Сибирский экологический журнал, 5 (2001) 631–636

УДК 57.087.1

Статическая аллометрия в случае существенно неоднородных выборок: опасность артефакта

Е. Л. ВОРОБЕЙЧИК

Институт экологии растений и животных УрО РАН
620144 Екатеринбург, ул. 8-го Марта, 202

АННОТАЦИЯ

В предположении, что выборка состоит из нескольких групп изоморфных объектов, получено соотношение, связывающее степенной коэффициент аллометрического уравнения для всей выборки и параметры групп. Степенной коэффициент – это композиция теоретического коэффициента изометрического роста и коэффициента регрессии средних групп; соотношение между вкладами этих величин зависит от структуры выборки. В случае существенно неоднородных выборок велика вероятность артефакта, когда совокупности, различающиеся по скорости изменения параметров, признаются одинаковыми, и наоборот. Предложено использовать вместо обычного степенного коэффициента коэффициент регрессии средних групп.

Более ста лет уравнение аллометрии вида $x = ax^b$ – одно из самых популярных в биологии. Первоначальная область его применения – описание изменений параметров в процессе роста организмов. В дальнейшем она была расширена вплоть до анализа размерной структуры таксонов ранга класса или типа. В труднообозримой литературе рассматриваются различные модификации аллометрической зависимости, обсуждается ее теоретический статус, связь с уравнениями роста и т. д. [1, 2]. Ряд работ посвящен анализу адекватности различных статистических процедур оценки коэффициентов уравнения [3]. Меньше внимания уделено интерпретации параметров, вычисляемых по результатам конкретных измерений. В то же время именно этот вопрос наиболее существен при практическом использовании аллометрического уравнения. Именно здесь могут быть "подводные камни", которые в силах кардинально изменить выводы всей работы.

ПОСТАНОВКА ЗАДАЧИ

По характеру получения эмпирических данных для оценки параметров уравнения различают онтогенетическую и статическую аллометрию [1]. Традиционная интерпретация коэффициентов (b – это отношение удельных скоростей изменения x и y, a – это коэффициенты формы, начальное отношение x и y) полностью справедлива в рамках онтогенетической аллометрии, когда измерению подвергается одна особь в процессе ее роста. При статической аллометрии анализируется выборка особей предположительно разного возраста. Соответственно, коэффициенты аллометрического уравнения оказываются в определенной степени "засорены" изменчивостью объектов, не связанной с возрастными изменениями. Если "засорение" очень сильное, например при существенно неоднородной выборке, истинное смещение коэффициентов может быть очень сильным. На данное обстоятельство неоднократно обращали внимание [1, 4]. Однако опасения по поводу возможной ненадежности коэффициентов уравнения в рамках схемы статической аллометрии оставались в основном в статусе эмпирического обобщения. В настоящей работе им придана аналитическая форма.

Нисколько не теряя в общности, будем излагать наш анализ аллометрического уравнения в терминах роста организмов. Нашим отправным положением служит принципиальное отличие статической аллометрии от онтогенетической: существование в выборке групп особей вместо одной особи. В данном случае под группой мы понимаем подмножество изоморфных объектов, т. е. объектов со строго одинаковым (или одинаковым в пределах ошибки измерений) характером зависимости между переменными x и y. Другими словами, внутри групп особи имеют одинаковую форму, но различаются абсолютными размерами. Поэтому в пределах группы переменные x и y связаны функциональной зависимостью, которая выводится из геометрического подобия тел. Следствие этого – разделение общих дисперсий переменных x и y на внутри- и межгрупповые компоненты.

ВЫВОД ОСНОВНОГО УРАВНЕНИЯ

Итак, пусть выборка состоит из K групп особей, в каждой из которых между x и y имеется функциональная зависимость (мы будем рассматривать ее в линейной форме после логарифмирования, оставив те же обозначения для параметров уравнения):

$$y_{ij} = a_j + b_T x_{ij}, \qquad (1)$$

где y_{ij} и x_{ij} – параметры i-й особи ($i = 1, 2, ..., n_j$) j-й группы, а j – коэффициент формы j-й группы, b_T –- теоретический коэффициент регрессии, определяемый исходя из геометрического подобия тел (коэффициент изометрического роста). Например, если x и y – линейные промеры, $b_T = 1$; если x – длина, y – объем, $b_T = 3$ и т. д.).

Наша задача сводится к отысканию степенного коэффициента аллометрического уравнения для всей выборки по известным параметрам групп. В соответствии с разложением общей дисперсии на внутри- и межгрупповую составляющие имеем:

$$b = \frac{\operatorname{cov}(x, y)}{D(x)} = \frac{\sum_{j=1}^{K} f_j \operatorname{cov}_{i=1}^{n_j}(x_{ij}, y_{ij}) + \operatorname{cov}(\overline{X}_j, \overline{Y}_j)}{\sum_{j=1}^{K} f_j \operatorname{D}_{i=1}^{n_j}(x_{ij}) + D(\overline{X}_j)}, \qquad (2)$$

где cov – ковариация, D – дисперсия, f_j – вес j-й группы, $\overline{X}_j$ и $\overline{Y}_j$ – средние арифметические параметров в j-й группе, n_j – численность j-й группы. Таким образом, в числителе записана сумма двух слагаемых – средневзвешенной ковариаций внутри групп и ковариации средних этих групп, а в знаменателе – средневзвешенной дисперсии внутри групп и дисперсии средних этих групп.

Поскольку было принято, что b_T строго одинаков для всех групп, имеем:

$$\sum_{j=1}^{K} f_j \operatorname{cov}_{i=1}^{n_j}(x_{ij}, y_{ij}) = b_T \sum_{j=1}^{K} f_j \operatorname{D}_{i=1}^{n_j}(x_{ij}). \qquad (3)$$

Введем также коэффициент, представляющий собой долю межгрупповой дисперсии в общей:

$$\gamma = \frac{D(\overline{X}_j)}{D(\overline{X}_j) + \sum_{j=1}^{K} f_j \operatorname{D}_{i=1}^{n_j}(x_{ij})}. \qquad (4)$$

Комбинируя (2), (3) и (4), получаем следующее соотношение:

$$b = (1 - \gamma) b_T + \gamma b_{\overline{Y}/\overline{X}}, \qquad (5)$$

где $b_{\overline{Y}/\overline{X}}$ – коэффициент регрессии средних групп $\overline{Y}_j$ на $\overline{X}_j$.

Выражение (5) – основной результат нашей работы. Фактически, это частный случай смещения мер связи при переходе от индивидуальных к агрегированным данным [5]. Его смысл заключается в следующем: степенной коэффициент аллометрического уравнения для всей выборки – это композиция теоретического коэффициента изометрического роста и коэффициента регрессии средних составляющих выборку групп. Соотношение между вкладами b_T и $b_{\overline{Y}/\overline{X}}$ в общий коэффициент выборки зависит от величины γ. При $\gamma = 1$, когда нет внутригрупповой дисперсии, имеет место случай онтогенетической аллометрии и отсутствует "засорение" истинного коэффициента, отражающего изменение формы в процессе роста. Чем меньше γ, тем больше истинный коэффициент $b_{\overline{Y}/\overline{X}}$ оказывается затушеван теоретическим коэффициентом b_T.

Различные модификации степенного коэффициента аллометрического уравнения не сказываются на нашем основном выводе. В частности, предлагавшийся рядом авторов коэффициент "функциональной регрессии" [6, 7] $\tilde{b} = \sqrt{D(y) / D(x)}$ преобразованиями, аналогичными описанным выше, может быть также представлен как композиция теоретического и истинного коэффициентов

$$\tilde{b} = \sqrt{(1-\gamma)b_T^2 + \gamma\tilde{b}^2_{\overline{Y}/\overline{X}}}, \tag{6}$$

где $\tilde{b}_{\overline{Y}/\overline{X}}$ – коэффициент функциональной регрессии средних.

Также принципиально не влияет на наш основной вывод смягчение условия о строгом равенстве коэффициента b_T во всех группах. Пусть в каждой группе будет свой коэффициент и они различаются не очень сильно (например, из-за погрешностей измерений). Представив $b_j = b_T + e_j$, где e_j – случайная величина с нулевым математическим ожиданием, имеем

$$b \approx (1+\gamma)b_T + \gamma b_{\overline{Y}/\overline{X}} + \\ + \frac{\mathrm{cov}(e_j, D(x_j)) + \left(\frac{1}{K}\sum_{j=1}^{K}\overline{X}_j\right)\mathrm{cov}(e_j, \overline{X}_j)}{\sum_{j=1}^{K} f_j D(x_{ij}) + D(\overline{X}_j)}. \tag{7}$$

Делая естественное допущение о некоррелированности случайной погрешности e_j с дисперсиями и средними групп, получаем равенство (5).

АНАЛИЗ ОСНОВНОГО УРАВНЕНИЯ

Основная задача использования аллометрического уравнения – определить, насколько велика разница между степенным коэффициентом аллометрического уравнения и теоретическим коэффициентом изометрического роста, т. е. сделать вывод, изменяется или нет в процессе роста форма тела. Из (5) с учетом (4) легко получить, что

$$b_{a/\overline{X}} = b_{\overline{Y}/\overline{X}} - b_T = \frac{1}{\gamma}(b - b_T), \tag{8}$$

где $b_a / \overline{X}$ – коэффициент регрессии коэффициента формы a_j на $\overline{X}_i$, т. е. истинное отличие роста от изометрического прямо пропорционально разнице $(b - b_T)$. Следовательно, степенной коэффициент выборки действительно отражает аллометричность роста. Но абсолютная величина его отличия от теоретического коэффициента зависит от γ, что может сказываться на выводе о величине и достоверности разницы $(b - b_T)$. Более того, из (8) следует, что сравнивать степенные коэффициенты двух выборок можно только при одинаковых γ (однако на практике это имеет место далеко не всегда). В противном случае можно прийти к неверным заключениям. Покажем это. Если выполняется равенство

$$\frac{b^{(2)}_{\overline{Y}/\overline{X}} - b^{(1)}_{\overline{Y}/\overline{X}}}{b_T - b^{(1)}_{\overline{Y}/\overline{X}}} = 1 - \frac{\gamma_1}{\gamma_2}, \tag{9}$$

угол между прямыми равен нулю, хотя коэффициенты регрессии средних этих выборок $b^{(1)}_{\overline{Y}/\overline{X}}$ и $b^{(2)}_{\overline{Y}/\overline{X}}$ различны. Также нетрудно найти, что при равенстве коэффициентов регрессии средних обеих выборок угол между прямыми может быть отличен от нуля, поскольку

$$tg\alpha = \\ = \frac{|\gamma_1 - \gamma_2|}{\gamma_1\gamma_2(b_T - b_{\overline{Y}/\overline{X}}) - (\gamma_1 + \gamma_2)b_T + b_T^2/(b_T - b_{\overline{Y}/\overline{X}})}.$$

Таким образом, если оперировать обычными степенными коэффициентами, можно прийти к ложным заключениям о характере изменения формы тела в процессе роста: выборки, различающиеся по данному признаку, признать одинаковыми, и наоборот.

Принципиально важно, что величина коэффициента γ не имеет никакого отношения к изменению формы тела в процессе роста. Коэффициент γ характеризует структуру выборки, по которой рассчитываются параметры аллометрического уравнения, и существенно зависит от многих параметров (соотношения масс, количества групп, объема выборки, дисперсии средних групп и т. д.). Следовательно, исследователь, по-разному формируя выборку, может произвольно (или неконтролируемо, что аналогично) изменять величину γ, а через него – и значения степенного коэффициента.

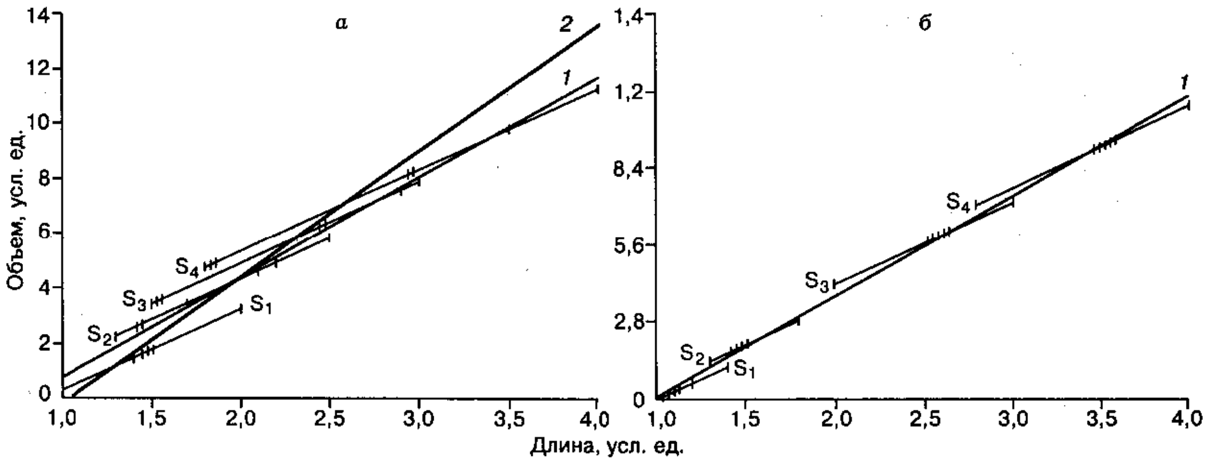
Таким образом, мы приходим к нашему основному выводу, вынесенному в заголовок статьи: использование степенного коэффициента аллометрического уравнения при анализе существенно неоднородных совокупностей таит опасность артефакта, вероятность которого тем выше, чем больше различия в структуре сравниваемых выборок. Чем менее широкий интервал значений средних, соответственно, чем больше доля в общем варьировании внутригрупповой дисперсии, тем менее надежно заключение относительно отклонения роста от изометрического. Наш вывод прямо следует из полученных соотношений (4) и (5). К аналогичному заключению, но на основе эмпирического анализа данных, приходили и другие исследователи. Например, Е. В. Балушкина [8], базируясь на изучении обширных материалов по соотношению длины и массы у личинок хирономид, делает вывод: "чем уже интервал измерений, тем более случайные значения параметров ... (аллометрического уравнения – Е. В.) могут быть получены" (с. 14).

Пример

Проиллюстрируем рассмотренную ситуацию следующим условным примером (см. рисунок). Пусть особи представляют собой вытянутые эллипсоиды вращения. Каждая из двух выборок состоит из четырех групп (S_1, S_2, S_3, S_4); каждая из групп включает по 6 особей, для которых измерены длина и объем. Средняя длина последовательно возрастает от первой группы к четвертой. Общий размах линейных размеров одинаков в обеих выборках, но различна скорость изменения формы с длиной. В выборке *а* "вытянутость" (отношение минимального промера к максимальному) распределена по группам следующим образом: S_1 = 0,35, S_2 = 0,60, S_3 = 0,80, S_4 = 1,00; в выборке *б* соответственно S_1 = 0,30, S_2 = 0,39, S_3 = 0,55, S_4 = 0,70. Сравниваемые совокупности очень сильно различаются по структуре: выборка *а* значительно более неоднородна (γ_A = 0,41, γ_B = 0,94). Рассчитанные обычным образом коэффициенты аллометрического уравнения (b = 3,65) дают одинаковые для обеих выборок прямые (прямая *1*). В то же время, коэффициент регрессии средних в выборке *а* равен 4,60, что дает существенно отличающуюся прямую (прямая *2*). Таким образом, если пользоваться обычным коэффициентом аллометрического уравнения, две выборки следует признать одинаковыми по скорости изменения формы с увеличением длины тела, тогда как они существенно различаются (в выборке *а* особи с увеличением длины меняются от вытянутого эллипсоида до шара, а в выборке *б* – лишь от более до менее вытянутого эллипсоида). Подводя итог рассмотрению данного примера, достаточно красноречиво иллюстрирующего основной вывод нашей работы, можно повторить слова Р. Мэя, сказанные по другому, но аналогичному поводу, – "это предостережение, вселяющее ужас".

ЧТО ДЕЛАТЬ?

Каков может быть выход (кроме тривиального решения – отказа от степенного коэффициента в рамках схемы статической аллометрии)? На наш взгляд – в избавлении от смеще-



Условный пример, иллюстрирующий возникновение артефакта при оперировании степенным коэффициентом аллометрического уравнения (объяснения в тексте). S_1, S_2, S_3, S_4 – группы, составляющие выборки *а* (более неоднородная) и *б*. *1* – прямая, соответствующая обычному степенному коэффициенту, *2* – коэффициенту регрессии средних (для выборки *б* прямые *1* и *2* совпадают).

ния аллометрического коэффициента. Этого можно добиться, используя коэфициент регрессии средних групп вместо обчного степенного коэффициента. При этом группу можно рассматривать как "усредненную особь" с параметрами $\overline{X}_j$ и $\overline{Y}_j$, а уравнение, связывающее $\overline{X}_j$ и $\overline{Y}_j$, – как "квазионтогенетическую" аллометрию. Осуществить разбиение выборки на группы достаточно просто: для этого необходимо построить вариационный ряд выборки с достаточно узким межклассовым интервалом по коэффициенту формы (т. е. по величине $a = y - b_T x$). Средние $\overline{X}_j$ и $\overline{Y}_j$, вычисленные для каждого такого класса, и будут формировать новую выборку, по которой должны рассчитываться параметры уравнения. При этом попутно можно получить выигрыш в статистической "обоснованности" регрессии и робастности коэффициентов уравнения. Это следует, в частности, из того, что, вне зависимости от закона статистического распределения в группе, ее средняя арифметическая всегда будет распределена нормально, если она рассчитана не менее чем по 25 наблюдениям (если закон распределения в группах не слишком сильно отличается от нормального – то даже по пяти наблюдениям) [9].

Когда аллометрическое уравнение используется для анализа размерной структуры крупных таксонов, "естественными" группами для расчета средних могут выступать таксоны меньшего (или низшего) ранга. Как следует из нашего анализа, такой путь более обоснован, чем часто практикуемый способ механического объединения всех данных в одну выборку [напр.: 10, 11].

ЗАКЛЮЧЕНИЕ

В настоящей работе мы получили соотношение, достаточно красноречиво демонстрирующее, что оперирование обычным степенным коэффициентом аллометрического уравнения в случае сильно неоднородных выборок может приводить к ошибочным выводам о скорости изменения формы тела с изменением размеров. Мы не рассматриваем, насколько широко распространены подобные ситуации в исследовательской практике. Скорее всего, они встречаются не очень часто. Но даже принципиальная возможность артефакта должна вызывать определенную настороженность при использовании степенного коэффициента. Тем большая осторожность должна присутствовать, когда содержательная интерпретация результатов работы полностью базируется на множественных сравнениях степенных коэффициентов или разница между ними мала.

При выводе уравнения (5) мы осуществили все построения для наиболее простого случая – когда теоретический коэффициент аллометрического уравнения в составляющих выборку группах однозначно выводится из геометрического подобия тел. Расширение ситуации, когда в зависимости от принятой гипотезы могут конкурировать несколько теоретических коэффициентов (например, при анализе зависимости "масса тела – основной обмен"), требует дополнительных изысканий. Также делом дальнейших работ должна быть апробация предложенного способа "борьбы" с возможными артефактами.

ЛИТЕРАТУРА

1. М. В. Мина, Г. А. Клевезаль, Рост животных, М., Наука, 1976.
2. Г. Б. Кофман, Исследования динамики роста организмов, Новосибирск, 1981, 36–55.
3. K. D. Jürgens, *Comp. Biochem. Physiol.*, 1991, **100 C:** 1–2, 287–290.
4. К. Шмидт-Ниельсен, Размеры животных: почему они так важны, М., Мир, 1987.
5. И. И. Елисеева, Статистические методы измерения связей, Л., Изд-во ЛГУ, 1982.
6. В. С. Смирнов, *Журн. общ. биол.*, 1971, **32:** 6, 693–698.
7. У. Е. Рикер, Методы оценки и интерпретации биологических показателей популяций рыб, М., Пищевая пром-сть, 1979.
8. Е. В. Балушкина, Функциональное значение личинок хирономид в континентальных водоемах, Л., Наука, Ленингр. отд-ние, 1987.
9. Р. Джессен, Методы статистических обследований, М., Финансы и статистика, 1985.
10. G. Gowing, H. F. Recher, *Austral. J. Ecology*, 1984, **9:** 1, 5–8.
11. L. E. Rogers, W. T. Hinds, R. L. Buschbom, *Ann. Entomol. Soc. Amer.*, 1976, **69:** 2, 387–389.

Statistical Allometry in the Case of Considerably Heterogeneous Samples: a Risk of Artifact

E. L. VOROBEICHIK

Assuming that a sample consists of several groups of isomorphic objects, a relation has been obtained which associates the power coefficient of the allometric equation for the whole sample and the group parameters. The power coefficient is a composition of the theoretical coefficient of isometric growth and regression coefficient of groups' means; the proportion of contributions of these quantities depends on the sample structure. In the case of considerably heterogeneous samples, the probability of an artifact, when the sets differing in the rate of change of parameters are considered as equal, and vice versa, is high. It is proposed to use, instead of the usual power coefficient, the coefficient of regression of the groups' means.