

## Использование математических методов в биологических исследованиях школьников\*

**Мы завершаем тему занятия: «Математические методы в биологии». Биология — наука, использующая различные методы представления данных. В предыдущих номерах речь шла о математических методах. В этом номере освещаются методы многомерного анализа. Также представлены статистические таблицы и аннотация дополнительной литературы.**

Ш  
К  
О  
Л  
А  
Н  
А  
Ч  
И  
Н  
А  
Ю  
Щ  
Е  
Г  
О  
И  
С  
С  
Л  
Е  
Д  
О  
В  
А  
Т  
Е  
Л  
Я  
П  
Р  
И  
Р  
О  
Д  
Ы

**Хайтов Вадим Михайлович,**

кандидат биологических наук, заведующий лабораторией экологии морского бентоса (гидробиологии)

Санкт-Петербургского городского Дворца творчества юных, г. Санкт-Петербург

### Методы многомерного анализа

В этой главе мы коснемся задач, которые в строгом смысле не являются статистическими, так как в них не будет поставлена цель оценки генеральных параметров распределения. Вот пример такого рода задачи. Допустим, мы изучили видовой состав и показатели обилия (например, плотность поселения) видов в десяти водоемах, кроме того, мы оценили степень загрязненности берегов водоема антропогенным мусором. В каждом из водоемов видовой состав и показатели обилия видов разные, различаются и уровни загрязненности. Можно ли каким-либо способом проанализировать степень сходства и различия этих водоемов и выявить зависимость населения водоема от степени его загрязненности?

Можно, конечно, эту задачу решить простым корреляционным анализом, вычислив коэффициенты корреляции обилия каждого из видов с обилием мусора. Однако это даст ответ на вопрос о том, связано ли обилие конкретного вида с количеством мусора. В то же время нас может интересовать реакция всего населения водоема как целого. Для этого и требуются методы многомерного анализа. Для обсуждения этих методов нам сначала необходимо ввести чрезвычайно важное понятие — понятие *гиперпространства признаков*, или *n-мерного пространства признаков*.

Представим себе, что каждый изученный объект обладает всего двумя признаками. С такой ситуацией мы уже сталкивались, когда анализировали размерно-возрастную структуру популяции. Если мы построим скеттер-диаграмму, то точки, соответствующие наиболее сходным объектам, расположатся на диаграмме поблизости. Точки, соответствующие наименее похожим объектам, будут максимально удалены друг от друга. Теперь давайте представим, что мы изучили не два признака, а три. Тогда скеттер-диаграмма будет располагаться не в плоскости, а в объеме. Точки же в ней

\*  
Продолжение. Начало:  
ИРШ. 2008. № 1.  
С. 42–57; № 2. С. 24–48;  
№ 3. С. 32–34.

сформируют подобие облака (обычно так и говорят «облако точек»). Далее надо сделать очень простой ход — необходимо предположить, что мы изучили еще один признак. Таким образом, мы будем уже иметь не трехмерное пространство, а четырехмерное. Вообразить себе это невозможно, так как наш мир трехмерен, и мы адаптированы только к нему. Однако такая математическая абстракция вполне допустима. Если уж мы позволили себе работать с четырехмерным пространством, то почему бы не взять любое количество признаков — тогда мы получим  $n$ -мерное пространство. Координатами точек (объектов) в данном пространстве будут значения их признаков. Это и есть гиперпространство признаков. В нем наиболее сходные объекты (точнее — точки, им соответствующие) располагаются поблизости друг от друга, а наименее сходные — удалены<sup>1</sup>.

Выше мы говорили о сходстве объектов. Можно ли это сходство как-то выразить количественно? Этой цели служат многочисленные коэффициенты сходства или различия<sup>2</sup>. Для того, чтобы пояснить, как ими пользоваться, рассмотрим конкретный пример, который мы предложили в начале этой главы. Итак, пусть мы описали обилие видов, населяющих дно в десяти водоемах, и свели все данные в таблицу:

Таблица 28<sup>3</sup>.

**Плотность поселения донных животных в десяти водоемах, экз./м<sup>2</sup>**

Виды	Водоемы									
	1	2	3	4	5	6	7	8	9	10
1. Водяной ослик	300	220	2	0	0	12	448	114	228	1
2. Катушка роговая	2	46	1	1	0	173	85	31	34	5
3. Катушка килевая	2	26	2	1	1	31	16	24	26	0
4. Прудовик обыкновенный	1	10	0	0	0	0	0	0	0	0
5. Прудовик ушковый	22	12	0	0	0	10	11	34	88	1
6. Олигохеты	1000	123	1	10	15	188	180	200	568	0
7. Поденки Эфемеры	0	0	100	221	45	0	0	0	0	1451
8. Большая ложноконская пиявка	0	0	0	0	0	0	0	1	0	0
9. Улитковая пиявка	0	0	0	0	0	10	12	15	1	0

**1**

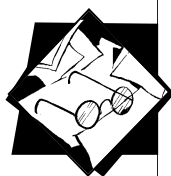
Все законы взаиморасположения точек, которые справедливы для трехмерного пространства и плоскости будут справедливы и для  $n$ -мерного пространства, поэтому далее некоторые рассуждения лучше иллюстрировать на рисунках в плоскости.

**2**

Таких коэффициентов огромное множество, можно, не сильно утруждаясь, придумать и свои собственные. Однако суть всех этих показателей одна — они, учитывая все признаки сравниваемых объектов, дают количественную меру сходства или различия между ними.

**3**

Нумерация таблиц и рисунков сквозная по всей статье с учетом начала в предыдущих номерах журнала.



В приведенном большом количестве чисел разобраться очень трудно (а ведь это пример учебный, в реальной работе количество видов может измеряться сотнями, а количество сравниваемых объектов — многими десятками), это многообразие необходимо упорядочить. Первый шаг на пути упорядочения — это вычисление коэффициентов сходства или различия.

Сразу надо оговориться, что принципиальной разницы между коэффициентами сходства или различия нет. Высокое сходство — есть низкое различие и наоборот. Поэтому выбор того или иного типа коэффициента — это дело исследователя. Однако чаще применяют коэффициенты, отражающие различия объектов.

Самый простой коэффициент различия между объектами знаком любому человеку, владеющему теоремой Пифагора. Этот коэффициент называется эвклидово расстояние (пояснения см. на рис. 9). Для вывода этого коэффициента предположим, что у двух изученных объектов мы изучили только два признака:  $X$  и  $Y$  (стало быть, мы будем рассматривать двумерный, плоский рисунок).

Как следует из теоремы Пифагора, расстояние между этими двумя точками будет вычисляться по следующей формуле:

$$R = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

Это и есть эвклидово расстояние. Поскольку законы плоскости и пространства эквивалентны, то добавление еще одного признака ( $Z$ ) принципиально ничего не изменит. Формула будет иметь такой вид:

$$R = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

Если же теперь рассмотреть общий случай, когда изучено  $n$  признаков, то формула приобретет следующий вид:

$$R = \sqrt{\sum_{i=1}^n (A_i - B_i)^2},$$

где  $A_i$  и  $B_i$  — значения  $i$  признаков объектов  $A$  и  $B$ . Чем меньше расстояние между объектами, тем выше между ними сходство. С этой величиной можно работать точно так же, как с обычным расстоянием в обычном геометрическом пространстве. Только надо помнить, что в нашем гиперпространстве признаков расстояние между точками измеряется не в метрах, а в каких-то абстрактных единицах.

Теперь вернемся к рассмотрению нашего примера с водоемами. Понятно, что объектами будут сами водоемы, а признаками — обилие видов в них. Для иллюстрации хода вычислений рассмотрим сравнение водоема № 1 и водоема № 2.

Таблица 29.

Ход вычисления эвклидова расстояния между  
водоемом № 1 и водоемом № 2

№ признака	Водоем 1 (А)	Водоем 2 (В)	$A_i - B_i$	$(A_i - B_i)^2$
1	300	220	80	6400
2	2	46	-44	1936
3	2	26	-24	576
4	1	10	-9	81
5	22	12	10	100
6	1000	123	877	769129
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
$\sum_{i=1}^n (A_i - B_i)^2$				778222
$R = \sqrt{\sum_{i=1}^n (A_i - B_i)^2}$				882

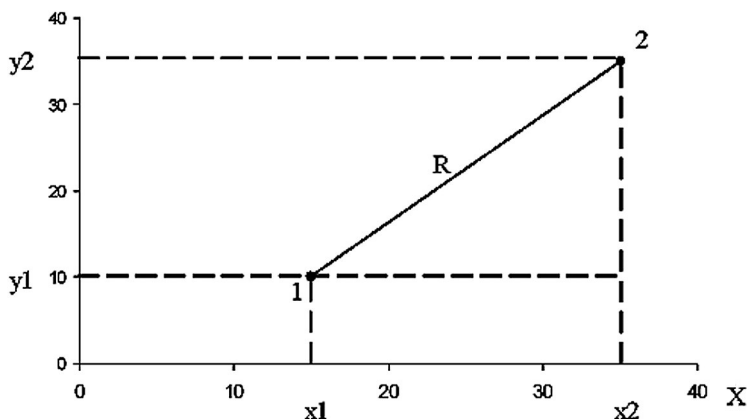


Рис. 9. Эвклидово расстояние (R) между двумя точками

Итак, эвклидово расстояние получено. Большое это расстояние или маленькое? Ответить на данный вопрос можно только тогда, когда мы сравним полученное значение расстояния с расстояниями между другими парами объектов. Самый правильный способ — это составление так называемой *матрицы расстояний*. Ниже приводится такая матрица, вычисленная для нашего примера (см. табл. 30).



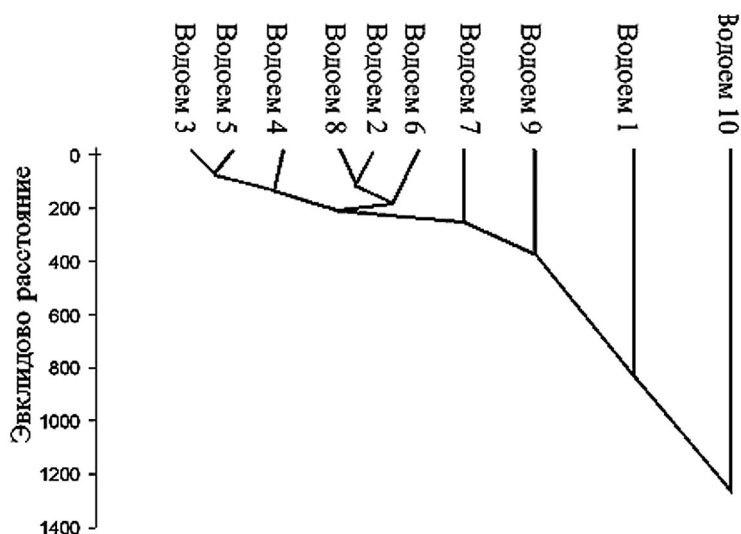
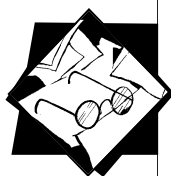


Рис. 10. Дендрограмма, отражающая степень сходства между водоемами

Таблица 30.

**Матрица эвклидовых расстояний между изученными водоемами**

	Водоем 1	Водоем 2	Водоем 3	Водоем 4	Водоем 5	Водоем 6	Водоем 7	Водоем 8	Водоем 9	Водоем 10
Водоем 1	0	882	1048	1058	1031	879	838	822	445	1788
Водоем 2	882	0	274	336	255	253	239	135	452	1473
Водоем 3	1048	274	0	121	57	275	498	255	626	1351
Водоем 4	1058	336	121	0	176	334	535	317	649	1230
Водоем 5	1031	255	57	176	0	251	488	228	608	1406
Водоем 6	879	253	275	334	251	0	445	177	465	1473
Водоем 7	838	239	498	535	488	445	0	340	456	1531
Водоем 8	822	135	255	317	228	177	340	0	389	1470
Водоем 9	445	452	626	649	608	465	456	389	0	1578
Водоем 10	1788	1473	1351	1230	1406	1473	1531	1470	1578	0

Нетрудно заметить, что эта матрица отражает расстояния во всех возможных парах сравнения. Кроме того, видно, что она симметрична относительно диагонали, где стоят нули. Последняя особенность позволяет использовать только одну ее половину (например, верхнюю). Если отбросить лишнее, то матрица принимает следующий вид (см. табл. 31).

После того, как все расстояния вычислены, можно приступить к выяснению того, насколько сходны или различны водоемы. Так, например, видно, что наиболее сходными будут водоемы № 3 и № 5, так как между ними наблюдается минимальное расстояние

( $R = 57$ ). Наименее похож на все остальные водоем № 10. Однако, если мы будем рассматривать результаты таким образом, то нам не хватит места. Для сжатого представления информации об уровне сходства между объектами необходимо провести так называемый *кластерный анализ*.

Таблица 31.

**Матрица эвклидовых расстояний между изученными водоемами, после удаления нулей и дублирующих значений расстояния**

	Водоем 1	Водоем 2	Водоем 3	Водоем 4	Водоем 5	Водоем 6	Водоем 7	Водоем 8	Водоем 9	Водоем 10
Водоем 1		882	1048	1058	1031	879	838	822	445	1788
Водоем 2			274	336	255	253	239	135	452	1473
Водоем 3				121	57	275	498	255	626	1351
Водоем 4					176	334	535	317	649	1230
Водоем 5						251	488	228	608	1406
Водоем 6							445	177	465	1473
Водоем 7								340	456	1531
Водоем 8									389	1470
Водоем 9										1578
Водоем 10										

Слово «*кластер*» происходит от английского «*cluster*» — группа, пучок, гроздь и т.п. Задача этого анализа — выявление групп сходных объектов и определение степени различия между этими группами. Существует много методов кластерного анализа, однако их процедуры достаточно трудоемки и обычно требуют вычислительной техники. Однако с принципами кластеризации можно познакомиться на самом простом методе, который называется «*методом ближайшего соседа*»<sup>4</sup>.

Для кластеризации по методу ближайшего соседа необходимо найти минимальное расстояние. В нашем примере это расстояние между водоемами № 3 и № 5. Далее заполняем следующую таблицу:

Таблица 32.

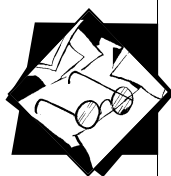
**Ход кластерного анализа**

Группируемые объекты	$R$
3–5	57

После этого находим следующее по величине значение расстояния — это расстояние между водоемами № 4 и № 3 ( $R = 121$ ). Однако водоем № 3 уже объединен в одну группу с водоемом № 5, поэтому на следующем шаге анализа мы записываем уже более длинную цепочку строящегося кластера:

**4**

К сожалению, этот метод дает наименее четкие результаты, поэтому большинство исследователей предпочитают им не пользоваться. Однако для объяснения принципа работы кластерного анализа он вполне годится. Если вы решите профессионально заниматься кластерным анализом собственного материала, то рекомендую использовать процедуру, реализованную в пакете Statistica for Windows. Опыт многих исследователей показал, что из всех методов кластерного анализа наиболее четкие результаты дает не метод ближайшего соседа, а два других — метод парно-групповой взвешенной средней и метод Варда.



Продолжение табл 32.

Группируемые объекты	$R$
3-5	57
3-5-4	121

Следующее по величине расстояние наблюдается между водоемами № 8 и № 2 ( $R = 135$ ). Эти два водоема еще не вошли ни в одну группу, поэтому они пока выделяются в отдельный кластер.

Группируемые объекты	$R$
3-5	57
3-5-4	121
8-2	135

Следующая пара — это водоемы № 5 и № 4 ( $R = 176$ ). Однако эти два водоема уже попали в один кластер при меньшем расстоянии. Поэтому данное значение пропускаем и берем следующее. Это расстояние между водоемами № 8 и № 6 ( $R = 177$ ). Соответственно, цепочка удлиняется.

Группируемые объекты	$R$
3-5	57
3-5-4	121
8-2	135
8-2-6	177

Далее мы видим, что водоем № 8 и № 5 имеют следующее по порядку значение расстояния ( $R = 228$ ). Однако эти водоемы пока относились к разным кластерам. Значит, при  $R = 228$ , эти два кластера могут быть объединены. Соответственно, можно записать следующую цепочку.

Группируемые объекты	$R$
3-5	57
3-5-4	121
8-2	135
8-2-6	177
3-5-4-8-2-6	228

Далее продолжаем аналогичные операции и получаем следующие цепочки.



Продолжение табл 32.

Группируемые объекты	R
3-5	57
3-5-4	121
8-2	135
8-2-6	177
3-5-4-8-2-6	228
3-5-4-8-2-6-7	239
3-5-4-8-2-6-7-9	389
3-5-4-8-2-6-7-9-1	822
3-5-4-8-2-6-7-9-1-10	1230

Когда все объекты попали в цепочку, процесс можно остановить. Теперь следует отразить результаты кластеризации с помощью специальной диаграммы, которая называется *дендрограммой*.

Приведенная на рис. 10 дендрограмма отражает, на каком уровне объединяются изученные объекты в группы. Однако на построении дендрограммы кластерный анализ заканчивается. Дендрограмма отражает объективный характер группировки объектов. Далее объекты надо разделить на дискретные кластеры. К сожалению, сделать это можно только достаточно субъективно.

Дендрограмма, полученная нами для данного примера, позволяет говорить о том, что отдельно от всех стоят водоемы № 10 и № 1, они формируют два отдельных кластера. В одну группу явно объединяются водоемы № 3, 4 и 5, а в другую — водоемы № 8, 2 и 6. Можно даже найти некоторую закономерность в этих группах. Так, водоем № 10 обладает наивысшей плотностью поселения поденок рода Эфемера. Водоемы № 8, 2 и 6 имеют большое количество олигохет, моллюсков, пиявок и водяных осликов. Иными словами, после группировки объектов можно искать причины их сходства. Однако это уже задача не математики, а биологии.

Вместе с тем, указанная субъективность в проведении границ между кластерами останавливает многих исследователей перед применением кластерного анализа в чистом виде. Он хорошо работает только в тех случаях, когда группы достаточно четко отличаются друг от друга. Если же все объекты формируют непрерывный ряд (как это наблюдается в нашем случае), то для этих целей более применим другой тип многомерного анализа, который называется *ординацией* или *шкалированием*.

Суть методов шкалирования достаточно проста: необходимо расположить все объекты вдоль некоторой оси, которая отражала бы сходство между ними. Самый простой способ шкалирования — это упорядочение объектов в соответствии со значениями какого-то одного признака. Например, наши водоемы можно ранжировать согласно обилию олигохет. Однако признаков у нас много, как быть в такой ситуации?







Для этого служат очень разные методы, самыми мощными из которых являются *методы многомерного шкалирования*<sup>5</sup> и разные формы *факторного анализа*. Однако суть этих методов достаточно сложна, поэтому поясним принципы ординации на самом простом методе, который получил название «*висконсинской полярной ординации*».

Давайте вновь обратимся к матрице сходства (табл. 31). Можно заметить, что среди всех пар водоемов наименее похожими являются водоемы № 1 и № 10. Между ними максимальное расстояние. Значит, они наиболее далеко отстоят друг от друга в гиперпространстве признаков. Почему они так различаются? Заглянем в исходную таблицу показателей обилий видов (табл. 28). Можно заметить, что в водоеме № 1 наиболее обильны олигохеты и водяные ослики, но полностью отсутствуют поденки. В водоеме № 10, напротив, нет олигохет, но многочисленны поденки. Столь высокая разница в населении, очевидно, определяется тем, что эти два водоема имеют какие-то принципиально разные условия. Значит, можно предположить, что эти два объекта находятся на двух противоположных полюсах некоторого градиента условий. Все остальные водоемы располагаются между этими полюсами. Как описать положение всех остальных водоемов в этом градиенте строго и объективно? Это опять чисто математическая задача, сводящаяся к умению использовать теорему косинусов.

Давайте, для примера, определим положение между означенными полюсами водоема № 2. Это можно сделать следующим образом. Представим, что в гиперпространстве признаков полюса (водоем № 1 и водоем № 10) соединены линией. Все остальные точки распределены где-то в пространстве вокруг этой линии. Расстояния от полюсов до каждой из этих точек нам известны (они приведены в матрице расстояний, табл. 31). Расстояние между полюсами составляет  $R_{1-10} = 1788$ , расстояние от водоема № 1 до водоема № 2 равно  $R_{1-2} = 882$ , а от водоема № 10 до водоема № 2 —  $R_{10-2} = 1473$ . Поскольку мы работаем с тремя точками (водоемы № 1, № 2 и № 10), мы можем провести через них плоскость и работать уже не в гиперпространстве признаков, а на плоскости (рисунок 11). Далее мы должны произвести проекцию точки, соответствующей водоему № 2, на линию, соединяющую полюса (водоем № 1 и № 10), и найти расстояние от любого из полюсов до точки проекции. Для этого надо провести очень нехитрые вычисления, суть которых поясняет рис. 11.

Итак, вычислив все необходимые величины, мы можем сказать, что проекция этого водоема ближе расположена к водоему № 1, чем к водоему № 10. Это означает, что население водоема № 2 больше похоже на население водоема № 1, чем на население водоема № 10 (если вы посмотрите на исходные данные, приведенные в таблице 28, то увидите, что это действительно так). Далее мы



5  
В компьютерных программах этот метод называется Multi Dimension Scaling (MDS).

должны провести аналогичные вычисления для всех остальных объектов. В результате можно сказать, насколько близок каждый из водоемов к тому или иному полюсу. В данном случае получились следующие результаты (см. табл. 33).

Таблица 33.

Расстояние от точек проекции водоемов до полюсов

	Расстояние от полюса до точки проекции
Водоем 10 (полюс 1)	0
Водоем 5	1004
Водоем 3	1097
Водоем 2	1150
Водоем 4	1283
Водоем 6	1285
Водоем 7	1309
Водоем 8	1353
Водоем 9	1535
Водоем 1 (полюс 2)	1788

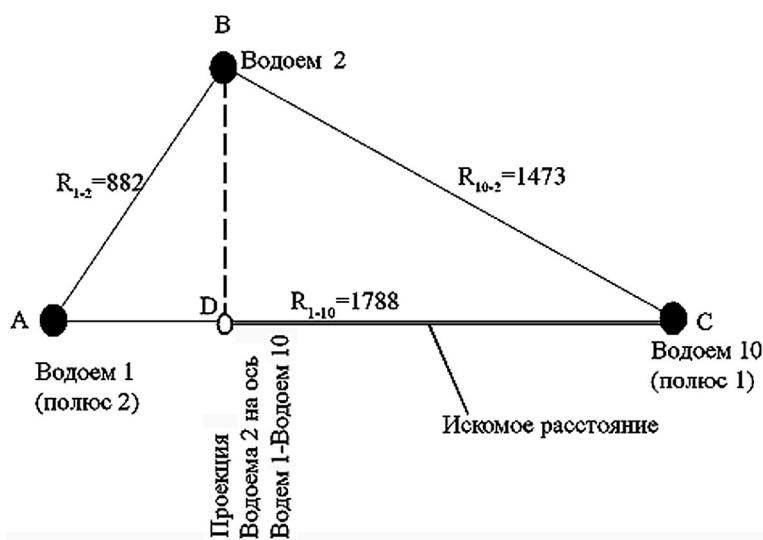
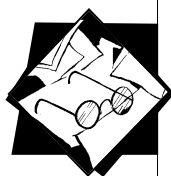


Рис. 11. Расположение трех точек относительно линии, соединяющей наименее похожие водоемы

**Пояснение.** Нам необходимо найти расстояние от одного из полюсов градиента до точки проекции. Все сводится к решению простейшей геометрической задачи.





Для привычности рассмотрим треугольник ABC. Длины сторон этого треугольника нам известны – это эвклидовы расстояния между водоемами. Нам нужно найти расстояние от одного из полюсов до точки проекции. Пусть это будет расстояние CD.

По теореме косинусов имеем:

$$\begin{aligned} AB^2 &= BC^2 + AC^2 - 2 \cdot AB \cdot AC \cdot \cos(ACB) \Rightarrow \\ 2 \cdot AB \cdot AC \cdot \cos(ACB) &= BC^2 + AC^2 - AB^2 \Rightarrow \\ \cos(ACB) &= (BC^2 + AC^2 - AB^2) / (2 \cdot AB \cdot AC). \end{aligned}$$

Однако, согласно свойствам прямоугольного треугольника BCD:

$$CD = BC \cdot \cos(DCB), \text{ но } DCB = ACB \Rightarrow CD = BC \cdot \cos(ACB) \Rightarrow \cos(ACB) = CD / BC.$$

Объединяя два выражения, получаем:

$$CD = (BC^2 + AC^2 - AB^2) / (2 \cdot AC).$$

Если вместо геометрических обозначений подставить значения эвклидовых расстояний, то можно вычислить расстояние от водоема № 10 (полюса 1) до точки проекции водоема № 2 и любого другого водоема.

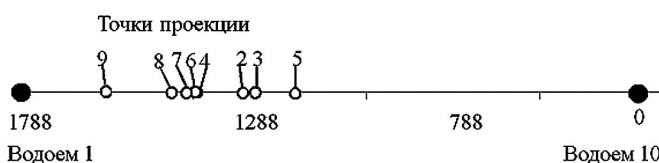


Рис. 12. Расположение точек проекции всех водоемов на линию, соединяющую два полюса

Взаиморасположение полюсов и точек проекций иллюстрирует рисунок 12. Полученные данные позволяют увидеть, что к водоему № 10 ближе всего водоемы № 5, 3, 2. Действительно, в этих водоемах водятся эфемеры и мало олигохет и водяных осликов. Значит, эти водоемы имеют сходные с водоемом № 10 условия. К водоему № 1, где много олигохет, близки водоемы № 9, 8, 7, 6 и 4, где население в чем-то похоже, там много олигохет и водяных осликов, и отсутствуют эфемеры.

Можно пойти дальше. Можно рассматривать полученные координаты как своего рода комплексные признаки, в которых в свернутом виде заключена информация обо всем населении. Далее можно проанализировать корреляцию этого признака с каким-то фактором. Например, со степенью загрязненности берегов. Пусть степень загрязненности будет следующей.

Таблица 34.

**Характеристика загрязненности берегов изученных водоемов**

Водоем	Расстояние от полюса до проекции	Степень загрязненности (количество пластиковых бутылок на 100 м береговой линии)
Водоем 1 (полюс 2)	1788	50
Водоем 2	1150	20
Водоем 3	1097	1
Водоем 4	1283	1
Водоем 5	1004	1
Водоем 6	1285	30
Водоем 7	1309	1
Водоем 8	1353	44
Водоем 9	1535	30
Водоем 10 (полюс 1)	0	0

Если вспомнить разговор о корреляционном анализе, то становится ясно, что в данной ситуации наиболее применим ранговый коэффициент Спирмена. В итоге, коэффициент корреляции оказался равен  $r = 0,834$  ( $N = 10$ ). Этот коэффициент выше порогового. Значит, между степенью загрязнения и положением водоема в выявленном градиенте есть корреляция. Стало быть, можно сделать вывод о том, что население водоема находится в зависимости от степени антропогенного загрязнения.

Завершая разговор о многомерных методах, замечу, что если объектом вашего исследования являются биоценозы, или вы занимаетесь вопросами систематики или какими-то другими исследованиями, требующими учета многих признаков, то освоение методов многомерного анализа для вас совершенно необходимо. Однако для более строгого анализа лучше использовать не описанные здесь простейшие (хотя иногда и весьма эффективные) методы, а более сложные, описание которых дается в специальной литературе, список которой приведен в части 4.



**Статистические таблицы**

Таблица 1.

**Пороговые значения t-критерия Стьюдента**

Число степеней свободы ( $\nu$ )	$t_{\text{порог}}$ при $P_{\text{дов}} = 95\%$	$t_{\text{порог}}$ при $P_{\text{дов}} = 99\%$
1	12,71	63,66
2	4,3	9,93
3	3,18	5,84
4	2,78	4,6
5	2,57	4,03
6	2,45	3,71

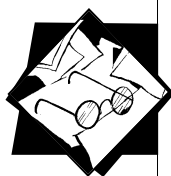


Таблица 1. (Окончание)

Число степеней свободы ( $\nu$ )	$t_{\text{порог}}$ при $P_{\text{дов}} = 95\%$	$t_{\text{порог}}$ при $P_{\text{дов}} = 99\%$
7	2,37	3,5
8	2,31	3,36
9	2,26	3,25
10	2,23	3,17
11	2,2	3,11
12	2,18	3,06
13	2,16	3,01
14	2,15	2,98
15	2,13	2,95
16	2,12	2,92
17	2,11	2,9
18	2,1	2,88
19	2,09	2,86
20	2,09	2,85
21	2,08	2,83
22	2,07	2,82
23	2,07	2,81
24	2,06	2,8
25	2,06	2,79
26	2,06	2,78
27	2,05	2,77
28	2,05	2,76
29	2,05	2,76
30	2,04	2,75
более 30	1,96	2,58

Таблица 2.

Пороговые значения критерия  $\chi^2$

Число степеней свободы ( $\nu$ )	$\chi^2$ при $P_{\text{дов}} = 95\%$	$\chi^2$ при $P_{\text{дов}} = 99\%$
1	3,84	6,63
2	5,99	9,21
3	7,81	11,34
4	9,49	13,28
5	11,07	15,09
6	12,59	16,81
7	14,07	18,48
8	15,51	20,09
9	16,92	21,67
10	18,31	23,21
11	19,68	24,72

Таблица 2. (Окончание)

Число степеней свободы ( $\nu$ )	$\chi^2$ при $P_{\text{дов}} = 95\%$	$\chi^2$ при $P_{\text{дов}} = 99\%$
12	21,03	26,22
13	22,36	27,69
14	23,68	29,14
15	25,00	30,58
16	26,30	32,00
17	27,59	33,41
18	28,87	34,81
19	30,14	36,19
20	31,41	37,57
21	32,67	38,93
22	33,92	40,29
23	35,17	41,64
24	36,42	42,98
25	37,65	44,31
26	38,89	45,64
27	40,11	46,96
28	41,34	48,28
29	42,56	49,59
30	43,77	50,89
40	55,76	63,69
50	67,5	76,15
60	79,08	88,38
70	90,53	100,42
80	101,88	112,33
90	113,14	124,12
100	124,34	135,81

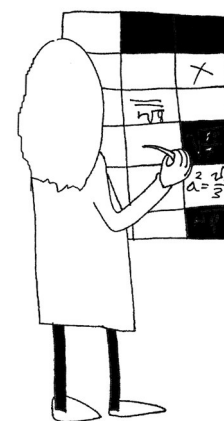


Таблица 3.

Пороговые значения коэффициентов корреляции ( $r_{\text{порог}}$ )

Число степеней свободы ( $\nu$ )	$r_{\text{s-порог}}$ при $P_{\text{дов}} = 95\%$	$r_{\text{s-порог}}$ при $P_{\text{дов}} = 99\%$	Число степеней свободы ( $\nu$ )	$r_{\text{s-порог}}$ при $P_{\text{дов}} = 95\%$	$r_{\text{s-порог}}$ при $P_{\text{дов}} = 99\%$
1	0,997	1,000	24	0,388	0,496
2	0,950	0,990	25	0,381	0,487
3	0,878	0,959	26	0,374	0,478
4	0,811	0,917	27	0,367	0,470
5	0,754	0,874	28	0,361	0,463
6	0,707	0,834	29	0,355	0,456
7	0,666	0,798	30	0,349	0,449
8	0,632	0,765	35	0,325	0,418

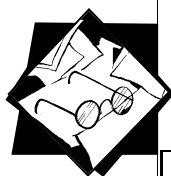


Таблица 3. (Окончание)

Число степеней свободы ( $\nu$ )	$r_{s\text{-порог}}$ при $P_{\text{дов}} = 95\%$	$r_{s\text{-порог}}$ при $P_{\text{дов}} = 99\%$	Число степеней свободы ( $\nu$ )	$r_{s\text{-порог}}$ при $P_{\text{дов}} = 95\%$	$r_{s\text{-порог}}$ при $P_{\text{дов}} = 99\%$
9	0,602	0,735	40	0,304	0,393
10	0,576	0,708	45	0,288	0,372
11	0,553	0,684	50	0,273	0,354
12	0,532	0,661	60	0,250	0,325
13	0,514	0,641	70	0,232	0,302
14	0,497	0,623	80	0,217	0,283
15	0,482	0,606	90	0,205	0,267
16	0,468	0,590	100	0,195	0,254
17	0,456	0,575	125	0,174	0,228
18	0,444	0,561	150	0,159	0,208
19	0,433	0,549	200	0,138	0,181
20	0,423	0,537	300	0,113	0,148
21	0,413	0,526	400	0,098	0,128
22	0,404	0,515	500	0,088	0,115
23	0,396	0,505	1000	0,062	0,081

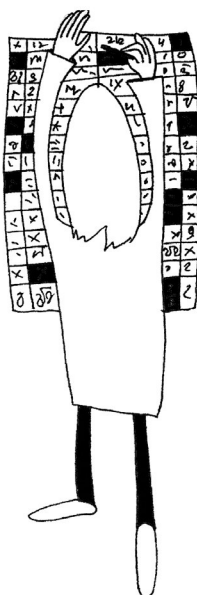


Таблица 4.

**Пороговые значения коэффициентов корреляции  
Спирмена ( $r_{s\text{-порог}}$ )**

Число степеней свободы ( $\nu$ )	$r_{s\text{-порог}}$ при $P_{\text{дов}} = 95\%$	$r_{s\text{-порог}}$ при $P_{\text{дов}} = 99\%$	Число степеней свободы ( $\nu$ )	$r_{s\text{-порог}}$ при $P_{\text{дов}} = 95\%$	$r_{s\text{-порог}}$ при $P_{\text{дов}} = 99\%$
1			24	0,389	0,502
2			25	0,382	0,493
3	0,941		26	0,375	0,484
4	0,848		27	0,368	0,476
5	0,779	0,932	28	0,362	0,468
6	0,724	0,879	29	0,356	0,460
7	0,679	0,833	30	0,350	0,453
8	0,642	0,794	35	0,325	0,422
9	0,610	0,760	40	0,305	0,396
10	0,582	0,729	45	0,288	0,375
11	0,558	0,702	50	0,274	0,356
12	0,537	0,678	60	0,250	0,327
13	0,518	0,656	70	0,232	0,303
14	0,501	0,636	80	0,217	0,284

Таблица 4. (Окончание)

Число степеней свободы ( $\nu$ )	$r_{s\text{-порог}}$ при $P_{\text{дов}} = 95\%$	$r_{s\text{-порог}}$ при $P_{\text{дов}} = 99\%$	Число степеней свободы ( $\nu$ )	$r_{s\text{-порог}}$ при $P_{\text{дов}} = 95\%$	$r_{s\text{-порог}}$ при $P_{\text{дов}} = 99\%$
15	0,485	0,617	90	0,205	0,268
16	0,471	0,600	100	0,195	0,255
17	0,458	0,585	125	0,174	0,229
18	0,446	0,570	150	0,159	0,209
19	0,435	0,557	200	0,138	0,181
20	0,424	0,545	300	0,113	0,148
21	0,415	0,533	400	0,098	0,129
22	0,406	0,522	500	0,088	0,115
23	0,397	0,511	1000	0,062	0,081

**Некоторые книги, которые рекомендуется прочитать для более глубокого освоения математических методов исследования**

Я хотел бы привести наиболее распространенные и наиболее простые для понимания (к сожалению, это не всегда совпадающие характеристики) издания, сопроводив их краткими замечаниями.

*Урбах В. Ю.* Биометрические методы. Статистическая обработка опытных данных в биологии, сельском хозяйстве и медицине. М.: Издательство «Наука», 1964.

Самое лучшее русскоязычное пособие по статистике для биологов. В книге приведены и объяснены выводы формул. Все написано простым и понятным языком. Однако обилие формул и строгих выводов может отпугнуть начинающего исследователя. Вместе с тем, эту книгу крайне рекомендуется прочитать.

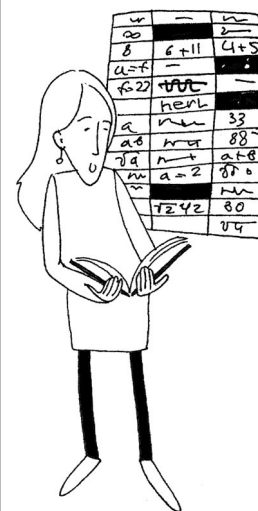
*Ивантер Э. В., Коросов А. В.* Основы биометрии. Введение в статистический анализ биологических явлений и процессов. – Петрозаводск: Издательство ПГУ, 1992.

Очень хорошее пособие, написанное просто и кратко. Опыт показал, что школьники осваивают его без особых усилий. В книге даются в доступной форме алгоритмы использования тех или иных методов. Однако это пособие, будучи изданным в Петрозаводске, практически неизвестно в других городах.

*Лакин Г. Ф.* Биометрия. М.: Издательство «Высшая школа», 1990.

*Рокицкий П. Ф.* Биологическая статистика. Минск: Издательство «Вышэйшая школа», 1967.

Эти два учебника для студентов вузов содержат все необходимые сведения по курсу биометрии. Они хороши именно как учебники для неспециалистов, для тех, кто не хочет глубоко







вникать в тонкости работы тех или иных методов, но стремится к их правильному использованию.


*Коросов А. В.* Экологические приложения компонентного анализа. Петрозаводск: Издательство ПГУ, 1996.

В книге дается самое простое описание одного из самых мощных методов многомерного анализа — метода главных компонентов. Пособие написано очень простым и понятным языком, изобилует примерами.

*Терентьев П. В., Ростова Н. С.* Практикум по биометрии. Л.: Издательство Ленинградского гос. университета, 1977.

Очень полный сборник описаний методов математического анализа биологических данных. Особенно хорошо изложен корреляционный анализ. Разбираются многочисленные примеры.

Help для пакета Statistica for Windows.

Очень хороший справочник. Его наличие делает продукт компании Statsoft, пожалуй, самым удобным орудием компьютерного статистического анализа данных. В тексте даются и ссылки на оригинальную литературу, в которой есть описание соответствующих методов. Единственный недостаток — это то, что все написано на английском языке. Однако компания Statsoft выпустила электронный учебник по статистике на русском языке. На мой взгляд, данный учебник ни в коей мере не заменяет описанных выше пособий — слишком уж он краток, написан интернет-языком (что, впрочем, многим может и понравиться). Поскольку продукт, наверняка, защищен всякими авторскими правами, то по вопросам его приобретения обращайтесь к разработчикам. 

## «Исследователь/Researcher»

Новое слово в освещении методики  
и практики организации  
исследовательской работы детей  
и юношества

<http://irsh.redu.ru>

Подписные индексы:

**39780** — полугодие,

**39781** — на год

