда имеется два пропущенных наблюдения (z_1 и z_2). Для этого случая необходимо взять частные производные по каждой из пропущенных величин и приравнять нулю. Таким образом, получаем два уравнения:

$$\frac{d(SS_0)}{dz_1} = 2z_1 - \frac{2T_1(z_1)}{\Pi_1} - \frac{2T_2(z_1)}{\Pi_2} - \dots - \frac{2T_n(z_1)}{\Pi_n} + \frac{2(n-1)T_{общ}(z_1, z_2)}{N} = 0,$$

$$\frac{d(SS_0)}{dz_2} = 2z_2 - \frac{2T_1(z_2)}{\Pi_1} - \frac{2T_2(z_2)}{\Pi_2} - \dots - \frac{2T_n(z_2)}{\Pi_n} + \frac{2(n-1)T_{общ}(z_1, z_2)}{N} = 0.$$

После ряда преобразований и подстановки значений Π_{τ} получим

$$\left. \begin{aligned} (N + n - 1 - \sum_{\tau=1}^n k_{\tau}) z_1 + (n - 1 - \sum_{l \in M_{12}} k_l) z_2 &= \sum_{\tau=1}^n k_{\tau} T'_{\tau}(z_1) - (n - 1) T'_{общ}, \\ (n - 1 - \sum_{l \in M_{12}} k_l) z_1 + (N + n - \sum_{\tau=1}^n k_{\tau}) z_2 &= \sum_{\tau=1}^n k_{\tau} T'_{\tau}(z_2) - (n - 1) T'_{общ}, \end{aligned} \right\} (5)$$

здесь M_{12} —множество, состоящее из совпадающих индексов пропущенных наблюдений z_1 и z_2 .

$$A_{12} = A_{21} = n - 1 - \sum_{l \in M_{12}} k_l,$$

$$C_2 = \sum_{\tau=1}^n k_{\tau} T'_{\tau}(z_2) - (n-1) T'_{\text{общ}},$$

и запишем систему уравнений (5) в виде

$$\left. \begin{aligned} A_{11}z_1 + A_{12}z_2 &= C_1, \\ A_{21}z_1 + A_{22}z_2 + C_2 &= 0 \end{aligned} \right\} \quad (6)$$

здесь

$$A_{11} = A_{22}.$$

Таким образом, пропущенные значения можно найти при решении системы (6):

$$z_1 = \frac{\Delta_1}{\Delta} = \frac{A_{11}C_1 - A_{12}C_2}{A_{11}^2 - A_{12}^2},$$

$$z_2 = \frac{\Delta_2}{\Delta} = \frac{A_{11}C_2 - A_{12}C_1}{A_{11}^2 - A_{12}^2}.$$

Распространим полученные результаты на общий случай, когда имеется m пропущенных наблюдений, запишем систему из m уравнений:

$$\left. \begin{aligned} A_{11}z_1 + A_{12}z_2 + \dots + A_{1m}z_m &= C_1, \\ A_{21}z_1 + A_{22}z_2 + \dots + A_{2m}z_m &= C_2, \\ \vdots & \\ A_{m1}z_1 + A_{m2}z_2 + \dots + A_{mm}z_m &= C_m, \end{aligned} \right\} \quad (7)$$

где

$$A_{11} = A_{22} = \dots = A_{mm} = n - 1 - \sum_{\tau=1}^n k_{\tau},$$

$$A_{12} = A_{21} = n - 1 - \sum_{l \in M_{12}} k_l,$$

$$A_{13} = A_{31} = n - 1 - \sum_{l \in M_{13}} k_l,$$

$$A_{(m-1)m} = A_{m(m-1)} = n - 1 - \sum_{l \in M_{(m-1)m}} k_l,$$

$$C_1 = \sum_{\tau=1}^n k_{\tau} T'_{\tau}(z_1) - (n-1) T'_{\text{общ}},$$

$$C_2 = \sum_{\tau=1}^n k_{\tau} T'_{\tau}(z_2) - (n-1) T'_{\text{общ}},$$

$$\vdots$$

$$C_m = \sum_{\tau=1}^n k_{\tau} T'_{\tau}(z_m) - (n-1) T'_{\text{общ}}.$$

Значения пропущенных наблюдений определяются из системы линейных уравнений, коэффициенты матрицы которой симметричны относительно ее диагонали.

После нахождения пропущенных значений z они считаются как бы действительно наблюдаемыми, но необходимо учесть, что полная и остаточная суммы квадратов отклонений теряют по 1, 2, ..., m степени свободы (в зависимости от числа пропущенных наблюдений).

Полученные результаты позволяют легко находить оптимальные оценки пропущенных наблюдений, проводить расчет по обычной схеме дисперсионного анализа и извлекать полезную информацию из результатов таких «плохих» экспериментов, анализ которых был сильно затруднен из-за нарушения условия ортогональности.

Предложенный метод использовался при решении задачи по изучению влияния ряда факторов на чувствительность методики титрования вируса гриппа В [1, 2]. Исследовалось влияние изменения рецепторной структуры нативных и стабилизированных эритроцитов под воздействием ряда ферментов, избирательно модифицирующих определенные рецепторные участки, на их способность к агглютинации. Одновременно изучалось действие некоторых углеводов и липидов на гемагглютинирующую активность вируса гриппа. Исследовались факторы, обозначенные в табл. следующим образом: X_1 —действие углеводов и липидов на 6-ти уровнях: физиологический раствор—0, глюкоза—1, манноза—2, фруктоза—3, лецитин—4, кефалин—5; X_2 —вид фермента на 5-ти уровнях: необработанные—0, трипсин—1, рибонуклеаза—2, гиалуронидаза—3, коллагеназа—4; X_3 —способ стабилизации эритроцитов на 4-х уровнях: нативные (НЭ)—0, акролеинизированные (АЭ)—1, глютарилованные (ГЭ)—2, формалинизированные (ФЭ)—3. Эксперимент состоял из 120 опытов ($6 \times 5 \times 4$), в которых использовались эритроциты человека; каждое титрование проводилось параллельно 3 раза.

Из-за нарушения нормального хода реакции (гемолиз эритроцитов) некоторые результаты эксперимента оказались потерянными. Для определения оценок пропущенных наблюдений вначале находились значения коэффициентов A и C по вышеприведенным формулам (здесь $k_1 = 6$, $k_2 = 5$, $k_3 = 4$):

$$A_{11} = A_{22} = A_{33} = A_{44} = 120 + 3 - 1 - 6 - 5 - 4 = 107,$$

$$A_{12} = A_{21} = 3 - 1 - 4 = -2,$$

$$A_{13} = A_{31} = 3 - 1 - 6 = -4,$$

$$A_{14} = A_{41} = 3 - 1 = 2,$$

$$A_{23} = A_{32} = 3 - 1 = 2,$$

Факторы		X_1					
X_2	X_3	0	1	2	3	4	5
0	0	5,3	5,7	6,7	6,0	0	5,3
0	1	0	3,3	6,0	5,0	6,0	1,7
0	2	z_1	2,5	7,7	8,3	8,0	5,0
0	3	2,7	5,0	6,3	4,3	0	3,3
1	0	5,7	6,0	7,0	6,0	3,3	5,0
1	1	8,3	7,7	9,0	9,0	6,0	0
1	2	10,0	10,3	11,0	10,7	10,0	z_2
1	3	8,7	9,0	10,0	10,0	0	5,3
2	0	5,7	6,7	7,0	6,0	5,3	3,3
2	1	7,3	5,0	7,5	11,0	8,0	3,3
2	2	9,0	7,3	8,7	9,0	7,3	3,3
2	3	z_3	5,0	8,0	z_4	7,0	5,0
3	0	5,7	6,0	8,0	7,0	6,3	5,0
3	1	1,7	5,0	6,7	6,3	8,0	1,7
3	2	9,0	8,7	10,3	8,0	7,7	5,0
3	3	0	4,3	8,7	9,0	2,5	5,0
4	0	5,0	5,3	6,3	5,3	0	1,7
4	1	1,7	5,0	7,3	6,0	1,7	0
4	2	9,0	6,7	10,3	10,0	6,7	3,3
4	3	6,7	5,7	8,5	7,0	0	3,3

$$A_{24} = A_{42} = 3 - 1 - 2,$$

$$A_{34} = A_{43} = 3 - 1 - 5 - 4 = -7.$$

$$C_1 = 6 \cdot 101,5 + 5 \cdot 104,1 + 4 \cdot 222,8 - (3 - 1) 685,7 = 649,3,$$

$$C_2 = 6 \cdot 65,5 + 5 \cdot 168 + 4 \cdot 222,8 - 2 \cdot 685,7 = 752,8,$$

$$C_3 = 6 \cdot 101,5 + 5 \cdot 145,7 + 4 \cdot 150,1 - 2 \cdot 685,7 = 566,5.$$

$$C_4 = 6 \cdot 143,9 + 5 \cdot 145,7 + 4 \cdot 150,1 - 2 \cdot 685,7 = 820,9.$$

Теперь систему линейных уравнений можно записать следующим образом:

$$107 z_1 - 2 z_2 - 4 z_3 + 2 z_4 = 649,3,$$

$$-2 z_1 + 107 z_2 + 2 z_3 + 2 z_4 = 752,8,$$

$$-4 z_1 + 2 z_2 + 107 z_3 - 7 z_4 = 566,5,$$

$$2 z_1 + 2 z_2 - 7 z_3 + 107 z_4 = 820,9.$$

В результате решения этой системы уравнений на ЭВМ «Мир-1» были определены значения: $z_1 = 6,28$; $z_2 = 6,90$; $z_3 = 5,91$; $z_4 = 7,81$. Подстановка найденных оценок пропущенных наблюдений позволила провести анализ эксперимента $6 \times 5 \times 4$, с использованием модификации алгоритма Йетса для несимметричных планов на ЭВМ «Мир-2» совместно с Кодкинд и Лисенковым [1].

Полученные результаты позволяют сделать вывод о том, что на гемагглютинирующую способность вирусов гриппа определенное действие оказывают углеводы и липиды. Добавление в физиологический раствор

маннозы и фруктозы приводит к четкому повышению средних геометрических титров вируса, а добавление липидов—к уменьшению титров.

Интересно отметить, что использование эритроцитов, стабилизированных формалином и глютаровым альдегидом, не снижает чувствительность методики титрования. Более того, при использовании глютаризованных эритроцитов наблюдается даже некоторое повышение средних геометрических титров вируса по сравнению с результатами опытов, в которых применялись нативные эритроциты.

Существенное влияние на результативный признак оказывает вид фермента. Причем необработанные или обработанные коллагеназой эритроциты человека дают невысокие титры, в то время как применение рибонуклеазы и трипсина значительно увеличивает средние геометрические титров.

Институт эпидемиологии, вирусологии и мед. паразитологии

МЗ Армянской ССР им. А. Б. Александяна

Поступило 8.I 1980 г.

ԲԺՇԿԱ-ԿԵՆՍԱԲԱՆԱԿԱՆ ՓՈՐՁԵՐԻ ԱՆԱԼԻԶԸ ԴԻՏԱՐԿՈՒՄՆԵՐԻ ՈՉ ԼՐԻՎ ԹՎԻ ԴԵՊՔՈՒՄ

Տ. Գ. ՕՀԱՆՅԱՆ

Բժշկա-կենսաբանական բազմագործոն հետազոտությունների ժամանակ միշտ չէ, որ ապահովվում են փորձերի հավասար թվով դիտարկումներ: Այդ հանգամանքը խիստ դժվարացնում է փորձերի արդյունքների մաթեմատիկական հետազոտությունը:

Աշխատանքում առաջարկված մեթոդը հնարավորություն է տալիս նշված դժվարությունների դեպքում կատարել արդյունքների մաթեմատիկական վերլուծություն:

Այդ մեթոդի օգնությամբ լուծված է կոնկրետ կենսաբանական խնդիր:

ANALYSIS OF MEDICO-BIOLOGICAL EXPERIMENTS WITH INCOMPLETE NUMBER OF OBSERVATIONS

T. G. OHANIAN

A method which allows to analyse the results of polyfactorous experiments when the number of observations is incomplete has been proposed.

Using the proposed method a problem of study of the effect of complex factors on the sensitivity of titration method has been solved.

Л И Т Е Р А Т У Р А

1. Грабовский П. М., Лисенков А. Н., Кодкин Г. Х., Оганян Т. Г., Козловская Г. И., Таранцева Л. П. В сб.: Вопросы медицинской вирусологии. М., 535, 1975.
2. Оганян Т. Г. Автореф. канд. дисс., М., 1978.
3. Хикс Ч. Основные принципы планирования эксперимента. М., 1967.
4. Шеффе Г. Дисперсионный анализ. М., 1963.