БИОЛОГИЯ МОРЯ, 2020, том 46, № 5, с. 349–356

ОРИГИНАЛЬНЫЕ СТАТЬИ

УДК 519.237.8+57.018.4

СРАВНЕНИЕ КОМПЬЮТЕРНЫХ МЕТОДОВ ДЛЯ ОПРЕДЕЛЕНИЯ ВОЗРАСТА ЖИВОТНЫХ

© 2020 г. В. В. Суханов[1, 2, *], Н. И. Селин[1]

[1]*Национальный научный центр морской биологии им. А.В. Жирмунского ДВО РАН, Владивосток 690041, Россия*

[2]*Дальневосточный федеральный университет, Владивосток 690001, Россия*

**e-mail: vsukhan@mail.ru*

Поступила в редакцию 04.07.2019 г.
После доработки 21.10.2019 г.
Принята к публикации 28.11.2019 г.

В статье обсуждаются достоинства и недостатки статистических методов, позволяющих определять возраст животных, у которых отсутствуют регистрирующие структуры. Обсуждение сопровождается примерами из материалов по росту раковины двустворчатого моллюска *Mizuhopecten yessoensis* (Jay, 1857). Возраст особей у данного вида хорошо определяется по морфологическим меткам на поверхности раковины. Это позволяет оценить процент ошибочного определения возраста при помощи той или иной статистической процедуры. Подробно протестированы следующие методы: построение гистограммы и кумуляты, дендрограммы, K-средних. Кратко описаны алгоритмы, положенные в их основу. Приведены рекомендации по применению данных методов на практике.

Ключевые слова: кластер, выделение возрастных групп, дендрограмма, метод K-средних

DOI: 10.31857/S0134347520050095

Изучение возрастной структуры популяции и роста составляющих ее особей представляет собой важную задачу, которую можно решить при помощи так называемых регистрирующих структур (Мина, Клевезаль, 1976). В этой роли выступают твердые нередко минерализованные части тела, не разлагающиеся в течение жизни животного: чешуя рыб, отолиты, зубы, кости, раковины и т.д. (Чугунова, 1959; Золотарев, 1989; Клевезаль, 2007; Клевезаль, Смирина, 2016, и др.). Годовые отметки на данных структурах, образующиеся из-за сезонной периодичности в процессе роста, позволяют определить возраст особи. Как правило, рост холодной зимой замедляется, а теплым летом ускоряется; пульсирующие изменения в скорости роста приводят к образованию годичных колец, напоминающих кольца роста на спилах деревьев.

Многие животные не имеют регистрирующих структур. Это членистоногие, у которых экзоскелет сбрасывается при каждой линьке, а также желетелые, иглокожие и отдельные мягкотелые, у которых твердые образования, хранящие следы прошлого роста, или вовсе отсутствуют, или с возрастом редуцируются. Возраст таких организмов, если они растут в природе, а не в лабораторных условиях, определить невозможно. На помощь может прийти следующее обстоятельство: в умеренных широтах процесс размножения животных в рамках года протекает в течение короткого периода времени. Как следствие, новорожденные особи компактной группой вливаются в популяцию почти одновременно. Прерывистый сезонный процесс размножения приводит к тому, что между соседними поколениями нередко существует разрыв, который обнаруживается в их распределении по размеру тела. Этот хиатус между генерациями позволяет трактовать такие размерные сгустки как одновозрастные группы. Возникает задача выделения данных групп в размерном распределении особей.

Поскольку в одновозрастных “пакетах” особи похожи друг на друга по размерам, можно использовать классификационные методы для выделения однородных групп. Сравнение этих методов и выявление среди них наилучшего – главная цель настоящей работы.

Из обсуждения исключены растения, поскольку понятие “особь” имеет для них расплывчатый смысл из-за вегетативного способа размножения.

Тестирование методов следует проводить на реальном примере. При этом мы должны знать точный возраст особей, чтобы можно было под-

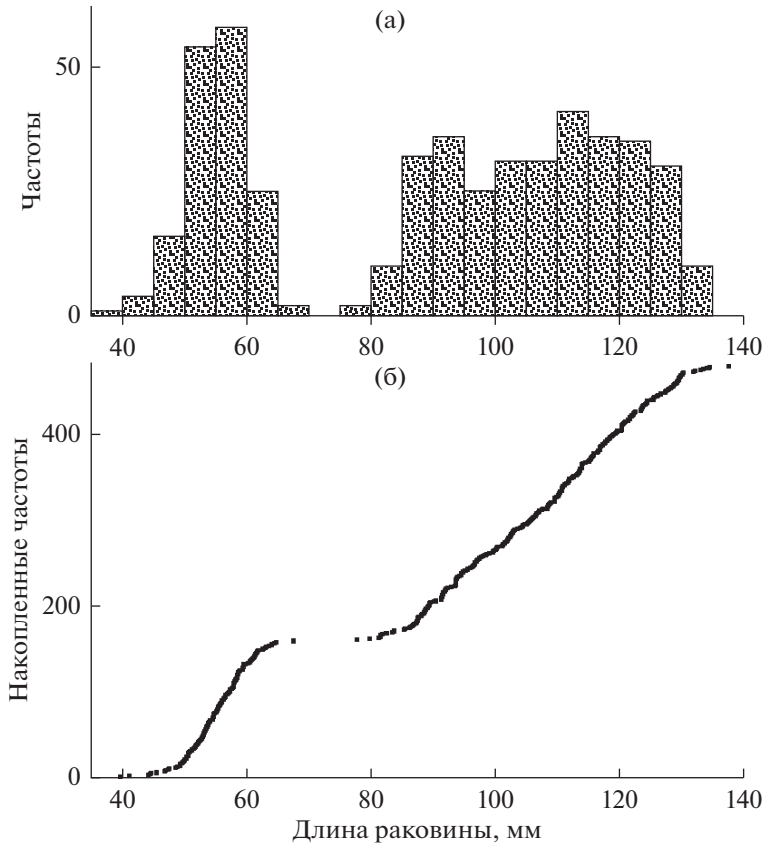

Рис. 1. Распределение числа особей *Mizuhopecten yessoensis* по длине раковины. а – гистограмма, б – кумулята.

считать процент ошибочных определений возраста при использовании конкретного метода. Можно было бы воспользоваться искусственными тестовыми данными, сгенерировав на компьютере некую воображаемую выборку с заданными свойствами. Однако у полевых зоологов эти данные вряд ли вызовут доверие, так как при компьютерном синтезе искусственной выборки может быть упущено что-то важное, но нам неизвестное. Поэтому мы воспользовались конкретным материалом, взятым непосредственно из природы.

Тест-объектом послужил двустворчатый моллюск приморский гребешок *Mizuhopecten yessoensis* (Jay, 1857) из сублиторального поселения в зал. Восток (зал. Петра Великого Японского моря). Выборка из 426 экз. была взята летом 1986 г. в течение трех дней, т.е. по сути является одномоментной. У каждой особи штангенциркулем с точностью до 0.1 мм измеряли два линейных параметра раковины – длину и высоту (см.: Скарлато, 1981). Индивидуальный возраст оценивали с помощью нескольких взаимодополняющих методов: по кольцам роста на нижней створке раковины и на лигаменте (Базикалова, 1934), а также по скульптуре верхней створки (Силина, 1978); последний метод ранее позволил оценить возрастные изменения размеров моллюсков из поселения в зал. Восток, т.е. рост (Селин, 1989). Результаты измерения длины и высоты раковины в нашем исследовании служили исходными данными для оценивания возраста разными статистическими методами.

ПРОСТО РАСПРЕДЕЛЕНИЕ

Сравнительный анализ методов начнем с построения гистограммы – графика частотного распределения особей по размеру их тела, в частности, по длине створки (рис. 1а). В идеальном случае размерные и соответствующие им возрастные группы должны образовывать компактные сгустки частот, отделенные друг от друга разрывами. На рисунке слева видна только одна такая группа, состоящая из самых мелких особей. По-видимому, это одногодовики. Правее по оси размеров едва заметны группы из более крупных особей. Но разрывы между ними выражены нечетко, а максимумы размыты, поэтому выделить и установить явные границы между возрастными когортами трудно. Очевидно, что гистограмма в примере с нашими данными неинформативна.

Существует другая форма графического представления частотных распределений. Это кумуля-

та. При ее построении все особи в выборке упорядочиваются по возрастанию величины признака. Затем формируется двухстолбцовая таблица: первый столбец – это номер особи в таком списке, второй столбец – значение самого признака (здесь – величина длины раковины). При построении графика кумуляты первый столбец обозначает “игреки”, второй – “иксы”, а сами особи изображаются точками.

Кумулята представляет собой возрастающую цепочку точек (рис. 1б). Это возрастание неравномерно. Можно увидеть ступеньку, разделяющую два участка, поднимающихся слева направо. Эта ступенька указывает на группу годовиков, также отчетливо видную на гистограмме (рис. 1а). В идеале при хорошо различимых размерных группах вся кумулята должна представлять собой лестницу, составленную из нескольких смежных ступенек. Но правый возрастающий участок здесь оказался однородным и не позволил выявить структуру у старших размерных групп. Следовательно, кумулята в нашем примере также не позволяет получить объективную информацию.

Построение гистограмм и кумулят исторически представляет собой первый способ обнаружения неоднородностей в размерных распределениях организмов. Математики называют такую задачу проблемой разделения смеси из нескольких нормальных распределений. Для этого разработаны расчетные методы и даже вспомогательные таблицы, давно вошедшие в математические энциклопедии, руководства и практику биологов (см., например: Harding, 1949; Cassie, 1954; Максимович, Погребов, 1986). Поводы для публикаций находятся и сейчас, особенно если удается найти материал для красивой цветной картинки полимодального распределения (Зуенко, 2009).

В 1980-е годы произошла компьютерная революция. За прошедшее время статистики и специалисты в области анализа данных создали немало программ для классификации объектов. Большинство этих программ имеет дружественный интерфейс, удобный и понятный для неискушенных пользователей. С их помощью можно быстрее, эффективнее и точнее решать проблему разделения смеси как для единственного признака (у нас – размер раковины) и нормальных распределений (нам все равно, какие они у *M. yessoensis*), так и для расчета статистических моментов (объема возрастных групп, средней и дисперсии по размеру).

В классической задаче разделения смеси не достигается одна из основных целей – присвоение найденного возраста каждой конкретной особи в выборке. Если бы эта цель могла быть достигнута, то возраст можно было бы привязать к другим признакам особи – измеренным, но не привлеченным к задаче возрастной классификации. Тем самым можно было бы расширить круг вопросов для исследования собранного материала.

Мы склонны считать практически бесполезными методы непосредственного изучения реальных размерных распределений. В настоящий обзор они включены лишь для полноты картины, из почтения к исторически первому подходу к решению данной группы задач. Подобное мы испытываем в музее, глядя на абак (деревянные бухгалтерские счеты) и удивляясь тому, что в течение сотен лет вычислительная поддержка цивилизации обеспечивалась таким архаичным инструментом.

НАСТРОЙКА МЕТОДОВ

Упомянутые далее методы желательно подвергнуть предварительной настройке. Во-первых, можно выбрать способ обезразмеривания признаков, характеризующих особей. В простейшем случае эти признаки измеряются в одних и тех же единицах. В нашем примере длина и высота раковины моллюска измеряются в миллиметрах. В случае полноценного морфометрического исследования список измеряемых признаков может достигать нескольких десятков. При этом размерности признаков могут быть несравнимыми. Например, длина тела измеряется в сантиметрах, масса в граммах, а зрелость гонад в баллах. Чтобы стандартизировать количественные признаки, по умолчанию используют следующий прием. Из каждого конкретного измерения вычитают среднюю арифметическую по всем особям, а результат делят на среднеквадратическое отклонение, вычисленное также по всем особям в выборке. Итоговая переменная имеет среднюю, равную нулю, и стандартное отклонение, равное единице. После использования таких обезразмеренных признаков легко вернуться к исходным размерностям, выполнив цепочку обратных шагов. Существуют и другие способы стандартизации признаков. Чтобы снизить влияние так называемых выбросов (резко выделяющихся значений признаков), можно вместо непосредственных измерений использовать их логарифмы. Если в данных есть нули, то такой прием становится невозможным. В этом случае можно использовать преобразование двойного квадратного корня типа sqrt(sqrt(...)). Если выбросов или пропущенных значений в массиве измерений слишком много, лучше изъять этих особей из анализа. Нередко такие аномалии представляют собой ошибки набора цифр при подготовке материала к компьютерной обработке.

В связи с этим важное значение может иметь погрешность прибора или метода, выбранного для измерения признаков. Пусть мы измеряем длину школьной линейкой с погрешностью 0.5 мм (половина цены деления линейки). Средняя длина

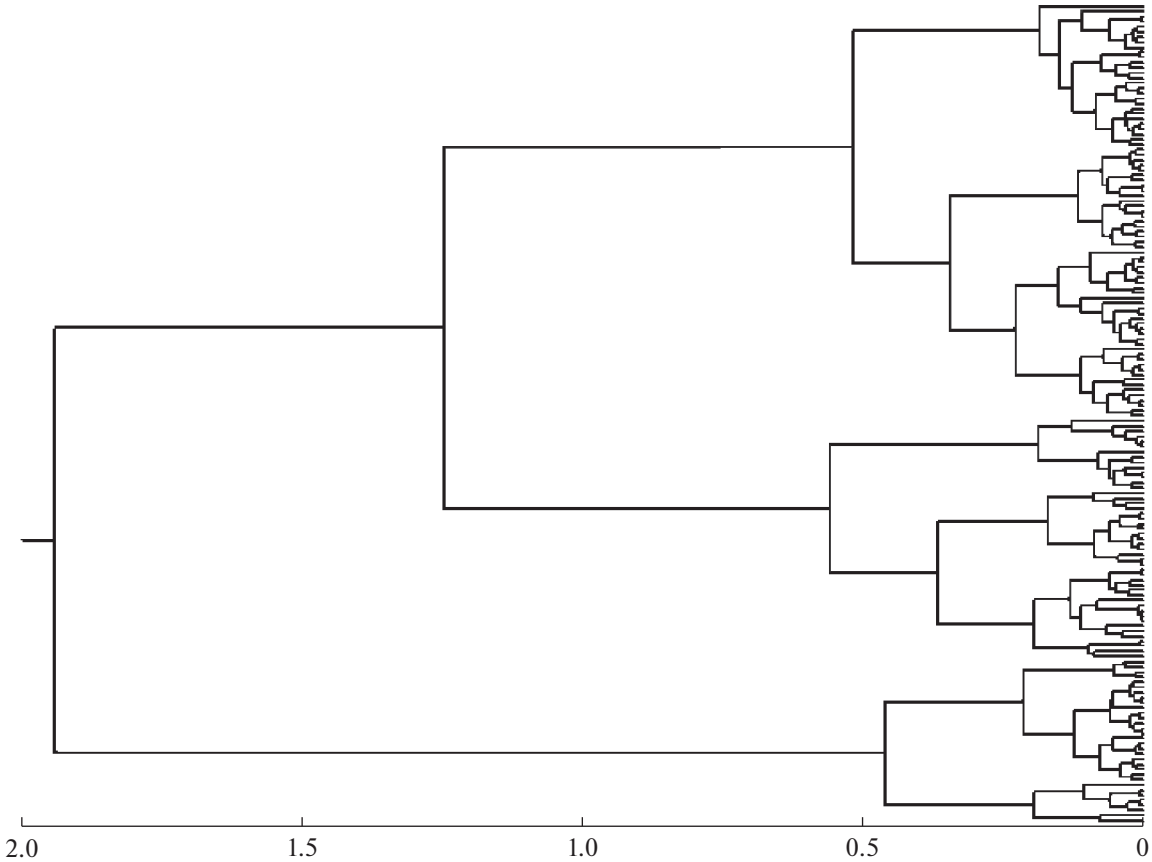

Рис. 2. Дендрограмма различия особей приморского гребешка по размерам раковины.

тела рыбы равна 10 см, а средняя, скажем, высота анального плавника составляет 5 мм. Тогда относительные погрешности измерения у этих признаков будут равны 0.5 и 10%. Второй признак, скорее всего, будет влиять на результаты классификации, уменьшая хиатусы между размерными группами, размывая их и увеличивая вероятность ошибки в определении возраста.

Во-вторых, можно выбрать метрику – способ измерения различий между особями и их групповыми объединениями. По умолчанию предлагается евклидово расстояние – обычное интуитивно понятное для всех расстояние. На прямой линии (когда признак единственный) это разница между значениями признака у двух сравниваемых особей. В двумерном случае (как в нашем примере с длиной и высотой раковины) – это расстояние на плоскости между двумя точками, изображающими сравниваемых особей. В трехмерном случае и более формула для расчета евклидова расстояния остается прежней. Существуют и другие способы измерения расстояний.

ДЕНДРОГРАММА

Метод построения дендрограмм позволяет строить древовидные изображения для описания сходства или различия как непосредственно между особями, так и между их объединениями в однородные группы. Эти группы называют кластерами (гроздьями, пучками), а процедуру их построения – иерархической кластеризацией.

Иерархия здесь появляется из-за того, что в процессе кластеризации наиболее похожие по размерам особи объединяются в первичные группы, при этом значения признаков в данных группах пересчитываются по заданному правилу. В свою очередь, кластеры первого порядка объединяются в группы второго порядка, затем третьего и т.д. Так вырастает иерархическая древовидная структура. На каком-то уровне различия или сходства найденные кластеры интерпретируются как одновозрастные группы (рис. 2).

Количество возрастных групп в выборке определяется числом кластеров, обнаруженных на заданном уровне сходства/различия. На горизонтальной шкале на рис. 2 внизу отложены значения уровня различия. Если выбрать уровень, равный 0.3, то возрастных групп (ближайших кластеров, находящихся правее этого уровня) насчитывается 8; если уровень равен 0.4, то таких групп 6; если он равен 0.5, то число групп равно 5, а при уровне 0.7 их 3. Можно воспользоваться методом “каменистой осыпи”, чтобы выбрать уро-

вень, возле которого скорость объединений кластеров достигает максимума. Но эта рекомендация не гарантирует правильности такого выбора. Объективных общепринятых методов, позволяющих выбрать нужный уровень сходства/различия, пока не найдено. Данная неопределенность в трактовке результатов снижает ценность метода дендрограмм.

Еще один серьезный дефект метода дендрограмм может сильно усложнить идентификацию особей, попавших в разные кластеры. Эти особи изображены в правой части нашего графика, где представлены конечными "листьями" в кроне построенного дерева. Каждый лист дерева сопровождается номером особи в их общем списке. По идее эти номера позволяют опознать всех особей, включенных в тот или иной кластер. Но когда объем массива исходных данных превышает несколько десятков особей, их номера располагаются очень тесно и на графике налагаются друг на друга, не давая возможности прочитать числа. Именно поэтому номера особей мы не приводим. Протокол объединений кластеров, выводимый в выходной файл с результатами, в принципе позволяет проследить путь каждой особи от исходного листочка до корня дерева. Но такая расшифровка невозможна для больших массивов данных. Можно указать еще на один недостаток метода дендрограмм: после того как два объекта объединены, они не могут быть вновь разделены. Так, Кауфман и Рауссиюв (Kaufman, Rousseeuw, 2005) отмечают, что "если дефект уже сделан, его никогда не исправить" ("once the damage is done, it can never be repaired").

В методе дендрограмм имеются и другие проблемы. Они обсуждаются в работе Кафанова с соавторами (2004). На наш взгляд, самая серьезная из них заключается в принципиальной невозможности вмешаться в работу иерархической структуры классификации. Даже если реальная иерархия в данных отсутствует, она всегда будет фигурировать в дендрограммах принудительно. Это может порождать странные и трудноуловимые артефакты.

K-СРЕДНИЕ (K-MEANS)

Данный метод включен во многие статистические компьютерные программы. Он характеризуется устойчивостью получаемых результатов, часто не требует жестких ограничений на объем обрабатываемого материала и очень прост по замыслу. В контексте нашей задачи существо метода K-средних можно описать следующим образом (Hartigan, 1975). Сначала весь массив отсортированных по возрастанию наблюдений "нарезают" на заранее предопределенное число размерных классов. Для каждого класса вычисляют среднюю арифметическую. Затем каждое наблюдение проверяют на его близость к каждой из этих подсчитанных средних. Его помещают в тот размерный класс, средняя арифметическая которого оказалась самой близкой к данному наблюдению. При этом некоторые наблюдения покидают прежние размерные классы и перемещаются в другие группы. Формируется новое содержимое размерных классов, для которых необходимо вновь пересчитывать средние арифметические. Такая процедура начинается заново, и этот процесс повторяется до тех пор, пока новые итерации не перестанут корректировать установившееся равновесное состояние. Таким образом, алгоритм метода K-средних стремится минимизировать сумму квадратов отклонений между отдельными наблюдениями внутри каждого размерного класса и между средней арифметической для этого же класса.

Здесь существуют три методические проблемы. Во-первых, не всегда понятно, с какого возраста следует начинать отсчет возрастной структуры в изучаемой коллекции. Во-вторых, почти всегда заранее неизвестно количество возрастных классов (кластеров), на которые нужно разбивать материал. В-третьих, не установлены начальные значения границ, разделяющих эти кластеры. Первая проблема может показаться надуманной. Беглый взгляд на рис. 1 приводит к правдоподобному выводу: левый изолированный кластер особей обозначает годовиков. Но по принятым здесь правилам информация о возрасте приморского гребешка полностью отсутствует и нельзя определить минимальную границу для всего диапазона возрастов. В таких случаях можно использовать грубые оценки по косвенным данным. Так, в качестве предварительной прикидки может оказаться полезной линейная регрессия вида: $\ln(D) = 1.13 + 0.80\ln(L)$, где D – среднее время генерации в популяциях данного вида (годы), L – средний размер тела (м). Эту регрессию мы рассчитали по табл. 1 из книги Боннера (Bonner, 1965). Регрессия довольно груба, но охватывает большой диапазон размеров тела организмов, равный девяти порядкам: от одного микрона для *Staphylococcus aureus* до 80 м для *Sequoya gigantea*.

По предлагаемой здесь регрессии при известном среднем размере раковин можно вычислить среднее время генерации у приморского гребешка. Оно должно быть равно 1.34 годам. В качестве грубой прикидки данная оценка вполне годится. Поскольку из-за сезонности роста возрастные группы образуют ряд целых чисел 1, 2, 3, ..., минимальная возрастная группа – это годовики.

Вторая проблема решается следующим образом. Перед запуском на счет метода K-средних число кластеров устанавливается директивно. Статистические тесты показывают, что чем большее число групп задается, тем "лучше" оказываются результаты разбиения и меньше становится

Таблица 1. Статистические показатели возрастных групп двустворчатого моллюска *Mizuhopecten yessoensis*

Показатель	Возраст				
	1	2	3	4	5
Объем	160/160	128/85	96/67	64/91	32/77
Ср. длина	55.3/55.3	93.2/89.6	110.8/101.9	121.4/113.6	132.0/125.9
Ср. высота	56.8/56.8	93.7/89.3	109.4/101.2	119.0/111.8	126.4/123.0
СрКвОткл длины	4.96/4.96	6.41/4.29	4.48/3.50	3.63/3.38	2.72/4.29
СрКвОткл высоты	4.83/4.83	5.94/3.79	4.25/3.32	3.34/3.39	2.44/4.83

Примечание. Перед чертой – по реальным данным, за чертой – по кластерам.

дисперсия в группах. Однако задание чрезмерно большого числа классов может привести к тому, что в роли реальных объективно существующих кластеров на самом деле будут фигурировать чисто случайные статистические флюктуации материала. Поэтому для правильного определения числа групп используется следующий прием. Сначала метод K-средних запускается на счет с малым числом заданных кластеров. Затем проводятся дополнительные запуски, при этом каждый раз число заданных кластеров увеличивается на единицу. Параллельно отслеживается скорость падения дисперсии в группах (рис. 3). Темп снижения дисперсии у створок раковин приморского гребешка ослабевает при числе кластеров, расположенных между 4 и 6. Этот вывод дополнительно можно проверить при помощи дендрограммы.

Третья проблема также оказалась разрешимой. Алгоритм K-средних многократно стартовал из разных случайно выбранных начальных разбиений материала по классам. Затем по статистике *F*-ratio (об этом далее) среди них находился наилучший результат. Обычно такая селекция проводится автоматически средствами самой статистической программы.

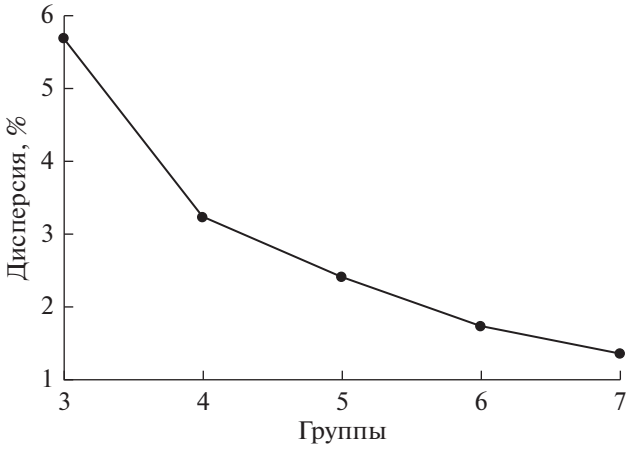

Рис. 3. Средняя дисперсия в кластерах для разного числа групп.

Подчеркнем, что метод K-средних не относится к иерархическим методам классификации (дендрограммам), хотя последние могут быть использованы здесь как вспомогательные процедуры. Иными словами, в результате работы этого метода получается разбиение, не обладающее какой-либо навязанной извне многоуровневой структурой. В нашем случае такое свойство метода K-средних можно считать его достоинством. По мнению статистиков, лучше всего метод K-средних работает, когда число наблюдений велико, а количество предполагаемых классов незначительно.

Подсчет ошибок в определении возраста моллюсков привел к результатам, которыми, очевидно, нельзя гордиться. Таких ошибок оказалось 26%. Улучшить этот результат не удалось. Мы проверили предположение о том, что использование не одного, а сразу нескольких (у нас двух) признаков может улучшить результаты классификации. Эта гипотеза подтвердилась. Когда для определения возраста привлекали лишь высоту раковины, то доля ошибочных определений достигала 32%, а когда лишь длину раковины – 40%. В процессе классификации тесно скоррелированные друг с другом признаки ослабляют влияние на результаты сильно отклоняющихся от нормы измерений.

Степень влияния каждого признака из комплекса на качество классификации можно оценить, используя критерий Фишера. С его помощью проверяется нулевая гипотеза о том, что полученное разбиение на кластеры достоверно не отличается от чисто случайного. Чем больше статистика Фишера (*F*-ratio) для данного признака, тем лучше. В нашем случае вероятность нулевой гипотезы была исчезающе малой для каждого из обоих признаков. Тем не менее для определения возраста высота раковины оказалась немного лучше, чем длина. Такой вывод подтверждают и проценты ошибок у этих признаков.

Распределение особей приморского гребешка в пространстве признаков (рис. 4) отчетливо демонстрирует замедление роста по мере взросления особей (снизу слева, вверх направо). Корре-

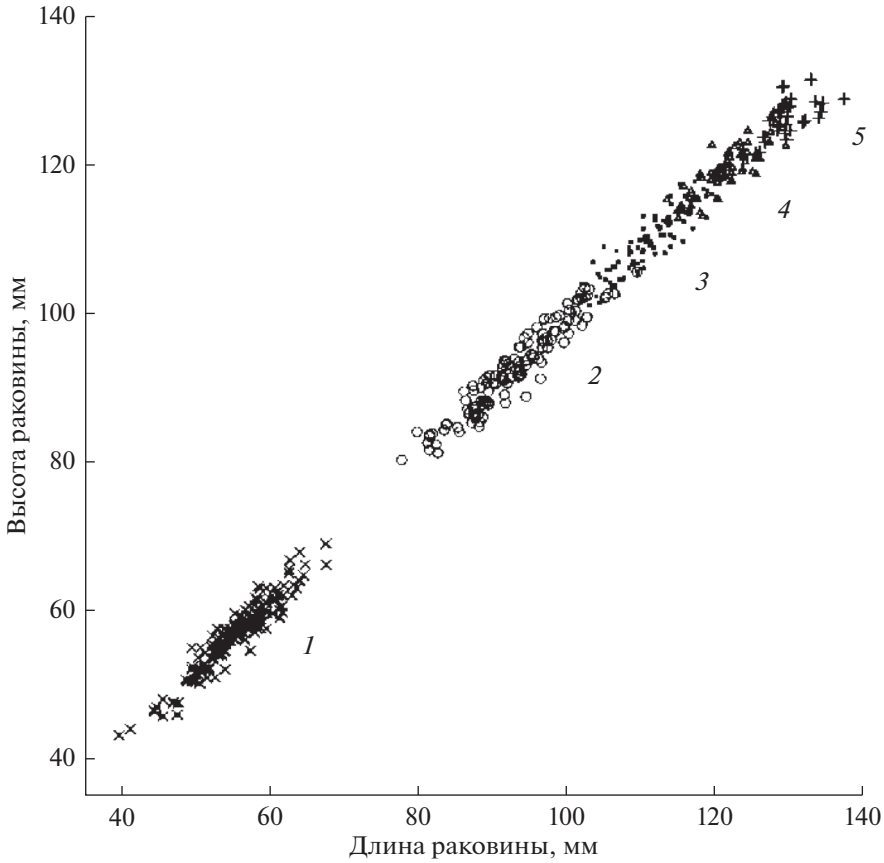

Рис. 4. Распределение особей приморского гребешка в пространстве признаков. Возраст особей (годы) указан справа от корреляционных эллипсов.

ляционные эллипсы становятся менее вытянутыми, более округлыми и все сильнее "въезжают" друг в друга. Особи из соседних возрастных групп перемешиваются особенно интенсивно как раз на стыках возрастных кластеров. Поэтому 26% ошибок в нашем случае – это вполне приемлемый результат.

Можно заметить (табл. 1), что рассчитанные по кластерам средние значения по размерам немного занижены по сравнению с реальными. При построении возрастного ключа ("размер → возраст") это может приводить к завышению оценки возраста приблизительно на полгода для старших особей. Среднеквадратические отклонения в размерах (в табл. 1 обозначены как СрКвОткл, StDev) завышены для 5-годовиков. Объемы выборок занижены для 2- и 3-годовиков, а завышены для старших возрастов. Годовики определены безупречно.

Напомним еще о двух методах кластеризации, которые иногда встречаются в статистических пакетах. Это методы медоидов и нечетких множеств.

Медоиды (Medoids). Алгоритм пытается минимизировать не сумму квадратов отклонений внутри групп, как в методе K-средних, а среднее несходство между всеми особями в группе.

Нечеткие множества (Fuzzy Sets). Здесь индивидуальный объект отнесен не к единственному, а ко всем кластерам одновременно, но с разными вероятностями принадлежности. При использовании нечетких множеств легко обнаруживаются аномальные особи, с трудом поддающиеся кластеризации: вероятности принадлежности такого объекта к разным кластерам примерно одинаковы, и с уверенностью нельзя выбрать среди них наилучший вариант. Оба метода – суть непринципиальные модификации метода K-средних. Поэтому анализ этих непопулярных методов мы опустим.

Таким образом, по удобству использования и легкости трактовки лучшую работу показал метод K-средних. Ошибки в определении возраста особей приморского гребешка – это немалые 26%. Однако, если ничего лучшего нет, это неплохой результат. Данный метод широко представлен в статистических пакетах и имеет хорошую репутацию у специалистов. По их мнению, лучше всего метод K-средних работает, когда количество наблюдений велико, а число предполагаемых клас-

сов незначительно. Именно с этим сталкиваются зоологи при попытках определения возраста у животных без регистрирующих структур.

Качество работы метода K-средних улучшается, если использовать не один, а сразу несколько количественных признаков. Чем точнее измерены признаки, тем меньше ошибок в определении возраста животного.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ НОРМ

Все применимые международные, национальные и/или институциональные принципы ухода и использования животных были соблюдены.

БЛАГОДАРНОСТИ

Авторы благодарят анонимных рецензентов за полезные критические замечания.

СПИСОК ЛИТЕРАТУРЫ

Базикалова А.Я. Возраст и темп роста *Pecten yessoensis* Jay // Изв. АН СССР. Отд. мат. и естеств. наук. 1934. № 2/3. С. 389–394.

Золотарев В.Н. Склерохронология морских двустворчатых моллюсков. Киев: Наукова думка. 1989. 112 с.

Зуенко Ю.И. Опыт использования функции нормального распределения для решения некоторых задач морской биологии // Изв. ТИНРО. 2009. Т. 156. С. 236–246.

Кафанов А.И., Борисовец Е.Э., Волвенко И.В. О применении кластерного анализа в биогеографических классификациях // Журн. общ. биол. 2004. Т. 65. № 3. С. 250–265.

Клевезаль Г.А. Принципы и методы определения возраста млекопитающих. М.: Товарищество науч. изд. КМК. 2007. 283 с.

Клевезаль Г.А., Смирина Э.М. Регистрирующие структуры наземных позвоночных. Краткая история и современное состояние исследований // Зоол. журн. 2016. Т. 95. № 8. С. 872–896.

Максимович Н.В., Погребов В.Б. Анализ количественных гидробиологических материалов: учеб. пособ. Л.: ЛГУ. 1986. 97 с.

Мина М.В., Клевезаль Г.А. Рост животных. М.: Наука. 1976. 291 с.

Селин Н.И. Распределение, структура поселений и рост приморского гребешка в заливе Восток Японского моря // Биол. моря. 1989. № 5. С. 24–29.

Силина А.В. Определение возраста и темпа роста приморского гребешка по скульптуре поверхности раковины // Биол. моря. 1978. № 5. С. 29–39.

Скарлато О.А. Двустворчатые моллюски умеренных широт западной части Тихого океана. Л.: Наука. 1981. 479 с.

Чугунова Н.И. Руководство по изучению возраста и роста рыб [Текст]. (Метод. пособ. по ихтиологии). М.: Изд-во АН СССР. 1959. 164 с.

Bonner J.T. Size and Cycle: An Essay on the Structure of Biology. Princeton, N.J.: Princeton Univ. Press. 1965. 219 p.

Cassie R.M. Some uses of probability paper in the analysis of size frequency distributions // Aust. J. Mar. Freshwater Res. 1954. V. 5. P. 513–522.

Harding J.P. The use of probability paper for the graphical analysis of polymodal frequency distributions // J. Mar. Biol. Assoc. U. K. 1949. V. 28. P. 141–153.

Hartigan J.A. Clustering Algorithms. New York: Wiley. 1975. 351 p.

Kaufman L., Rousseeuw P.J. Finding Groups in Data: An Introduction to Cluster Analysis. New York: Wiley. 2005. 342 p.

A Comparison of Computer-Based Methods for Estimating Age of Animals

V. V. Sukhanov[a, b] and N. I. Selin[a]

[a] *A.V. Zhirmunsky National Scientific Center of Marine Biology, Far Eastern Branch, Russian Academy of Sciences, Vladivostok 690041, Russia*

[b] *Far Eastern Federal University, Vladivostok 690001, Russia*

Advantages and disadvantages of the statistical methods for estimating age of animals that lack recording structures are discussed in the article. The discussion is supported by the examples from studies on growth of the bivalve mollusk *Mizuhopecten yessoensis* (Jay, 1857). Age of individuals of this species can easily be determined by morphological marks on the shell surface. This makes it possible to calculate the percentage of erroneous age estimates through one or another statistical procedure. The following methods were comprehensively tested: building a histogram, cumulative frequency curve, and dendrogram, as well as K-means. The algorithms put at the basis of these methods are described in brief. Some practical recommendations on their use are provided.

Keywords: cluster, identification of age groups, dendrogram, K-means method