ՀԱՅԱՍՏԱՆԻ ԿԵՆՍԱԲԱՆԱԿԱՆ ՀԱՆԴԵՍ
БИОЛОГИЧЕСКИЙ ЖУРНАЛ АРМЕНИИ

XXXII, 2, 1979

УДК 57.087

НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ АНАЛИЗА МНОГОФАКТОРНЫХ МЕДИКО-БИОЛОГИЧЕСКИХ ЭКСПЕРИМЕНТОВ

Т. Г. ОГАНЯН

Для проведения множественных сравнений, а также анализа ряда типовых задач планирования медико-биологического эксперимента предлагается применять наиболее чувствительные непараметрические критерии типа критериев нормальных меток и Ван дер Вардена.

Если результаты дисперсионного анализа указывают на существование значимого различия в средних для разных популяций, то часто такого общего вывода бывает недостаточно и нужно решить, какие именно из изучаемых групп отличаются от других. Фишер предлагал использовать в схемах дисперсионного анализа t-критерий для проведения множественных сравнений между уровнями фактора [1]. Наиболее часто после обычного анализа по F-критерию с целью выявления значимых различий в уровнях факторов и контрастах используются критерии множественных сравнений Шеффе и Тьюки, а также процедуры Ньюмена-Кейльса и Дункана [1]. Процедуры, использующие критерий ранговых сумм Вилкоксона, $S=\Sigma R_j$, для множественных сравнений всех $K(K-1)/2$ плотностей для K исследуемых популяций, рассматривались в некоторых работах [3, 4].

Непараметрические методы анализа многофакторных экспериментов, основанные на критериях нормальных меток и Ван дер Вардена, рекомендуется использовать в случае, когда результаты медико-биологического эксперимента не могут быть обработаны обычными методами дисперсионного и регрессионного анализа из-за нарушения соответствующих предпосылок (нормальность распределения отклика, однородность дисперсий и т. п.). Кроме того, часто в медико-биологических исследованиях используются не только количественные определения, но и полуколичественные и качественные—тогда непараметрические методы являются единственными, позволяющими проводить анализ.

Будем использовать после ранжирования значений выходного показателя в едином ряду вместо меток $a_N(R_j)$ следующие преобразования: а) $a\ (R_j)=E[R_j,\ N]$, где $E[R_j,\ N]$ —так называемые нормальные метки, являющиеся математическими ожиданиями R_j-тых порядковых статистик выборки объема N из нормальной совокупности. Значения нормальных меток задаются выражением [2]:

$$E[R_j, N] = \frac{N!}{(R_j-1)!(n-R_j)!}\int_{-\infty}^{\infty}\xi[F(\xi)]^{R_j-1}[1-F(\xi)]^{n-R_j}dF(\xi) =$$

$$= \frac{N!}{(R_j-1)!(n-R_j)!}\int_{-\infty}^{\infty}\xi[F(\xi)]^{R_j-1}[1-F(\xi)]^{n-R_j}f(\xi)\,d\xi,$$

где $F(\xi) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\xi} e^{-\frac{x^2}{2}}dx$ — функция нормального распределения,

$f(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}}$ — функция плотности нормального распределения.

б) $a_N(R_j) = W^{-1}\left(\frac{R_j}{N+1}\right)$, где W^{-1} функция, обратная интегралу вероятностей нормального распределения.

Из критериев Фишера-Петса-Терри-Гёфдинга и Ван дер Вардена известно, что статистики типа $S = \sum a_N(R_j)$ имеют асимптотически нормальное распределение [5]. Линейные комбинации таких сумм для соответствующих контрастов,

$$c_\lambda \equiv \frac{\sum_i S_i}{\sum_i n_i} - \frac{\sum_{i'} S_{i'}}{\sum_{i'} n_{i'}};\ \lambda = 1, 2, \cdots, P,$$

также будут иметь асимптотически нормальное распределение. Дисперсия каждого контраста c_λ будет:

$$D[c_\lambda] = \sum_{j=1}^{N} a_N(R_j)\left[\frac{1}{\sum_i n_i} + \frac{1}{\sum_{i'} n_{i'}}\right] \Big/ (N-1),$$

где $a_N(R_j)$ — нормальные метки, если используется преобразование (а) или значения функции обратной интегралу вероятностей нормального распределения, в случае (б) $N = \sum_i n_i + \sum_{i'} n_{i'}$.

Образуя отношения

$$z_\lambda = c_\lambda/[D(c_\lambda)]^{1/2}$$

для каждого из P интересующих контрастов, можно проверить нулевую гипотезу, сравнивая эти отношения с $z_{1-\frac{\alpha}{2p}}$, где $1-\frac{\alpha}{2p}$ — точка стандартного нормального распределения. Эффект контраста положителен, если расчетное значение $z_\lambda > z_{1-\frac{\alpha}{2p}}$, отрицателен, если $z_\lambda < -z_{1-\frac{\alpha}{2p}}$, и равен нулю, если $-z_{1-\frac{\alpha}{2p}} < z_\lambda < z_{1-\frac{\alpha}{2p}}$. Если нулевая гипотеза верна, т. е. все К популяции идентичны, вероятность сделать правильное решение о Р-контрастах равна, по крайней мере, $1-\alpha$ [3].

На основе рассмотренных процедур множественных сравнений можно разработать метод анализа планов 2^k. В этом случае для множества ортогональных контрастов плана выражения для соответствующих эффектов запишутся следующим образом:

$$B_\lambda = \frac{\sum\limits_u S_{1u}}{\sum\limits_u n_{1u}} - \frac{\sum\limits_u S_{2u}}{\sum\limits_u n_{2u}},$$

где n_u — число дублирований u-го опыта, $N = 2^k$ — число опытов плана, $\sum\limits_u S_{1u}$ и $\sum\limits_u S_{2u}$ — суммы меток, полученные для верхнего уровня фактора (или взаимодействия) и нижнего соответственно.

Значения B_λ делятся на соответствующие среднеквадратичные ошибки

$$\sigma_\lambda = \sqrt{\sum_{j=1}^{\tilde{N}} a^2_{\tilde{N}}(R)_j \left[\frac{1}{\sum\limits_u n_{1u}} + \frac{1}{\sum\limits_u n_{2u}}\right] / (\tilde{N}-1)},$$

где $\tilde{N} = \sum\limits_u n_u$.

Эти процедуры можно обобщить в случае ортогональных многоуровневых планов, когда, кроме проверки эффектов факторов и их взаимодействий, необходима проверка гипотезы о наличии линейных, квадратичных и т. д. составляющих (тренда). Для такого анализа используются статистики, подобные предложенным в работе Оганяна и Кодкинд [6].

$$B^k_\lambda = \sum_{i=1}^{l} \frac{\Theta_{ki} S_i}{n_i},$$

где n_i — число дублирований i—того опыта, l — число уровней исследуемого фактора, Θ_{ki} — значения ортогональных полиномов, которые задаются в следующем виде [6]:

$$\Theta_{ki} = \alpha + \beta x_i + \gamma x_i^2 + \cdots + \eta_l x_i^k,$$

где $x_i = 0, 1, 2, \cdots, (l-1)$; $k = 1, 2, \cdots, (l-1)$; константы α, β, $\gamma, \cdots, \eta_l$ определяются из условий ортогональности:

$$\sum \Theta_{ki} = 0;\ \sum \Theta_{1i}\,\Theta_{ki} = 0; \cdots,\ \sum \Theta_{k-1,\,i}\,\Theta_{ki} = 0.$$

Статистики B^k_λ имеют асимптотически нормальное распределение с математическим ожиданием $M\{B^k_\lambda\} = 0$ и дисперсией

$$D\{B^k_\lambda\} = \frac{1}{(\tilde{N}-1)} \sum_{i=1}^{l} \frac{\Theta^2_{ki}}{n_i} \sum_{j=1}^{n_i} a^2_{\tilde{N}}(R_{ij}).$$

Проверка гипотезы о наличии тренда k-го порядка сводится к вычислению отношений $z^k_\lambda = B^k_\lambda / \sigma^k_\lambda$ и сравнению их с критическим значением стандартного нормального распределения $z_1 - \frac{\alpha}{2\tilde{p}}$, где $\tilde{p}$ — число оцениваемых эффектов B_{λ}^{k}.

Рассмотрим пример применения предложенной методики при изучении влияния низких температур на активность дегидрогеназ моллюсков видов Bradybaena fruticum и Bradybaena schrencke [7]. Независимые переменные приведены в табл. 1. Матрица эксперимента дана в табл. 2.

Таблица 1

Факторы и их уровни

Независимые переменные	$\tilde{X}_1$ — принадлежность к виду	$\tilde{X}_2$ — морфа	$\tilde{X}_3$ — особенности препарата	$\tilde{X}_4$ — температурный режим
Верхний уровень	B. fruticum (1)	полосатая (1)	гомогенат (1)	(3) — 15° (2) — 8°
Нижний уровень	B. schrencke (0)	бесполосая (0)	гиалоплазма (0)	(1) —1—1° (0) 8°

Таблица 2

Матрица плана 2^3, совмещенная с 2×4 латинским прямоугольником и результаты эксперимента

Кодированные переменные					Тетразольвосстанавливающая активность гомогената и гиалоплазмы, г/мл в атмосфере (аргон)			
№	X_1	X_2	X_3	X_4	$У_1$	$У_2$	$У_3$	$У_4$
1	1	1	1	3	0,240	0,225	0,200	0,215
2	1	0	!	2	1,460	1,200	1,265	1,235
3	0	1	1	1	0,330	0,300	0,405	0,375
4	0	0	1	0	0,180	0,170	0,275	0,250
5	1	1	0	0	0,230	0,215	0,275	0,280
6	1	0	0	1	0,765	0,690	0,710	0,675
7	0	1	0	2	0,180	0,130	0,185	0,175
8	0	0	0	3	0,120	0,155	0,120	0,155

Необходимость применения непараметрических методов анализа в этой задаче была вызвана тем, что при проверке по W-критерию [8] не подтвердилась гипотеза нормальности: $W\ (7) = 0{,}771 < W_{0.1}\ (4) = 0{,}792$; $W\ (8) = 0{,}731 < W_{0.05}\ (4) = 0{,}748$.

Кроме нарушения условия нормальности распределения отклика, наблюдается резкое отличие двух значений выходной величины (2-я и 6-я строки) от остальных (табл. 2) На ЭВМ «Мир-2» были подсчитаны коэффициенты регрессии, однако при проверке по F-критерию модель признана неадекватной: $F_p = 45{,}7 > F_{0{,}05}\ (2{,}24) = 3{,}4$. Не удалось получить адекватного описания и после преобразования отклика Y на lnY.

Для решения этой задачи далее применялся разработанный непараметрический метод анализа (табл. 3).

Таблица 3

Математические ожидания порядковых статистик для результатов многофакторного эксперимента

№	I	II	III	IV	$S_i = \sum_{j=1}^{n_i} E[R_{ij}, \tilde{N}]$
1	−0,04	−0,20	−0,44	−0,32	−1,00
2	2,07	1,22	1,65	1,40	6,34
3	0,44	0,36	0,62	0,53	1,95
4	−0,67	−0,94	0,16	0,04	−1,41
5	−0,12	−0,32	0,16	0,28	0
6	1,07	0,82	0,94	0,72	3,55
7	−0,67	−1,40	−0,53	−0,82	−3,42
8	−1,86	−1,15	−1,86	−1,15	−6,02

Вначале значения отклика заменялись соответствующими математическими ожиданиями порядковых статистик выборки объема $\tilde{N} = 32$ из стандартной нормальной совокупности.

Значение эффекта первого фактора подсчитывалось следующим образом: $B_1 = \frac{2}{32}(-1+6{,}34-1{,}95+1{,}41+0+3{,}55+3{,}42+6{,}02) = 1{,}112.$

Дисперсия статистики B_1 равна

$$D\{B_1\} = \frac{4 \cdot 29{,}42}{32(32-1)} = 0{,}119.$$

Далее определялось

$$z_1 = \frac{1{,}112}{\sqrt{0{,}119}} = 3{,}23.$$

Аналогично вычислялись Z—отношения для остальных факторов: $z_2 = -0{,}89$; $z_3 = 2{,}14$; $z_4^1 = -1{,}45$; $z_4^2 = -3{,}01$; $z_4^3 = 0{,}16$.

Табличное значение $z_1 - \frac{\alpha}{2\tilde{p}}$ при $\alpha = 0{,}05$ и $\tilde{p} = 7$ оказалось равным 2,91 [9]. При сравнении вычисленных значений z_λ^k с табличными значимыми оказались два эффекта, B_1 и B_4^2 ($z_1 = 3{,}23$; $z_4^2 = -3{,}01$).

Таким образом, удалось установить, что на тетразольвосстанавливающую активность гомогената и гиалоплазмы существенное влияние оказывает эффект I фактора (принадлежность к виду) и квадратичный эффект IV фактора (температурный режим).

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

Поступило 10.XI 1978 г.

ԲԺՇԿԱ-ԿԵՆՍԱԲԱՆԱԿԱՆ ԲԱԶՄԱՖԱԿՏՈՐ ՓՈՐՁԵՐԻ ՎԵՐԼՈՒԾՈՒԹՅԱՆ ՈՉ ՊԱՐԱՄԵՏՐԻԿ ՄԵԹՈԴՆԵՐԸ

Տ. Գ. ՕՀԱՆՅԱՆ

Հայտնի է, որ բազմաֆակտոր փորձի պլանավորման և վերլուծության մեթոդները բարձրացնում են գիտական հետազոտության արդյունավետությունը 2—10 անգամ։ Սակայն այդ մեթոդների կիրառումը կենսաբանության և բժշկության բնագավառներում կապված է որոշակի դժվարությունների հետ։ Այդ սահմանափակումները հաջողվեց հաղթահարել ոչ պարամետրիկ մեթոդների օգնությամբ, որոնց կիրառումը չի պահանջում նորմալ բաշխման կատարում, ցրման համասեռություն և այլն։

Հետազոտության արդյունքները հնարավորություն են տալիս ստանալու բանաձևեր, որոնք ապահովում են բժշկա-կենսաբանական փորձերի պլանավորման վերլուծությունը։ Առաջարկվող մեթոդի արդյունավետությունը ցույց է տրված կենսաբանական որոշակի խնդրի լուծման ժամանակ։

NONPAMETRICAL METHODS OF MULTIFACTOR MEDICO-BIOLOGICAL EXPERIMENT ANALYSIS

T. G. OGANIAN

It is offered to use the most sensitive nonparametric criterions of normal label and of Van der Varden types to carry out multiple comparisons, as well as to analyse medico-biological experiment planning problems.

ЛИТЕРАТУРА

1. *Дэвид Г.* Сб. Введение в теорию порядковых статистик. 135—151, М., 1970.
2. *Огава Дз.* Сб. Введение в теорию порядковых статистик. 21—28, М., 1970.
3. *Dunn O. J.* Technometrics, 6, 241—252, 1964.
4. *Лисенкоз А. Н., Никитина Е. П., Оганян Т. Г.* Непараметрические методы анализа многофакторных экспериментов. М., 1976.
5. *Гаек Я., Шидак З* Теория ранговых критериев. М., 1971.
6. *Оганян Т. Г., Кодкинд Г. Х.* Сб. Статистические методы планирования и анализа медико-биологических экспериментов. 23—27, Киев, 1975.
7. *Рункова Г. Г., Оганян Т. Г.* Сб. Биохимические аспекты некоторых проблем популяционной экологии. 32—45, Свердловск, 1978.
8. *Хан Г., Шапиро С* Статистические модели в инженерных задачах. М., 1969.
9. *Урбах В. Ю.* Математическая статистика для биологов и медиков. М., 1963.