6. Лебедева М. Н., Анищенко Э. Я. 1966. Допустимо ли применение батометров Нансена для количественных микробиологических исследований? «Гидробиол. ж.», 2, 4.
7. Могилевский Г. А., Кузнецова З. И. 1949. Батометр. Авт. св-во № 75408 от 31 янв. 1949 г.
8. Романенко В. И., Младова Г. А. 1969. Шарообразные стеклянные баллоны для стерильного отбора проб воды с больших глубин. Инф. бюлл. «Биол. внутр. вод», 4, изд-во «Наука», Л.
9. Сорокин Ю. И. 1960. Батометр для отбора проб воды на бактериологический анализ. «Бюлл. Ин-та биол. водохр.», 6.
10. Его же. 1962. Вопросы методики отбора проб при изучении морской микрофлоры. «Океанология», 2, 5.
11. Его же. К вопросу о методике микробиологических работ в море в свете современных задач морской микробиологии. «Океанология», 4, 2.
12. Столбунов А. К., Рябов Ф. П. 1963. Новая модель бактериологического батометра. «Лабораторное дело», 10.
13. Helland-Nansen B., Nansen F. 1925. The Eastern North Atlantic. «Geophys. Publ.», 4, 6—7.
14. Jannasch H. W., Maddux W. S. 1967. A Note on Bacteriological Sampling in Seawater. «J. Marine Res.», 25, 2.
15. Lewis W. M., McNail O. D., Summerfelt R. C. 1963. A devise for taking water samples in sterille bottles at various depths. «Ecology», 44, 1.
16. Niskin S. J. 1962. A water sampler for microbiological studies. «Deep-Sea Res.», 9.
17. Schegg E. 1970. A new bacteriological sampling bottle. «Limnolog. and Oceanogr.», 15, 5.
18. Seki H. 1971. Microbial clumps in seawater in the euphotic zone of Saanich Inlet (British Columbia). «Mar. Biol.», 9, 1.
19. Sieburth J. McN. 1965. Bacteriological samplers for air-water and water-sediment interfaces. «Ocean. Scie. Ocean Engineer. Trans. MTS-ASLO Conf. Wash.», 2.
20. Van Dorn W. G. 1956. Large volume water samplers. «Trans Am. Geophys. Union», 37.
21. Willingham C. A., Buck J. D. 1965. A preliminary comparative study of fungal contamination in non sterile water samples. «Deep-Sea Res.», 12, 5.
22. Zo Bell C. E. 1941. Apparatus for collecting water samples from different depths for bacteriological analysis. «J. marine Res.», 4.

Поступила 27.XII 1971 г.

УДК 577.472:51

ОПРЕДЕЛЕНИЕ ОПТИМАЛЬНОГО ЧИСЛА ПРОБ ПРИ ИССЛЕДОВАНИИ МНОГОКОМПОНЕНТНЫХ БИОЛОГИЧЕСКИХ СИСТЕМ

В. Л. АНДРЕЕВ, В. Е. МОЛОТКОВ

(Лаборатория математического моделирования экологических систем ДВНЦ АН СССР, Владивосток)

При изучении экологических систем исследователю, как правило, приходится прибегать к выборочному методу. При этом его естественным желанием является установление точности, с которой по показателям выборки можно судить о показателях природной системы. Кажется достаточно очевидным, что, чем больше размер пробы, тем точнее оцениваются искомые показатели. С другой стороны, если все объекты внутри какой-либо генеральной совокупности (г. с.) приблизительно одинаковы по изучаемому признаку, то нет необходимости тратить время и средства на изучение выборки большого размера: увеличение точности оценки не оправдает средств, затраченных на ее получение.

Таким образом, возникает типичная задача оптимизации размера пробы.

Будем считать оптимальной такую выборку, которая характеризует показатели г. с. с заданной наперед точностью и малой вероятностью ошибочного суждения о ней.

Одномерный случай (когда изучается единственная компонента) — признак — г. с.) хорошо известен [4]. А именно: оценка среднего значения компоненты в г. с. с заданной вероятностью находится в интервале *:

$$|\bar{x}-\hat{x}|=t_\alpha\sigma_{\bar{x}}=\frac{t_\alpha\sigma}{\sqrt{N}}, \tag{1}$$

где $\bar{x}$ и $\hat{x}$ соответственно средние арифметические выборки и г. с.; σ — среднеквадратичное отклонение от $\hat{x}$ компоненты в г. с.; t_α — величина, указывающая «степень доверия» к вычисленному интервалу (обычно в экологических исследованиях $t_\alpha=2$, что соответствует вероятности ошибочного суждения о доверительном интервале $\bar{x}\pm m$ около 5%). Из формулы (1) легко установить, что:

$$N=\frac{t_\alpha^2\sigma^2}{m^2}, \tag{2}$$

где $m=|\bar{x}-\hat{x}|$ — заданная наперед точность оценки $\hat{x}$.

Пример 1. Численность планктеров группы Rotatoria (на 1 $м^3$) в пробах из выростного пруда соответствует нормальному распределению со следующими оценками параметров:

$\bar{x}=12{,}5$ тыс. шт.; $S=10{,}024$ тыс. шт.

Определить число станций, необходимое для оценки $\hat{x}$ с точностью 6 тыс. шт. и вероятностью ошибки не более 5%.

Решение.

$$N=\frac{t_\alpha^2S^2}{m^2}\approx\frac{4\cdot 105{,}1}{36}\approx 11.$$

К сожалению, подобные задачи (когда исследуется единственная компонента г. с.) на практике встречаются редко. Чаще всего исследователям приходится сталкиваться с проблемой планирования наблюдений сразу по нескольким переменным. Но именно данный случай (насколько известно авторам) не нашел своего отражения в биологической литературе. Целью настоящей работы и является устранение этого досадного пробела.

Если в г. с. изучается несколько признаков, то для количественной характеристики их изменчивости недостаточно знания одних только дисперсий (σ^2), необходимо учитывать также и их взаимную сопряженность (ковариацию).

Допустим, что изучаемые $к$ признаков в г. с. имеют следующий набор (вектор) средних значений: $\hat{x}_1$, $\hat{x}_2 \ldots$ $\hat{x}_k$. Тогда их изменчивость можно охарактеризовать квадратной таблицей (матрицей порядка $к$), в которой по диагонали расположены дисперсии, а недиагональные элементы представляют собой ковариации. Например, для $к=3$:

$$\|C\|=\left\|\begin{matrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{matrix}\right\|,$$

* В предположении, что изучаемый признак распределен по нормальному закону.

где C_{ij} — ковариация между i-тым и j-м признаками

$$(i=1,2,\ldots,k;\ j=1,2,\ldots,k);\quad C_{ij}=\frac{1}{M}\sum_{n=1}^{M}(x_{in}-\hat{x}_i)(x_{jn}-\hat{x}_j). \tag{3}$$

$C_{ii}=\sigma_i^2$ — дисперсия i-того признака:

$$C_{ii}=\frac{1}{M}\sum_{n=1}^{M}(x_{in}-\hat{x}_i)(x_{in}-\hat{x}_i)=\frac{1}{M}\sum_{n=1}^{M}(x_{in}-\hat{x}_i)^2 \tag{4}$$

(M — численность г. с.).

Если из данной совокупности производится выборка, то, как доказано [2], наилучшими оценками $\hat{x}_1$, $\hat{x}_2 \ldots$ $\hat{x}_k$ являются $\bar{x}_1$, $\bar{x}_2 \ldots$ $\bar{x}_k$, а оценками дисперсий и ковариаций:

$$S_{ij}=\frac{1}{N-1}\sum_{n=1}^{N}(x_{in}-\bar{x}_i)(x_{jn}-\bar{x}_j). \tag{5}$$

Доверительную область для $\bar{x}_1$, $\bar{x}_2 \ldots$ $\bar{x}_k$ находим из следующей формулы [1]:

$$T^2 \geqslant N\sum_{i=1}^{k}\sum_{j=1}^{k} m_i m_j S^{ij}, \tag{6}$$

где S^{ij} — элементы матрицы, обратной $\|S\|$. T^2 — многомерный аналог t в формуле (1). При этом величина T^2 распределена как F-критерий [1] с k и $N-k$ числом степеней свободы. Из формулы (6) легко находится N.

Пример 2. При гидробиологическом обследовании выростного пруда № 2 (см. рисунок, табл. 1) карпового хозяйства «Ханкайский» Приморского края от 9 августа 1970 г. в шести точках (станциях) получены следующие показатели численности трех групп планктеров (табл. 1) — Cladocera, Copepoda, Rotatoria (тыс. *экз/м*^{3}).

Требуется определить число станций, необходимое для оценки среднего числа указанных групп планктеров с точностью $m_1=9$ тыс. шт., $m_2=10$ тыс. шт., $m_3=7$ тыс. шт. и вероятностью ошибки не более 5%.

Решение:

1. Подсчитаем значения ковариационной матрицы. Для этого предварительно каждую варианту убавим на величину ее средней арифметической, а затем перемножим варианты в строках, соответствующих номерам элементов матрицы. Получим следующий результат:

$$\|S\|=\begin{matrix} S_{11}=1159{,}02 & S_{12}=\ 667{,}067 & S_{13}=124{,}99 \\ S_{21}=\ 667{,}067 & S_{22}=1350 & S_{23}=358{,}56 \\ S_{31}=\ 124{,}99 & S_{32}=\ 358{,}56 & S_{33}=105{,}1 \end{matrix}$$

Например, первый элемент подсчитывается следующим образом:

$$S_{11}=\frac{1}{6-1}[(20-38{,}55)^2+(2-38{,}55)^2+$$
$$+(49-38{,}55)^2+\ldots+(99-38{,}55)^2]=1159{,}02;$$

второй элемент:

$$S_{12}=\frac{1}{6-1}\left[(20-38{,}55)(30-65{,}93)+(2-38{,}55)(52-65{,}93)+\ldots+(99-38{,}55)(115-65{,}93)\right]=667{,}067$$

(т. е. перемножаются значения первой и второй строчек) и т. д. Ясно, что матрица ковариаций всегда симметрична ($S_{ij}=S_{ji}$) и, следовательно, можно подсчитывать не k^2 значений, а всего лишь $k(k+1)/2$.

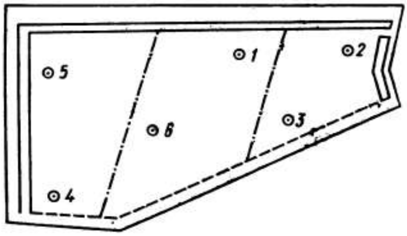

Схема расположения стандартных станций в выростном пруду № 2 (Ханкайский р-н Приморского края).

2. Обращение матрицы ковариаций * дало следующие результаты:

$$\|S^{ij}\|=\begin{Vmatrix} 0{,}00181 & -0{,}00342 & 0{,}00953 \\ -0{,}00342 & 0{,}01437 & -0{,}04496 \\ 0{,}00953 & -0{,}04496 & 0{,}152 \end{Vmatrix}$$

3. Подсчитаем некоторую величину D^2 **:

а) находим вспомогательные величины A_i ***:

$$A_1=m_1S^{11}+m_2S^{12}+m_3S^{13}=0{,}0495$$

$$A_2=m_1S^{21}+m_2S^{22}+m_3S^{23}=0{,}20136$$

$$A_3=m_1S^{31}+m_2S^{32}+m_3S^{33}=0{,}7008$$

б) $$D^2=m_1A_1+m_2A_2+m_3A_3=3{,}34.$$

4. По таблице распределения F-критерия (см. например, [4]) находим критические значения F, соответствующие $k=3$ и $N-k=3^1$ степеням свободы и определяем

$$T^2_\alpha=\frac{(N-1)k}{N-k}\cdot F_{k,N-k}=\frac{5\cdot 3}{3}\cdot 9{,}28=46{,}45.$$

Таблица 1

Численность планктеров (тыс. *экз/м*^{3}) **в пробе 9 августа 1970 г.** (выростной пруд № 2)

Группы планктона	Станции						$\bar{x}$
	1	2	3	4	5	6	
Cladocera	20	2	49	41,3	20	99	39,55
Copepoda	30	52	66,2	28,4	104	115	65,93
Rotatoria	3	7	11	4	28	22	12,50

* Приемы обращения матриц можно найти в любых справочниках по высшей математике.

** Эта величина называется «обощенным расстоянием Махаланобиса».

*** Величины A_i являются коэффициентами т. н. «дискриминантной функции».

5. Искомое число станций определяем так:

$$N = T_{\alpha}^{2}/D^{2} \approx 14.$$

В этом примере следует обратить внимание на следующее обстоятельство. Если подсчитывать оптимальное число станций по формуле (1), т. е. для каждой группы планктеров в отдельности, то получим следующие результаты: $N_1=56$, $N_2=54$, $N_3=8$. Иначе говоря, число станций оказывается в среднем гораздо большим, чем при предлагаемом методе для той же точности и вероятности ошибки. Происходит это потому, что в первом случае изменения численности каждой группы видов от станции к станции считаются независимыми. Иначе говоря, заранее утверждается, что наличие определенного числа экземпляров какой-либо группы на станции не несет никакой информации о численности других групп. На самом же деле это не так. Если подсчитать коэффициент корреляции, например, между второй и третьей группами, то он окажется равным 0,95. Из этого следует, что, если на данной станции численность Cladocera велика, то с большей вероятностью можно ожидать здесь же и большого количества Rotatoria. То же происходит и в отношении других групп планктеров. Таким образом, оказывается, что даже по численности в пробах одной группы мы получаем некоторое количество информации и о других группах. Именно это обстоятельство и учитывается в предлагаемом методе и именно оно позволяет существенно сократить усилия по сбору первичной информации

В самом деле, произведя соответствующие расчеты для проб планктона из выростного пруда № 2 Ханкайского района за летне-осенний сезон, воспользовавшись обоими методами, мы получили следующие результаты (табл. 2).

Точность для всех средних арифметических задавалась равной 35% их значений. Данные табл. 2 показывают, что во всех анализируемых случаях оптимальное число станций. определяемое по предлагаемому методу, существенно ниже результатов традиционного метода. Так, если для одномерного случая среднее число станций колеблется от 13 до 132, то для многомерного — от 5 до 24. Верхний предел этих колебаний для настоящего примера указывает на то, что в отдельных съемках при числе станций, равном 6, точность оценки численности групп планктеров весьма низка, и, следовательно, число станций в этот период следует существенно увеличить.

Полученные данные могут навести на мысль, что, чем больше будет учтено признаков (например, если учитывать не группы видов, а каждый вид в отдельности), тем эффективнее окажутся результаты. Однако это может произойти только в том случае, если число станций достаточно велико. Для практических приложений метода следует руководствоваться следующим правилом: число признаков должно быть примерно в два раза меньше числа станций. В противном случае матрица ковариаций может оказаться положительно не определенной [3] и исследователь получит неверные результаты.

Существует ли возможность увеличить точность искомых оценок за счет рационального распределения усилий во времени и пространстве?

Такая возможность в данном случае существует, если изменчивость изучаемых признаков разбивается на пространственно-временные «слои», или зоны.

Рассмотрим конкретные замерения, проведенные в 1970 г. в упомянутом водоеме, и подсчитаем дисперсию признаков за весь сезон на каждой стадии в отдельности. Для этого составим ковариационные матрицы всех признаков для каждой стадии и вычислим их детерминанты (определители):

Станции	1	2	3	4	5	6
Определители ковариаций матрицы	$12{,}6\cdot 10^{10}$	$178\cdot 10^{10}$	$177\cdot 10^{10}$	$375\cdot 10^{10}$	$489\cdot 10^{10}$	$86{,}8\cdot 10^{10}$

Сопоставляя приведенные результаты и данные рисунка, можно заметить, что наибольшей изменчивостью признаков отличаются станции № 4 и 5 (зона 1), наименьшей — № 1 и 6 (середина пруда). Станции № 2 и 3 (зона 3) занимают промежуточное положение. Это и понятно, т. к. в местах наибольшей изменчивости числа планктеров (зона 1) осуществляется поступление воды из реки в течение года, в то время как в остальных зонах пруда условия более стабильны.

Таблица 2

Оптимальное число станций, определенное разными способами
(выростной пруд № 2 Ханкайского рыбхоза Приморского края)

Дата	«Одномерный» способ	«Многомерный» способ
10.VI	$\bar{x}_1=90{,}8$ $\bar{x}_2=55$; $\bar{x}_3=395{,}6$ $N_1=41$; $N_2=3$; $N_3=74$; $\bar{N}=36$	$N=18$
20.VI	$\bar{x}_1=46{,}7$; $\bar{x}_2=110$; $\bar{x}_3=334{,}8$; $N_1=21$; $N_2=29$; $N_3=28$; $\bar{N}=26$	$N=24$
30.VI	$\bar{x}_1=79{,}3$; $\bar{x}_2=87$; $\bar{x}_3=23{,}2$ $N_1=12$; $N_2=43$; $N_3=1$; $\bar{N}=18$	$N=6$
9.VII	$\bar{x}_1=113{,}6$; $\bar{x}_2=28{,}3$; $\bar{x}_3=4{,}8$ $N_1=92$; $N_2=2$; $N_3=13$; $\bar{N}=36$	$N=5$
19.VII	$\bar{x}_1=17{,}3$; $\bar{x}_2=48{,}8$; $\bar{x}_3=392$ $N_1=10$; $N_2=34$; $N_3=100$; $\bar{N}=48$	$N=8$
29.VII	$\bar{x}_1=221{,}8$; $\bar{x}_2=62{,}2$; $\bar{x}_3=448{,}5$ $N_1=112$; $N_2=6$; $N_3=200$; $\bar{N}=106$	$N=46$
9.VIII	$\bar{x}_1=38{,}5$; $\bar{x}_2=65{,}9$; $\bar{x}_3=12{,}7$ $N_1=56$; $N_2=54$; $N_3=8$; $\bar{N}=37$	$N=14$
19.VIII	$\bar{x}_1=29$; $\bar{x}_2=23$; $\bar{x}_3=3{,}1$ $N_1=350$; $N_2=40$; $N_3=8$; $\bar{N}=132$	$N=19$
29.VIII	$\bar{x}_1=42{,}7$ $\bar{x}_2=82{,}7$; $\bar{x}_3=6{,}7$ $N_1=40$; $N_2=24$; $N_3=3$; $\bar{N}=22$	$N=13$
8.IX	$\bar{x}_1=124$; $\bar{x}_2=162$; $\bar{x}_3=94{,}6$ $N_1=60$; $N_2=28$; $N_3=39$; $\bar{N}=42$	$N=24$
18.IX	$\bar{x}_1=19{,}6$; $\bar{x}_2=200$; $\bar{x}_3=88{,}5$ $N_1=257$; $N_2=20$; $N_3=2$; $\bar{N}=18$	$N=10$

Таким образом, пространство водоема по этому признаку можно приблизительно разбить на три части, как показано пунктирными линиями на рисунке. Усредняя дисперсии в каждой зоне, получаем: $D_1=43\,2\cdot 10^{10}$; $D_2=49{,}7\cdot 10^{10}$; $D_3=177{,}5\cdot 10^{10}$.

Теперь можно получить лучшую репрезентативность вторичных данных (по сравнению со случайным расположением станций внутри всего водоема), если распределять объем выборки так, чтобы ее части были пропорциональны значениям вариабельности:

$$N_i = N \frac{D_i}{\sum_{i=1}^{r} D_i},$$

где r — число зон; D_i — дисперсия признаков i-той зоны.

Например, 10 станций для выростного пруда № 2 нужно распределить следующим образом:

$$N_1 = 7 \text{ ст.}, \ N_2 = 1 \text{ ст.}, \ N_3 = 2 \text{ ст.},$$

т. е. в первой зоне распределяются семь станций, во второй — одна, в третьей — две.

Аналогично можно провести распределение числа станций и во времени, однако для этой цели в данном случае требуется многолетний материал, которым мы в настоящее время не располагаем *.

Естественно, что описываемый метод применим не только для оценки численности планктеров в водоеме, но также во всех тех случаях, когда исследователю приходится обращаться к выборочному методу при одновременном изучении нескольких признаков генеральной совокупности.

ЛИТЕРАТУРА

1. Андерсон Т. 1963. Введение в многомерный статистический анализ. Изд-во «Наука», М.
2. Рао С. Р. 1968. Линейные статистические методы и их применение. Изд-во «Наука», М.
3. Уилкс С. С. 1967. Математическая статистика. Изд-во «Наука», М.
4. Урбах В. Ю. 1964. Биометрические методы. Изд-во «Наука», М.

Поступила 24. XI 1971 г.

УДК 551.482.214

К МЕТОДИКЕ ОБЪЕМНОГО ОПРЕДЕЛЕНИЯ СУЛЬФАТ-ИОНОВ В ПРИРОДНЫХ ВОДАХ РАЗЛИЧНОЙ МИНЕРАЛИЗАЦИИ

И. Г. ЕНАКИ, Б. И. НАБИВАНЕЦ

(Институт гидробиологии АН УССР, Киев)

Наиболее часто применяемый в гидрохимической практике классический весовой метод определения сульфатов, хотя и довольно точен, но трудоемок и требует ряда продолжительных операций. Поэтому при массовых анализах его стремятся заменить менее продолжительными и трудоемкими объемными или фотоэлектротурбидиметрическими методами. Преимущество первых в том, что они не требуют специального оборудования (аналитических весов, муфельных печей, оптических приборов и т. д.) и могут быть применены в полевых условиях.

* Кроме того, необходимо, чтобы изменения численности планктона из года в год в течение сезона происходили примерно одинаково.