

Использование многомерного анализа при изучении гибридов рыб — объектов аквакультуры

Е.В. Микодина, А.Г. Новосадов (ВНИРО, г. Москва)

Use of the multivariate analysis in studying of fishes hybrids as aquaculture objects

E.V. Mikodina, A.G. Novosadov (VNIRO, Moscow)

The opportunity of application of the multivariate analysis in studying of fishes and their hybrids is shown. Cluster analysis may be used as basis for an alternative technique on distinguishability, uniformity and stability.

Одним из важнейших этапов ихтиологических исследований является обработка полученных данных, которая в большинстве случаев сводится к статистическому анализу. Такие показатели, как средняя, среднеквадратичное отклонение, коэффициенты вариации, асимметрии и эксцесса, позволяют достаточно полно и всесторонне охарактеризовать изучаемую совокупность. В аквакультуре, особенно при исследовании перспективных для выращивания гибридов, такие исследования не проводили. Цель данной работы состояла в попытке оценки сходства/различия между двумя видами осетровых рыб (сибирский осетр ленской популяции и белуга), которые используются в аквакультуре при тепловодном садковом выращивании, и их гибридом иным методом, а именно с помощью многомерного кластерного анализа.

Статистические данные, представленные обычным (табличным) способом, весьма подробно характеризуют исходные виды и гибридную форму. Однако такая форма подачи данных не обладает достаточной наглядностью, не позволяет проводить сравнение по комплексу признаков, а также не учитывает индивидуальные свойства особей, что может привести к искажению общего представления об исследуемой генеральной совокупности. Кроме этого, в ряде случаев желательнее визуализировать полученные данные, например степень сходства гибридов рыб с исходными видами. На наш взгляд, иерархический кластерный анализ с построением дендрограммы может оказаться весьма перспективным для этих целей.

Методы многомерного анализа обычно используются в работах по систематике и таксономии рыб. Использование кластерного анализа для анализа гибридов рыб предлагали довольно давно, но до настоящего времени таких работ нами не найдено. Существует множество алгоритмов кластерного анализа, подробное описание которых можно найти в соответствующей литературе. Выбор алгоритма целиком зависит от исходных данных и целей. По нашему мнению, в данном случае для решения поставленных задач более всего подходит иерархический агломеративный анализ, который характеризуется последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров.

В начале работы алгоритма все объекты являются отдельными кластерами. На первом шаге наиболее похожие объекты объединяются в кластер. На последующих шагах объединение продолжается до тех пор, пока все объекты не будут объединены в один кластер.

Расстояние между кластерами (уровень расхождения) определялось методом одиночной связи или «ближнего соседа»

$$\rho_{\min}(S_p, S_m) = \min \rho(x_i, x_j);$$
$$x_i \in S_p;$$
$$x_j \in S_m$$

При таком методе расстояние между кластерами определяется расстоянием между двумя наиболее близкими объектами в разных кластерах. Этот метод позволяет выделять кластеры сколь угодно сложной формы, какими обычно и являются биологические группы.

В качестве меры сходства используется обычное евклидово расстояние

$$\rho_E(x_i, x_j) = \sqrt{\sum_{l=1}^k (x_{il} - x_{jl})^2},$$

где x_i, x_j — величина l -й компоненты у i -го (j -го) объекта ($l = 1, 2, 3, \dots, k; i, j = 1, 2, 3, \dots, n$).

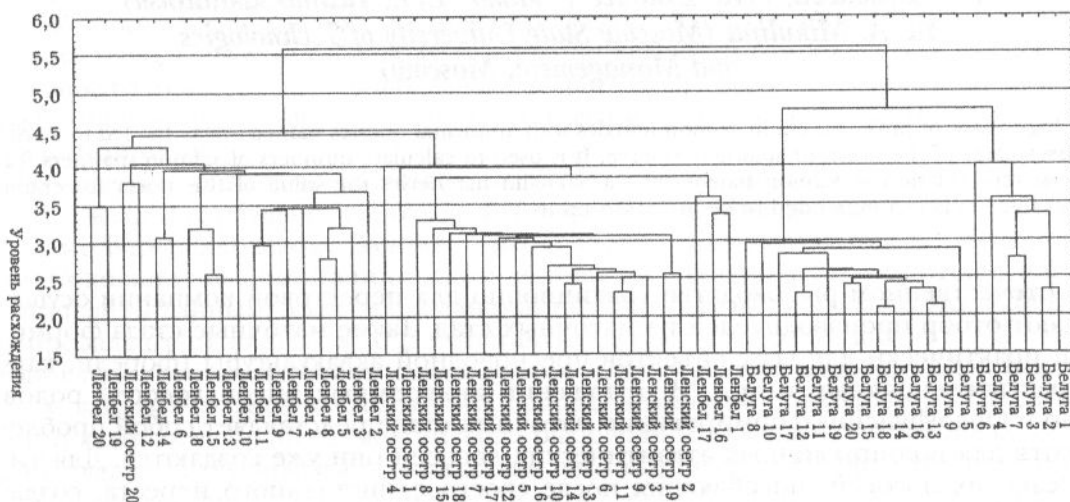
Очевидно, все показатели имеют различные единицы измерения. Чтобы избежать доминирования одних показателей над другими, прибегают к нормировке каждого признака путем деления централизованной величины на среднее квадратическое отклонение для каждого признака и переходят от матрицы X к нормированной матрице с элементами:

$$x_{il}^H = \frac{x_{il} - \bar{x}_l}{S_l},$$

где x_{il} — значение l -го признака i -го объекта; \bar{x}_l — среднее значение l -го признака; S_l — среднее квадратическое отклонение l -го признака.

Более интересную информацию о сходстве или различии можно получить, если подвергнуть кластеризации не средние показатели, а индивидуальные. На рисунке показан результат такого анализа для 20 особей каждого из исследуемых видов (ленского осетра, белуга) и их гибрида (ленбел) который несет информацию о степени сходства исследуемых объектов. Кроме этого, можно выделить еще некоторые закономерности. Так, измеренные сеголетки белуги в значительной мере консолидированы и поэтому все объединены в один кластер. Ленский осетр и ленбел также образуют два четко идентифицируемых кластера, но внутригрупповая дифференциация гибрида несколько выше, чем у исходных форм.

Дендрограмма сходства ленского осетра, белуги и их гибрида



Кроме этого, из приведенной дендрограммы следует, что отдельные особи (ленский осетр № 20, а также гибрид ленбел № 1, 16 и 17) принадлежат чужим кластерам, т. е. обладают большим сходством с другим видом, чем со своим. Анализ морфометрических данных показывает, что это сходство обусловлено в основном пластическими признаками (массой, длиной тела, высотой плавников и т.д.). Как известно, пластические признаки у осетровых рыб отличаются значительной вариабельностью и поэтому не могут являться основополагающими при идентификации видов. В данном случае гибрид достоверно отличается от ленского осетра лишь шириной перерыва нижней губы. Тем не менее, нами выявлены

дополнительные признаки, которые визуально позволяют достаточно точно отличить гибрид от исходных видов. Такими признаками являются некоторые экстерьерные показатели, например специфическая форма рта, похожая на белужью, форма головы.

Таким образом, многомерный кластерный анализ, хотя и не несет статистической информации, позволяет получить довольно точные ответы на вопросы о степени сходства гибридов и исходных видов, а также о степени внутригрупповой изменчивости гибридов и исходных видов. Многомерный анализ не является заменой статистических расчетов, а лишь позволяет привести статистические данные к более наглядному изображению.