Том LXV | Труды Всесоюзного научно-исследовательского института морского рыбного хозяйства и океанографии (ВНИРО) | 1969

УДК 591.524.1

ПРИМЕНЕНИЕ ГЕОМЕТРИЧЕСКОЙ СРЕДНЕЙ ПРИ ИЗУЧЕНИИ РАСПРЕДЕЛЕНИЯ ВОДНЫХ ОРГАНИЗМОВ

К. Н. НЕСИС

В гидробиологии и ихтиологии часто приходится оперировать со средними величинами численности, биомассы или частоты встречаемости водных организмов. В таких случаях обычно применяют среднюю арифметическую величину. Но правильно ли это?

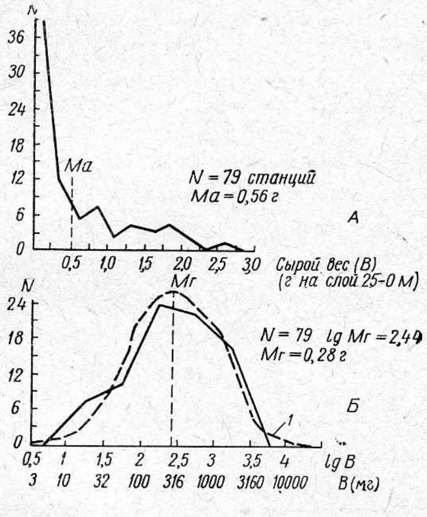

Рис. 1. Исходные данные и пример их логарифмического преобразования:
А — распределение исходных величин; *Б* — распределение логарифмов исходных величин; *1* — кривая нормального распределения; *N* — количество станций; *В* — сырая масса зоопланктона в слое 25—0 *м*, $M_а$ — средняя арифметическая, $M_г$ — средняя геометрическая.

Каким требованиям должна удовлетворять средняя величина? Во-первых, она должна быть характерна для значительной доли изучаемой совокупности, для «массы», иными словами, средняя должна быть близка к модальной величине. Во-вторых, средняя величина должна в возможно меньшей степени зависеть от случайных влияний, связанных, например, с неточностями сбора проб, с колебаниями условий среды и т. п. Средняя арифметическая удовлетворяет этим требованиям при условии нормального (гауссова) распределения данных. Если распределение существенно отличается от нормального, средняя арифметическая перестает удовлетворять этим требованиям.

При изучении численности, биомассы, частоты встречаемости и других количественных показателей распределения организмов, минимальные наблюденные значения показателей обычно равны или близки к нулю. При условии нормального распределения минимальные и максимальные значения отстоят от среднего значения приблизительно на одинаковую величину, средняя совпадает с модой, кривая симметрична. Таким образом, при нормальном распределении средние арифметические и модальные значения изучаемых показателей должны быть приблизительно равны половине максимальных значений. Каждый гидробиолог и ихтиолог по собственной практике, однако, знает, что в подавляющем большинстве случаев это не так — модальные значения обычно намного меньше половины максимальных! Как правило, кривые распределения численности, биомассы или частоты встречаемости организмов по числу станций, на которых наблюдались те или иные величины этих показателей, значительно отличаются от нормальных кривых — они асимметричны и мода сдвинута влево, в сторону малых значений (рис. 1, *А*). В таких случаях средняя арифметическая нехарактерна для основной доли изучаемой совокупности показателей и к тому же сильно зависит от величин максимальных значений или от числа проб, в которых наблюдались эти максимальные значения — а ведь такие экстремальные значения в наибольшей степени подвержены влиянию случайностей сбора. Значит средняя арифметическая для распределений, отклоняющихся от нормальных, не имеет биологического смысла и недостаточно устойчива, а потому применение ее для исследования первичных данных нецелесообразно.

Для того чтобы можно было оперировать со средней без риска впасть в ошибку, первичные данные необходимо преобразовать так, чтобы распределение приблизилось к нормальному. В очень многих случаях хорошие результаты получаются при логарифмическом преобразовании данных, т. е. если откладывать на оси абсцисс не сами величины численности, биомассы и т. п., а их логарифмы. При этом левая часть кривой распределения растягивается, а правая — сжимается и распределение приближается к нормальному (см. рис. 1, *Б*). Распределение, при котором нормально распределены не сами данные, а их логарифмы, принято называть *логарифмически нормальным*, или *логнормальным*. Средняя арифметическая из логарифмов данных будет удовлетворять требованиям, предъявляемым к средним величинам — она близка к модальной величине (см. рис. 1, *Б*) и достаточно устойчива. Например, если в приведенном на рис. 1 распределении величин сырой массы зоопланктона Ньюфаундленд-Лабрадорского района отбросим одно — максимальное значение, то средняя арифметическая из наблюдавшихся величин уменьшится на 6,1 (с 0,56 до 0,53 *г*), а средняя из логарифмов — лишь на 1,8% (с 0,28 до 0,27 *г*).

Средняя арифметическая из логарифмов данных — это логарифм средней геометрической. Средняя геометрическая определяется по формуле

$$M_{\text{г}} = \sqrt[n]{N_1 \cdot N_2 \cdot N_3 \ldots N_n},$$

где $N_1, N_2, \ldots, N_n$ — наблюдавшиеся значения изучаемой величины,

n — общее число наблюдений.

Средняя геометрическая взвешенная определяется по формуле

$$M_{\text{г. взв}} = \sqrt[n]{N_1^{m_1} \cdot N_2^{m_2} \cdot N_3^{m_3} \ldots N_{nn}^{m}},$$

где $N_1, N_2, \ldots, N_n$ — середины классовых промежутков,

$m_1, m_2, \ldots, m_n$ — масса, т. е. количество значений в каждом классовом промежутке,

$$n = \Sigma m_i.$$

Логарифмирование наблюденных значений и вычисление средней из логарифмов

$$\lg M_{\text{г. взв}} = \frac{\sum_{i=1}^{n} m_i \lg N_i}{n}$$

дает наиболее точное значение $M_г$, но при большом количестве данных такой расчет занимает слишком много времени. Для облегчения подсчетов можно воспользоваться следующим часто применяемым приемом: весь размах изменчивости изучаемой величины надо разбить на размерные классы, середины которых относятся друг к другу как $1:2:4:8:16$ и т. д., перенумеровать эти классы и вычислить среднюю арифметическую взвешенную для номеров классов. Если середины размерных классов представить как

$$a \cdot 2^k,$$

где k — номер класса, то искомая средняя геометрическая будет равна

$$a \cdot 2^M,$$

где M — средняя арифметическая взвешенная для номеров классов.

Например, для распределения численности калянуса в Лабрадорском море (рис. 2).

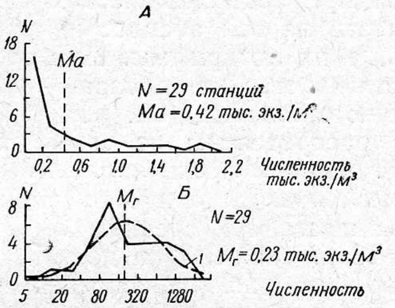

Рис. 2. Логарифмическое преобразование при помощи логарифмов с основанием 2. Распределение численности Calanus finmarchicus (*экз/м³*) в слое 200—0 *м*. Условные обозначения те же, что на рис. 1.

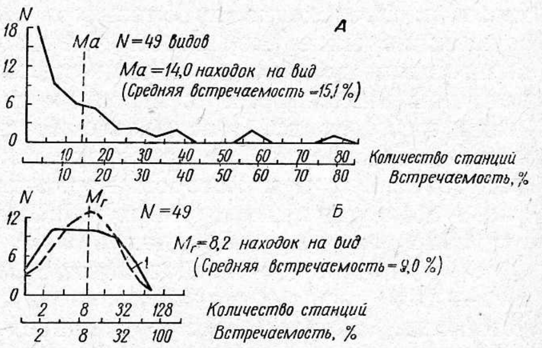

Рис. 3. Логарифмическое преобразование при помощи логарифмов с основанием 2. Распределение количества находок и частоты встречаемости (%) 49 видов иглокожих и десятиногих раков в южной, юго-восточной и восточной частях Баренцева моря, Воронке и Горле Белого моря. Данные различных рейсов 1957—1958 гг., сборы промысловым тралом и тралом Сигсби. Материалы автора. Условные обозначения те же, что на рис. 1.

Номер класса (k)	0	1	2	3	4	5	6	7	8	
Классовые промежутки	5	10	20	40	80	160	320	640	1280	2560 *экз/м³*.
Количество значений	0	1	1	4	8	4	4	4	3	

$$M = \frac{1\cdot 1 + 2\cdot 1 + 3\cdot 4 + 4\cdot 8 + 5\cdot 4 + 6\cdot 4 + 7\cdot 4 + 8\cdot 3}{1 + 1 + 4 + 8 + 4 + 4 + 4 + 3} = 4{,}93$$

$$a = \frac{5 + 10}{2} = 7{,}5 \text{ (середина классового промежутка нулевого класса)}$$

$$M_г = 7{,}5 \cdot 2^{4,93} = 230 \text{ (экз/м}^3\text{)}.$$

Значение средней геометрической, получаемое при таком методе расчета, может в результате объединения данных несколько отличаться от значения средней, полученного в результате прямого вычисления по логарифмам данных, но отклонение это невелико и лежит в пределах ошибки средней. Важно отметить, что при логнормальном распределении средняя геометрическая всегда меньше средней арифметической.

Применение средней геометрической необходимо и целесообразно только при логнормальном распределении значений численности, биомассы, частоты встречаемости и других показателей. Но, по-видимому, большинство распределений количественных показателей популяции вполне удовлетворительно аппроксимируется логнормальным распределением. Данные, приведенные на рис. 1—4, взяты без выбора из материалов лаборатории биологии моря Полярного научно-исследовательского института морского рыбного хозяйства и океанографии им. Н. М. Книповича (ПИНРО) и лаборатории нектона Института океанологии АН СССР. Можно было бы привести еще много примеров логнормального распределения подобных показателей. В частности, в предыдущей работе (Несис, 1964) было показано, что распределение биомассы бентоса Ньюфаундленд-Лабрадорского района Атлантики также логнормально. Логнормальное распределение с успехом применяли для анализа распределения численности и биомассы рыб, зоопланктона, наннопланктона, различных наземных животных (Беляева, 1964; Barnes, 1952; Cassie 1962; Preston, 1962 и др.).

Поскольку логарифм нуля равен минус бесконечности, то обычный метод вычисления средней геометрической и логарифмическое преобразование неприменимы в тех случаях, когда в изучаемой совокупности данных имеются нулевые значения численности, биомассы и т. п. Как поступать в таких случаях? Если нулевых значений немного, то каждое значение признака (численности, биомассы...) можно увеличить на небольшую постоянную величину, скажем, на 0,5 или 1 и таким образом избавиться от нулевых значений. Если нулевых значений много и они существенны для вычисления средней (например, если мы изучаем распределение одного какого-либо вида, встречающегося далеко не на всех станциях исследованного района), то такое распределение нельзя аппроксимировать логнормальным распределением. Если рассматриваемый вид животных относительно редок (скажем, встречается в количестве меньше 10 экз. в среднем на пробу) и не образует стай, то частоты его нахождений обычно бывают распределены по Пуассону: например, морские ежи, звезды и голотурии на дне Баренцева моря (Пропп, 1964) или чайки и крачки в колониях (Ивлев, Ивлева, 1960). Если вид образует стаи, скопления, то частоты его нахождений соответствуют одному из так называемых контагиозных распределений, примером которых могут служить довольно часто используемое при анализе распределения организмов *отрицательно-биномиальное* распределение или изученное Р. М. Касси (Cassie, 1962, 1963) *пуассон-логнормальное* распределение. С точки зрения биолога пуассон-логнормальное распределение можно рассматривать как результат влияния фактора случайности при сборе проб на частоты исходного логнормального распределения численностей, биомасс и т. п. Для анализа подобных распределений предложены особые методы, на которых мы здесь не останавливаемся.

Для того чтобы проверить, является ли такое распределение логнормальным, или оно существенно отличается от логнормального, можно применить *логарифмически-вероятностный* график. Для построения этого графика надо расположить все рассматриваемые значения численностей, биомасс... в ряд от наименьших к наибольшим и составить ряд кумулятивных (накопленных) частот в процентах от общего числа проб: к частоте наименьшего класса прибавляем частоту второго, затем третьего и так далее — до последнего класса, значение которого соответствует 100% проб. Кумулятивные значения частот откладываются на оси ординат графика против откладываемых на оси абсцисс значений численности, биомассы..., которым соответствуют накопленные значения частот. Масштаб по оси абсцисс — логарифмический, по оси ординат — вероятностный, представляющий собой как бы проекцию нормальной кривой на ось ординат. Вероятностная шкала применяется, например, в пробит-методе, широко используемом в биохимии, фармакологии, радиобиологии и т. д. для изучения зависимостей типа «доза — эффект», принципы и методы построения вероятностных графиков изложены в ряде курсов биологической статистики (Finney, 1952; Weber, 1957 и др.).

Точки, соответствующие данному значению признака и накопленному числу проб со значением признака, равным или меньшим данного значения, наносятся на график последовательно, от наименьших к наибольшим. Нулевые значения отбрасываются, если их можно считать несущественными, или устраняются описанным выше методом прибавления к каждому значению признака малого постоянного числа. Последнее значение, соответствующее 100% проб, не наносится, так как эта точка на вероятностной шкале уходит в бесконечность. Если изучаемое распределение соответствует логнормальному, то точки на графике расположатся вдоль прямой (рис. 4). С такого графика легко снять значения средней геометрической (на оси абсцисс против частоты 50%) и среднего квадратичного отклонения (против частот от 16 до 50% и от 50 до 84%). При пуассон - логнормальном распределении нижняя левая часть прямой будет несколько изогнута влево, при отрицательно-биномиальном распределении точки на графике расположатся вдоль изогнутой книзу кривой. Применение вероятностных и логарифмически-вероятностных графиков при изучении гидробиологических и ихтиологических данных описано в работах Касси (1954, 1962) и др.

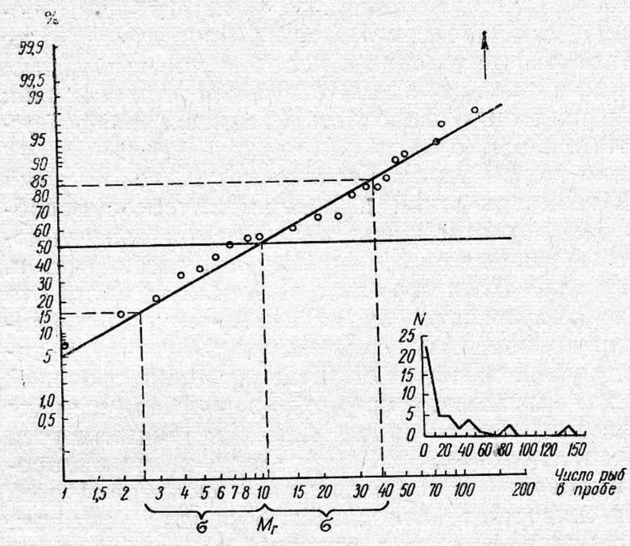

Рис. 4. Логарифмическое преобразование на вероятностно-логарифмическом графике. Распределение количества светящихся анчоусов Benthosema glaciale в северной части Атлантического океана. Э/с «П. Лебедев», рейсы 1, 2, 4, 1961, 1962, 1964 гг., трал Айзекс-Кидда, сборы Ю. Г. Чиндоновой и Н. А. Рудяковой. Материал В. Э. Беккера.

$M_г$ — средняя геометрическая; σ — среднее квадратичное отклонение; стрелкой отмечено максимальное значение количества рыб в уловах; на врезке — распределение исходных величин: N — число проб.

Применение геометрической средней при логнормальном распределении не является чем-то новым. (Andrewartha, 1961; Andrewartha, Birch, 1954; Barnes, 1952; Cassie, 1954, 1962, 1963; Preston, 1962). Этому вопросу посвящена монография Эйчисона и Броуна (Aitchison, Brown, 1957). Логарифмическое преобразование широко применяется в геологических науках (Колмогоров, 1941; Разумовский, 1940; Родионов, 1961; Ястрембский, 1932 и другие) и многих биологических науках

(Каминский, 1964; Любищев, 1958; Ушаков, Чернокожева, 1964; Hemmingsen, 1934 и др.).

Логнормальное распределение численностей, биомасс, частот встречаемости и других показателей, по-видимому, является обычным случаем, если не правилом для совокупностей водных животных. Поэтому практика логарифмического преобразования и расчета средней геометрической должна войти в арсенал математических приемов каждого гидробиолога и ихтиолога.

ЛИТЕРАТУРА

Беляева Н. В. Распределение планктонных фораминифер в водах и на дне Индийского океана. Труды Института океанологии АН СССР. Т. 68, 1964.

Ивлев В. С., Ивлева И. В. Опыт статистического анализа размещения птиц в колониях. Применение математических методов в биологии. Сб. 1, изд-во ЛГУ. Л., 1960.

Каминский Л. С. О применении геометрической средней в биологии. Применение математических методов в биологии. Сб. 3, изд-во ЛГУ, Л., 1964.

Колмогоров А. Н. О логарифмически нормальном законе распределения размеров частиц при дроблении. Доклады АН СССР, 31, № 2, 1941.

Любищев А. А. К методике количественного учета и районирования насекомых. Изд-во АН КиргССР, Фрунзе, 1958.

Несис К. Н. Степень достоверности величин биомассы по пробам дночерпателя «Океан-50». «Океанология». Т. 4, Вып. 6, 1964.

Пропп М. В. Применение статистических оценок в изучении экологии сублиторальных беспозвоночных. Труды Мурманского морского биологического института. Вып. 6 (10), 1964.

Разумовский Н. К. Характер распределения содержаний металлов в рудных месторождениях. Доклады АН СССР, 28, № 9, 1940.

Родионов Д. А. К вопросу о логарифмически нормальном распределении содержаний элементов в изверженных горных породах. «Геохимия», № 4, 1961.

Ушаков Б. П., Чернокожева Н. С. Изменение уровня теплоустойчивости мышечной ткани головастиков лягушек Rana temporaria L. в результате температурного воздействия на сперматозоиды. Сб. «Клетка и температура среды», М., «Наука», 1964.

Ястрембский Б. С. О формуле связи между средней арифметической и средней геометрической. Проблемы учета и статистики, № 1, 1932. Цитировано по Б. С. Ястрембскому. Избранные труды. М., изд-во «Статистика».

Aitchison J., Brown J. A. C. The lognormal distribution. Cambridge Univ. Press, 1957.

Andrewartha H. G. Introduction to the study of animal populations. Chicago Univ. Press, 1961.

Andrewartha H. G., Birch L. C. The distribution and abundance of animals. Chicago Univ. Press, 1954.

Barnes H. The use of transformation in marine biological statistics. J. du Cons., 18, N. 1, 1952.

Cassie R. M. Some uses of probability paper in the analysis of size frequency distribution. Austral. J. Mar. Freshw. Res. **5**, N. 3, 1954.

Cassie R. M. Frequency distribution models in the ecology of plankton and other organisms. J. Anim. Ecol., **31**, N. 1, 1962.

Cassie R. M. Microdistribution of plankton. Oceanogr. Mar. Biol. Vol. 1, 1963.

Finney D. J. Probit analysis. Cambridge Univ. Press. 1952.

Hemmingsen A. M. A statistical analysis of the differences in body size of related species. Vid. Medd. Dansk Naturhist. Foren., **98**, 1934.

Preston F. W. The canonical distribution of commonness and rarity. Ecology. **43**, No. 2, 3. 1962.

Weber E. Grundriss der biologischen Statistik für Naturwissenschaftler, Landwirte und Mediziner. 3te Aufl., Jena 1957.