

ЛЕНИНГРАДСКИЙ ОРДЕНА ЛЕНИНА
И ОРДЕНА ТРУДОВОГО КРАСНОГО ЗНАМЕНИ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ имени А. А. ЖДАНОВА

В. М. ШМИДТ

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В БОТАНИКЕ

Учебное пособие

КОНТРОЛЬНЫЙ
ЭКЗЕМПЛЯР



ИЗДАТЕЛЬСТВО
ЛЕНИНГРАДСКОГО УНИВЕРСИТЕТА
1984

Г 85
1369

Печатается по постановлению
Редакционно-издательского совета
Ленинградского университета

УДК 578.087.1 : 580(075) .

Шмидт В. М. Математические методы в ботанике: Учеб. пособие. —
Л.: Изд-во Ленингр. ун-та, 1984. — 288 с.

После рассмотрения необходимого минимума общепометрических приемов анализа биологического материала в книге излагается ряд специальных методов, используемых в ботанике и других отраслях биологии (способы математического изучения роста и органогенеза у растений, таксономический анализ, приемы анализа корреляционной структуры биологических объектов, способы статистического анализа и сравнения флор, а также флористического районирования и др.).

Для студентов и аспирантов биологических факультетов университетов, педагогических и сельскохозяйственных институтов, специализирующихся в различных отраслях ботаники. Может быть рекомендовано также преподавателям, читающим в вузах курсы по применению математических методов в биологии, и научным сотрудникам.

Библиогр. 150 назв. Ил. 49. Табл. 96 + Приложение.

Рецензенты: д-р биол. наук *В. И. Василевич* (БИН АН СССР),
канд. биол. наук *Н. С. Ростова* (ЛГУ).

ИБ № 1931

Владимир Михайлович Шмидт

Математические методы в ботанике

Редактор *И. П. Дубровская*
Художественный редактор *О. Н. Советникова*
Технический редактор *А. В. Борщева*
Корректоры *Е. К. Терентьева, Н. В. Ермолаева*

Сдано в набор 26.03. 84. Подписано в печать 29.10. 84. М-30952. Формат бум. 60×90^{1/16}. Бумага тип. № 2. Гарнитура литературная. Печать высокая. Усл. печ. л. 18,0. Усл. кр.-отт. 18,25. Уч. изд. л. 17,81. Тираж 2921 экз. Заказ 344. Цена 80 коп.
Издательство ЛГУ имени А. А. Жданова, 199164, Ленинград,
университетская наб., 7/9.

Сортавальская книжная типография Государственного комитета Карельской АССР по делам издательств, полиграфии и книжной торговли. 186750 Сортавала, Карельская, 42.

Ш 2001040000—193
076(02)—84 120—84

© Издательство
Ленинградского
университета,
1984 г.

69/3 85
ГНТБ СО АН СССР
Гос. публ. библиот.-техн.
библиот. в

E5

ПРЕДИСЛОВИЕ

В первой половине 60-х годов автору не раз доводилось выступать в довольно редкой для того времени роли пропагандиста математических методов в ботанике. За прошедшее время положение изменилось настолько, что отсутствие математической обработки данных все чаще воспринимается как недостаток той или иной ботанической работы. Процесс внедрения математических методов в биологию, в частности в ботанику, идет полным ходом.

Необходимость рекламировать биометрию отпала, зато представляется своевременным предостеречь от некоторых крайностей в ее применении. Иногда в лавине цифр и формул теряется биологическая суть вопроса. В других случаях без достаточных на то оснований используются наиболее сложные и трудоемкие математические методы, тогда как задача может быть решена значительно проще и экономичнее. Такие работы, как правило, остаются малопонятными широкому кругу исследователей и вопреки добрым намерениям авторов не содействуют росту авторитета математических методов в биологии.

Последовательность действий биометрика всегда складывается из следующих трех этапов. Вначале исходные биологические данные переводятся на язык математических символов и формул. Этот процесс кодирования биологической информации позволяет сконцентрировать и формализовать ее до такой степени, что на втором этапе оказывается возможным пустить в ход математический механизм ее обработки и получить на выходе некоторый формальный результат. Биологическая интерпретация этого математического результата составляет третий, заключительный, этап работы.

Из сказанного видно, что математические методы — не самоцель, а только мощное средство познания биологических объектов и явлений. Ясно также, что сложность избираемого ма-

тематического приема должна в разумной мере соответствовать трудности биологической задачи и степени точности исходных данных. Если последняя невысока (а это нередко бывает в биологии), то нет смысла стремиться к использованию самых сложных и трудоемких математических методов. Короче говоря, в каждой конкретной ситуации надо уметь выбрать адекватный и наиболее рациональный математический прием, гарантирующий успешное решение биологической задачи с наименьшей затратой труда и времени.

Главная цель данной книги и состоит в том, чтобы помочь студентам, а также аспирантам и научным работникам сознательно ориентироваться в основных математических методах и приемах, используемых современной биологией.

В основу книги положен специальный курс лекций, читаемый автором на кафедре ботаники Ленинградского университета с 1961 г. Он соответствует курсу «Математические методы в биологии», введенному в учебные планы биологических факультетов во всех университетах СССР. Несмотря на то, что этот курс построен автором на ботанических примерах и включает в себя ряд специфических для ботаники приемов, его содержание представляет непосредственный интерес для биологов любой специальности.

Общая часть книги (главы 1—5) содержит сжатое изложение основ биометрии. Ни одним образом не подменяя более полные руководства по биометрии, эта часть знакомит читателя лишь с теми общими методами, которые имеют самостоятельное значение или готовят его к сознательному восприятию специальной части пособия.

Специальная часть (главы 6—8) посвящена математическим методам изучения роста и органогенеза растений, применению их в систематике и сравнительной флористике. Описываемые в этих главах методы впервые собраны в рамках одного учебного пособия. Между тем именно эти методы, разбросанные по страницам журналов и малодоступных изданий, как раз и представляют для биологов наибольший интерес. Надо оговорить, что в книгу включены только наиболее важные и надежно зарекомендовавшие себя на практике методы. Специфические для геоботаники математические методы в книгу не включены, так как их можно найти в ряде руководств (например, [Грейг-Смит, 1967; Василевич, 1969, и др.]).

В биометрии еще не выработались унифицированная терминология и символика. Мы будем придерживаться тех терминов и буквенных обозначений, которые использовались П. В. Терентьевым, наиболее привычны автору и почти совпадают с принятыми в «Биометрии» Н. А. Плохинского [1970]. Читатели, достаточно знакомые с биометрией, могут обращаться сразу к второй части книги.

Недостатком имеющихся пособий по биометрии является то, что среди цифр и формул остается мало места для объяснения биологической сути дела. Поэтому одной из важных задач было восполнить указанный пробел и тем самым сделать книгу более доступной для широкого круга биологов.

Автор сознает, что его попытка создать учебное пособие несколько иного типа, чем общепринятые, не может быть свободна от недостатков, и заранее благодарен тем, кто пришлет ему критические замечания по адресу: 199164, Ленинград, Университетская набережная, 7/9, кафедра ботаники Ленинградского университета.

ОБЩАЯ ЧАСТЬ

Глава I

ИЗУЧЕНИЕ ИЗМЕНЧИВОСТИ ПРИЗНАКОВ

1.1. СУЩНОСТЬ БИОМЕТРИЧЕСКОГО МЕТОДА. ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ И ВЫБОРКА

Сущность биометрического метода вытекает из специфических особенностей, отличающих его от чисто биологических приемов изучения материала.

Первой такой особенностью является использование выборочных приемов наблюдения и особой техники обработки данных, применяемых в математической статистике.

Биометрическому анализу подвергаются чаще всего не одиночные особи, а статистические выборки, составленные по определенным правилам из некоего множества экземпляров генеральной совокупности. Объем генеральной совокупности зависит от биологической задачи исследования. Например, если ставится цель изучить изменчивость признаков данного вида, то под генеральной совокупностью следует понимать все множество особей, произрастающих в пределах его ареала. Если же биологической задачей исследования является изучение географической или экологической дифференциации вида, то популяции, произрастающие в разных частях ареала или в разных экологических условиях, составят каждая свою генеральную совокупность.

Сколько же экземпляров надо исследовать или сколько наблюдений произвести (иными словами, каков должен быть объем выборки), чтобы получить надежное представление о характере варьирования признака в генеральной совокупности? Ответ на этот вопрос в математической форме требует хотя бы ориентировочного знания меры изменчивости признака, которым мы далеко не всегда обладаем, но которое можем обрести путем «прикидки» некоторыми упрощенными способами (см. с. 46). С другой стороны, громадный опыт, накопленный биометрией, позволяет дать на этот счет некоторые практические рекомендации, которым мы считаем полезным предпослать следующее рассуждение.

Если бы все особи по своим признакам были тождественны, то для получения 100%-ной информации о генеральной совокупности достаточно было бы изучить один экземпляр. Но, как известно, в природе не существует даже двух абсолютно одинаковых особей данного вида. Это означает, что каждый экземпляр несет в себе лишь часть информации о генеральной совокупности. Эта часть, как правило, достаточно велика для определения его видовой принадлежности, но дает очень приблизительное представление о среднем значении признаков в генеральной совокупности и совершенно не отражает размаха их изменчивости.

Изучение второго экземпляра дает первое (но очень неточное!) представление о пределах изменчивости признаков. Привлечение в выборку последующих экземпляров поначалу существенно корректирует эти пределы, однако по мере того, как естественное разнообразие значений признаков постепенно исчерпывается, приток новой информации начинает асимптотически затухать. Теоретически этот процесс прекращается лишь с анализом последнего экземпляра генеральной совокупности.

Выход из положения подсказывает сам асимптотический характер затухания притока новой информации, в ходе которого неизбежно наступает такой момент, когда количество информации, привносимое каждым очередным экземпляром, становится настолько незначительным, что дальнейшее увеличение объема выборки делается нерациональным: расход труда и времени перестает оправдываться сколько-нибудь существенным повышением точности результата. Этот момент и определяет оптимальный объем выборки, но вся сложность заключается в том, что если в распоряжении исследователя нет никаких предварительных биометрических данных о генеральной совокупности (например, когда объект исследуется впервые), то уловить этот момент практически очень трудно.

Опыт показывает, что в подавляющем большинстве случаев достаточно точные результаты получаются при объеме выборок $n \approx 100$, но надо иметь в виду, что оптимальный объем выборки пропорционален степени изменчивости признака. Поэтому если заметна сильная изменчивость, то объем выборки желательно увеличить, если же признак варьирует слабо, то удовлетворительный результат может быть достигнут и при меньшем ее объеме (например, при $n = 20-50$). Во всяком случае имеется возможность определить точность полученного значения средней арифметической и тем самым проверить достаточность объема выборки (раздел 1.4.7).

Из сказанного ясно, что задачей выборочного исследования является получение достоверной информации о генеральной совокупности посредством изучения относительно небольшого количества особей, включенных в данную выборку. Ясно также, что каков бы ни был объем выборки, она не может дать нам

100% -ную информацию о генеральной совокупности. Более того, если изучить несколько выборок из одной и той же генеральной совокупности, то их параметры (среднее арифметическое, среднее квадратичное отклонение, коэффициент вариации) будут несколько отличаться друг от друга. Вместе с тем, если выборки были составлены правильно, то эти различия не могут быть слишком велики и каждое из них не может сильно отклоняться от истинного значения того же параметра в генеральной совокупности. Короче говоря, выборочные параметры могут служить приближенными оценками генеральных параметров.

Вторая специфическая особенность биометрического метода состоит в том, что с помощью теории вероятностей результаты анализа одной выборки могут быть распространены на всю генеральную совокупность с точно известной степенью достоверности. При этом мы определяем не сами параметры генеральной совокупности, а достаточно узкие пределы, в которых они заключаются и в которые попадают параметры всех других выборок из данной совокупности, сколько бы мы их ни делали.

Необходимость вероятностного подхода логически вытекает из самого характера изменчивости признаков у живых организмов. Известно, что проявление признака в фенотипе осуществляется не жестко (однозначно), а в пределах нормы реакции генотипа на условия среды, т. е. многозначно, в зависимости от случайного сочетания множества факторов, полный учет которых практически невозможен. Каждый наблюдаемый признак теоретически может быть описан многозначной функцией типа $y=f(x, z, \dots)$, причем точки, соответствующие значениям признака у разных особей, не лежат непосредственно на кривой, а образуют вокруг нее облако рассеивания. Изучение же меры рассеивания случайных величин входит в компетенцию теории вероятностей.

Итак, сущность биометрического метода состоит в том, что, изучив приемами математической статистики сравнительно небольшую по объему выборку из генеральной совокупности, мы можем с помощью теории вероятностей судить о значениях и характере варьирования признаков во всей генеральной совокупности.

Сравнивая разные выборки, мы можем говорить о степени их сходства или различия, о принадлежности их к одной и той же или к разным генеральным совокупностям (популяциям, видам и т. д.). При этом степень достоверности выводов может быть задана наперед, но обычно она определяется впоследствии. Биолог может быть удовлетворен, если вероятность ошибиться не превышает 5% (5 шансов из 100 за то, что вывод неверен), но он может ограничить себя в особо ответственных случаях и более высоким, например 1%-ным или 0,1%-ным, доверительным уровнем (при этом шансы ошибиться составят соответственно 1 из 100 и 1 из 1000).

Вместе с тем необходимо решительно подчеркнуть то, что, несмотря на свою специфику и обусловленную ею известную самостоятельность, биометрический метод немыслим в отрыве от чисто биологических методов. Он не только опирается на них, но на первых этапах исследования представляет собой математическую обработку данных, поставляемых морфологическим, анатомическим, географическим, эколого-биологическим и другими методами ботаники.

Дело в том, что математический аппарат сам по себе сугубо формален, и успех его применения полностью зависит от биологической доброкачественности исходных данных.¹ Следовательно, приступая к биометрическому анализу, необходимо обладать предварительным биологическим знанием объекта в степени, достаточной для обеспечения доброкачественности исходных данных, а сама схема биометрического анализа должна соответствовать поставленной биологом задаче.

Может показаться, что сказанное низводит биометрический метод до роли сугубо вспомогательного способа математической обработки данных, поставляемых классическими методами ботаники. На самом деле это далеко не так.

Биометрический метод в силу своей известной самостоятельности способен вскрывать новые биологические закономерности, ставить и разрешать такие проблемы, которые не могли быть изучены или вопрос о существовании которых не ставился классическими методами.

Биометрический метод способствует наиболее объективному разрешению ряда основных задач динамической морфологии, систематики, филогении, сравнительной флористики и ботанической географии. Решение большинства задач начинается с составления статистических выборок из соответствующих цели исследования генеральных совокупностей и биометрического изучения отдельных признаков в полученных выборках.

1.2. ПРАВИЛА СОСТАВЛЕНИЯ ВЫБОРОК

Для того чтобы выборочные параметры могли служить приближенными оценками параметров генеральной совокупности, выборка должна быть *репрезентативной* и *рэндомизированной*. Первое требование означает, что при составлении выборки необходимо принять специальные меры к обеспечению встречае-

¹ Здесь уместно привести слова выдающегося представителя прикладной математики академика А. Н. Крылова «Подобно тому, как для получения хорошей муки надо засыпать хорошее зерно, так и в приложениях математики к решению практических вопросов надо прежде всего заботиться, чтобы те приложения и допущения, которые делаются для придания практическому вопросу математической формы, соответствовали действительности в мере практической потребности» (газета «За социалистическую науку», № 9, 7 нояб. 1932 г.).

мости в ней по возможности всех реально существующих в природе значений признаков и притом в той пропорции, которая свойственна генеральной совокупности. Второе требование предостерегает от проявления какой-либо тенденциозности при составлении выборки.

Попадание в выборку растений с разнообразными значениями признаков обеспечивается достаточностью ее объема. Однако никакой объем выборки сам по себе не гарантирует соблюдения того соотношения различных значений признаков, которое реально существует в генеральной совокупности. Поэтому далеко не безразлично, каким путем производится отбор особей в выборку, какие именно экземпляры включаются в нее. Если отбор в каком-либо отношении и в какой-либо степени тенденциозен, то происходит направленное искажение выборочных параметров, и представление о средних значениях и изменчивости признаков изучаемого объекта может оказаться не соответствующим действительности.

Репрезентативность выборки достигается соблюдением определенных правил ее составления. В зависимости от задач исследования и условий сбора материала (или условий опыта) различают три основных типа отбора, которым соответствуют три типа выборок.

1. Случайный отбор. Отбор и выборка называются случайными, если все особи в момент отбора имеют равные шансы попасть в выборку. Следует рекомендовать бесповторный случайный отбор, когда отобранные экземпляры не возвращаются в генеральную совокупность.

Получить случайную выборку не так просто, как это кажется на первый взгляд. Дело в том, что никакой глазомерный способ отбора особей наудачу не может ликвидировать возможность смещения выборки по причинам, зависящим от личных качеств отбирающего.¹ Человек — весьма несовершенное орудие случайного отбора. В психике каждого из нас, даже при кажущейся беспристрастности, заложено подсознательное предпочтение определенного облика объекта, а значит, и стремление к тенденциозному подбору особей. Таким образом, как это ни парадоксально, случайный отбор должен вестись по определенным правилам, исключающим проявление предвзятости.

Простейший прием такого отбора состоит в том, что в выборку включают каждую 5-ю, 10-ю (или иную по счету) особь по выбранной наугад линии или натянув предварительно шнур. Этот способ при сборе массового материала обычно дает вполне удовлетворительные результаты, однако надо следить за тем, чтобы синхронно с номерами отбираемых экземпляров не повторялось бы какое-либо внешнее условие, влияющее на облик

¹ Другие возможные причины смещения выборки рассматриваются в следующем разделе.

растений. Как в природных условиях, так и на опытных делянках может, например, получиться, что все отбираемые особи выросли на периферии популяции, на краю поля или грядки, где условия влажного снабжения или освещения несколько иные, чем в центре. Это обстоятельство, если вовремя его не заметить, способно привести к смещению выборки. Исключить синхронное влияние условий среды можно посредством жеребьевки предварительно пронумерованных растений.

Более строгим способом случайного отбора следует считать использование таблицы случайных чисел (Приложение I), полученных на особой рандомизирующей машине. Любая цифра в каждом четырехзначном числе этой таблицы может считаться совершенно случайной. Пользоваться таблицей очень легко. Если, допустим, надо произвести случайную выборку 50 экз., то, начав с любой колонки таблицы и двигаясь по столбцам сверху вниз или, наоборот, учитывая первые или последние две цифры четырехзначного числа, выписывают числа, не превышающие 50. Это и будут номера растений, подлежащих включению в выборку.

2. Направленный (типический) отбор. На практике встречаются ситуации, в которых с биологической точки зрения случайная выборка не соответствует условиям сбора материала или самой задаче исследования.

Первое может произойти в обстоятельствах, когда биолог вынужден довольствоваться малым объемом ($n \leq 20$) выборки: именно в силу слепого случая в нее могут попасть экземпляры с преимущественно крайними значениями признаков, а это вызывает смещение выборочных параметров относительно генеральных. Разумеется, с увеличением объема выборки вероятность такого случая быстро уменьшается.

Несоответствие случайного отбора задаче исследования встречается гораздо чаще. Дело в том, что при сборе материала из природной популяции нередко приходится сталкиваться с ее явной неоднородностью. Часть растений популяции может вырасти на открытом месте, а часть — под пологом кустарника или леса. Расположение растений на разных уровнях склона, на склонах разной экспозиции, различия в микрорельефе, разная степень увлажнения, неоднородность эдафических факторов — все эти и другие условия могут незаметным для глаза образом сказываться на выраженности признаков растений. Если в таких условиях ботаника интересуется типичное состояние объекта, то и выборка должна состоять из растений, собранных на типичных с биологической точки зрения участках, занятых данной популяцией. Сама же типическая выборка из участка популяции производится при этом с соблюдением правил случайного отбора, так как и в пределах кажущегося типичным участка может иметь место незаметный для глаза градиент изменчивости признаков.

3. *Смешанный отбор.* Если, несмотря на явную неоднородность популяции, вызванную воздействием факторов среды, необходимо получить представление о популяции в целом, то наиболее правильным путем составления выборки будет смешанный отбор. Заключается он в том, что территорию, занятую популяцией, делят на участки, соответствующие грациям ее неоднородности, и отбирают растения случайным способом на каждом таком участке. Затем объединяют отобранные растения.

Итак, объективность случайного отбора обеспечивается равной вероятностью попадания в выборку любой особи. Поскольку в случае типического отбора эта вероятность неодинакова, то некоторые исследователи считают, что он предполагает внесение доли субъективизма в строгие математические построения. На самом деле это далеко не так. Случайный отбор является, в сущности, бессознательным. Равную вероятность попадания особей в выборку можно рассматривать как признание отсутствия всякой предварительной информации о генеральной совокупности. Когда это действительно так, то другого выхода, кроме случайного отбора, нет. Но на практике ботаник часто имеет некоторую априорную информацию об исследуемом объекте, например о степени типичности тех или иных участков популяции. В этих условиях отказ от использования имеющейся информации противоречит здравому смыслу, а типический или смешанный отбор дает лучшие результаты, нежели отбор случайный. Следовательно, в зависимости от условий сбора материала и биологической задачи исследования надо применять тот способ отбора, который наилучшим образом соответствует данной ситуации.

Как случайная, так и типическая выборка имеет свои достоинства и недостатки. При малом объеме случайная выборка именно в силу своей случайности может дать худшее представление о наиболее характерных значениях признаков, чем типическая, но при увеличении объема она лучше последней воспроизводит генеральную совокупность. И все-таки даже при большом объеме случайная выборка характеризует среднее значение признака обычно не лучше типической. Зато типическая выборка значительно слабее случайной воспроизводит истинный размах изменчивости признаков в совокупности. Поэтому выбор способа составления выборки может зависеть и от того, что больше — среднее значение или мера изменчивости признака — интересует исследователя. Если позволяют обстоятельства, то лучше всего производить смешанный отбор, который сводит к минимуму недостатки как случайной, так и типической выборки.

Все, что говорилось до сих пор, относится к составлению выборок из природных популяций. Но систематик в своей работе нередко должен использовать и гербарные материалы. Правила

составления выборок и в этом случае остаются в силе. Случайную выборку удобно производить с помощью жребия или таблицы случайных чисел, пронумеровав предварительно гербарные образцы. Типическую и смешанную выборки делают на основании этикеток, содержащих сведения географического или экологического характера.

Однако при использовании гербарного материала возникает опасность специфических для данного случая ошибок. Так, например, в гербарии непомерный удельный вес могут иногда иметь сборы из одного и того же места или одного и того же автора, занимавшегося изучением данного вида. Тогда получения правильного представления об изменчивости признаков необходим смешанный отбор с пропорциональным включением в выборку экземпляров, собранных в разных местах разными авторами.

Другая характерная при работе с гербарием опасность кроется в тенденциозной гербаризации некоторых объектов. Например, в силу чисто технических неудобств рослые экземпляры, плохо уместяющиеся на гербарном листе, имеют меньше шансов быть изъятыми из природной популяции. Без учета этого обстоятельства можно получить неверное представление не только о высоте растений данного вида, но и о других признаках, сильно коррелирующих с ней (например, о длине листьев, числе междоузлий, числе цветков).

1.3. ОШИБКИ ВЫБОРОЧНОГО ИССЛЕДОВАНИЯ

В процессе биометрического исследования могут возникать различные по своему происхождению и последствиям ошибки, ведущие к смещению выборочных показателей относительно параметров генеральной совокупности. Эти ошибки могут быть двоякого рода: устранимые и неустраиваемые.

Устранимые ошибки. Сюда относятся погрешности, предотвратить появление которых необходимо правильной организацией биометрического исследования. К их числу принадлежат: методические огрехи в постановке опыта или полевого эксперимента, несоблюдение изложенных выше правил составления выборки, использование грубых измерительных инструментов, недостаточная точность самих измерений, ошибки в записях наблюдений и в выполнении расчетов.

Появления измерительных и расчетных ошибок можно избежать, руководствуясь следующими советами. Измерение морфологических признаков лучше всего производить штангенциркулем. За неимением его следует пользоваться измерителем и тщательно отградуированной линейкой. Точность измерений зависит от задачи исследования, объема выборки и размеров измеряемого органа. Обычно для таксономических целей при измерении морфологических признаков точность измерения со-

ставляет 0,1 мм. При небольшом объеме выборки точность измерений лучше снизить (0,2—0,5 мм). Если размеры органа (признака) велики, точность измерений может быть и еще ниже (например, высоту травянистых растений можно в ряде случаев измерять с точностью до 1 мм или даже до 1 см). Измерение микропризнаков (анатомических структур, размеров пыльцевых зерен, хромосом и т. п.) следует производить с помощью проверенного окулярмикрометра. Очень важно при обработке всего материала пользоваться одним и тем же измерительным прибором, так как каждый прибор имеет свою систематическую (механическую или оптическую) ошибку.

Что касается биометрических расчетов, то каждый расчет, особенно если он производится вручную, во избежание ошибки следует повторять дважды. Точность расчетов (количество значащих цифр) должна соответствовать не только задаче исследования, но и технике обработки данных. Для таксономических целей выборочные параметры должны иметь точность в 1—2 десятичных знака, но в ходе самих расчетов эта точность при некоторых методиках должна достигать иногда 4—6 знаков, ибо в противном случае может накапливаться ошибка за счет промежуточных округлений. Вообще при расчетах всегда лучше иметь большее, чем это нужно для окончательного результата, количество знаков и округлять лишь окончательное число.

Неустраняемые ошибки (их называют также ошибками репрезентативности) не могут быть устранены сознательными усилиями, так как они вытекают из самой сущности выборочного метода исследования: никакой отбор не может обеспечить абсолютного совпадения выборочных и генеральных параметров. Тем не менее при правильной организации отбора и достаточном объеме выборки рассматриваемые ошибки невелики и обеспечивают получение узкого интервала, в котором должны лежать значения генеральных параметров. Существуют простые способы вычисления ошибок репрезентативности основных выборочных параметров (раздел 1.4.7), с помощью которых можно определить тот интервал, в котором заключаются параметры генеральной совокупности.

1.4. ВАРИАЦИОННЫЙ РЯД И ЕГО ОБРАБОТКА

1.4.1. Запись наблюдений. Результаты измерения и подсчета признаков при обработке выборки записываются в виде таблицы. Каждый экземпляр получает свой номер, а в соответствующих графах таблицы проставляются значения всех признаков каждой особи (табл. 1). Такой порядок записи наблюдений обеспечивает не только биометрическую обработку отдельных признаков, но и выполнение в дальнейшем корреляционного и регрессионного анализов, потребность в которых может возникнуть в ходе работы.

Другим возможным способом записи наблюдений является их нанесение на перфорационные карты. В этом случае на каждую особь заводится своя перфокарта, имеющая номер соответствующего экземпляра, а значения признаков наносятся по периферии карты путем ее прокола в определенных точках или надреза готовых перфораций. Перфокарты могут использоваться не только для машинной обработки: ими удобно пользоваться и вручную, нанизывая их на обыкновенную спицу и отбирая особи с одинаковым значением признака.

1.4.2. Типы варьирования признаков. Варьированием (изменчивостью) признака называется его способность принимать у разных особей различные значения. По своему характеру признаки могут быть качественными и количественными.

Таблица 1. Примерная схема записи наблюдений

№ особи	Признаки					Примечания
	A	B	C	...	N	
1	A ₁	B ₁	C ₁	...	N ₁	
2	A ₂	B ₂	C ₂	...	N ₂	
...	
n	A _n	B _n	C _n	...	N _n	

Примером качественного признака может служить окраска венчика. Качественным признакам соответствует *альтернативный* тип изменчивости, при котором имеет место небольшое число (обычно 2—3) градаций или классов признака (например, венчик белый — венчик розовый). Качественные признаки обычно не требуют измерения, но в особо ответственных случаях их градации могут определяться по цветовой шкале или с помощью колориметра.

Количественным признакам, требующим подсчета или измерения, соответствуют два типа изменчивости: *дискретный* и *непрерывный*.

При дискретной изменчивости разница между соседними значениями признака не может быть меньше единицы. Например, число цветков в соцветии выражается членами натурального ряда 1, 2, 3, 4, 5... и не может быть дробным.

При непрерывной изменчивости соседние значения признака, например размеры органа или всего растения в целом, теоретически могут сколь угодно мало отличаться друг от друга. Известная дискретность соседних значений зависит при этом только от принятой точности измерений, обрывающей число десятичных знаков. Если записано, что длина листа равна 3,6 см,

а принятая точность измерения 0,1 см, то в действительности точное значение признака может быть в пределах округления несколько иным (например, 3,55 или 3,64 см). Однако размеры органов, закончивших свой рост, отличаются друг от друга на дискретную величину, кратную размеру кванта роста (раздел 6.1), а само явление квантированности роста свидетельствует о том, что в природе нет резкой грани между дискретным и непрерывным типами изменчивости признаков. Но практически их различать надо, ибо техника обработки данных при этом несколько различна.

При решении задач прикидочного характера (например, для грубого определения наличия или отсутствия корреляционной связи между признаками) возможны некоторые искусственные преобразования. Так, если округлить до целой величины (скажем, до 1 мм) точность измерения длины листа, то от непрерывного типа изменчивости мы перейдем к дискретному. В свою очередь, от дискретного типа можно перейти к альтернативному. К примеру, вместо того, чтобы учитывать число ветвей у растения, можно разделить все экземпляры на два класса: ветвистые и неветвистые. В некоторых случаях оказывается полезным даже непрерывно варьирующие признаки (размеры органов) сводить к альтернативе по принципу «большой — маленький», взяв за разграничительную линию среднее арифметическое или медиану.

Надо иметь в виду, что подобного рода «преобразования» не меняют истинного типа варьирования признака и, будучи искусственными, в значительной мере огрубляют результат. Поэтому прибегать к ним следует только в рекогносцировочных целях. При детальном изучении варьирования отдельных признаков такие преобразования недопустимы.

1.4.3. Группировка данных. Статистическая обработка данных даже в выборках относительно небольшого объема технически трудно выполнима без предварительной их группировки в классы. В зависимости от типа изменчивости признака эта группи-

Таблица 2. Распределение ветвистых и неветвистых экземпляров в выборке *Odontites verna* (Bell.) Dum. (альтернативный тип изменчивости признака)

x	f	% от n
Ветвистые	x_1 91	60,7
Неветвистые	x_2 59	39,3
n	150	100

ровка может быть естественной и искусственной, но в последнем случае необходимо соблюдать определенные правила, позволяющие избежать смещения выборочных показателей.

Каждое частное значение ($x_1, x_2, x_3, \dots, x_n$), которое способен принимать данный признак (x), называется *вариантой*. Количество экземпляров, обладающих данным значением признака, составляет *частоту* (f) встречаемости данной варианты в выборке. Совместный ряд вариантов и соответствующих им частот ($x; f$) образует *вариационный ряд*, подлежащий дальнейшей обработке.

Группировка данных при альтернативной изменчивости не представляет затруднений, так как число естественных классов варьирования признака совпадает с числом вариантов (табл. 2). Первоначальная обработка такого ряда состоит в вычислении процента встречаемости каждого значения признака в выборке.

Группировка данных при дискретном варьировании признака обычно также является естественной, хотя количество классов возрастает (табл. 3).

Таблица 3. Вариационный ряд числа междоузлий у *Odontites litoralis* Fr. (дискретный тип изменчивости признака)

x	Разноска вариантов	f
3	☒ : .	13
4	☒ ☒ ☒ ☒ ☒ ☒ ☒ ☒ .	85
5	☒ ☒ ☒ ☒	40
6	☒ :	12
		$n=150$

В некоторых случаях при дискретной изменчивости признака вариационный ряд получается непомерно длинным за счет большого размаха варьирования признака. Примером может служить изменчивость числа лучей первого порядка в соцветиях зонтичных. Так, у *Pimpinella saxifraga* L. (окрестности ст. Сиверская Ленинградской области, 1965 г.) число лучей варьирует от 10 до 33, т. е. составляет 24 естественных класса изменчивости. Обработка такого ряда вручную затруднительна, и выход из положения состоит в объединении вариантов по две:

x :	10—11	12—13	14—15	16—17	18—19	20—21	22—23
f :	4	38	134	181	206	182	103
x :	24—25	26—27	28—29	30—31	32—33		
f :	50	18	10	3	1	(n=930)	

Теперь вариационный ряд состоит из 12 классов и легко может быть обработан, однако эту возможность мы обрели ценою жертвы части информации, и, следовательно, результаты вычисления выборочных параметров будут несколько огрубленными.

Более сложна и должна выполняться с минимумом потери информации группировка вариант в классы при непрерывном типе изменчивости признака. В этом случае признак в выборке характеризуется большим количеством разных вариант, и может даже возникнуть положение, когда все значения признака отличаются друг от друга. Если к тому же объем выборки достаточно велик, то обработка вариационного ряда без группировки вариант в классы возможна лишь с помощью вычислительных машин. При обработке ряда вручную группировка неизбежна, а ее суть состоит в объединении близких по своему значению вариант в классы и замене попадающих в каждый класс вариант усредненной величиной — центральным классовым значением ($x_{ц}$). Потеря информации при этом будет невелика, если число классов в определенной мере соответствует объему выборки и степени изменчивости признака. Число классов должно быть тем больше, чем крупнее выборка, и тем меньше, чем сильнее изменчивость признака.

Согласно правилу Старджеса, число классов (k) в зависимости от объема выборки (n) определяется формулой $k=1+\sqrt{3,3 \log n}$, исходя из которой следует придерживаться следующей рекомендации:

n : 10—50	50—100	100—400	400—1000	1000—2000
k : 4—6	6—8	8—9	9—11	11—12

Избрав число классов, устанавливают их границы, для чего сначала определяют величину классового интервала (i), который должен быть одинаковым на протяжении всего вариационного ряда. Величину классового интервала получают посредством деления размаха варьирования признака (разница между максимальной варианттой x_{\max} и минимальной варианттой x_{\min}) на выбранное число классов (k):

$$i = \frac{x_{\max} - x_{\min}}{k}$$

Полученное значение i для удобства округляют в большую сторону до целого числа (или с точностью до 0,5). Левую границу первого класса назначают с таким расчетом, чтобы значение x_{\max} (или x_{\min} , если ряд строится в убывающем порядке) попало внутрь первого класса, но в то же время сама граница была обозначена удобным (лучше всего целым или с точностью до 0,5) числом.

Рассмотрим произвольный пример. Пусть в выборке объемом $n=100$ $x_{\max}=100,3$ мм, а $x_{\min}=69,7$ мм. Вариационный ряд желательно разбить на 8 классов. Находим классовый интервал:

$$i = \frac{100,3 - 69,7}{8} = 3,84 \simeq 4,0.$$

Отодвигая левую границу первого класса до 69,0, определяем левые границы остальных классов путем последовательного прибавления к каждой из них величины классового интервала: 69,0—73,0—77,0—81,0—85,0—89,0—93,0—97,0.

Правые границы классов надо указать таким образом, чтобы они отличались на единицу точности наблюдения от левой

Таблица 4. Пример группировки вариантов в классы при непрерывной изменчивости признака

x	$x_{ц}$	Разность вариант	f
69,0 - 72,9	70,95	—	5
73,0 - 76,9	74,95	☒ :	12
77,0 - 80,9	78,95	☒ ☒ : :	23
81,0 - 84,9	82,95	☒ ☒ ☒	30
85,0 - 88,9	86,95	☒ ☒	20
89,0 - 92,9	90,95	☐	7
93,0 - 96,9	94,95	:	2
97,0 - 100,9	98,95	.	1

границы следующего класса, иначе неизвестно, к какому классу следует относить граничные значения признака. В нашем произвольном примере окончательная разбивка ряда на классы при точности измерений 0,1 мм приведена в табл. 4.

Центральные классовые значения ($x_{ц}$), которыми заменяются все значения признака, входящие в данный класс, рассчитывают, прибавляя к левой границе класса половину разности между его границами. Например, для первого класса $x_{ц} = 69,0 + (72,9 - 69,0) : 2 = 70,95$. Для получения центральных значений остальных классов достаточно последовательно прибавлять классовый интервал к центральному значению предыдущего класса.

1.4.4. Графическое изображение вариационного ряда. Существуют три способа изображения вариационного ряда на гра-

фике: вариационная кривая (кривая распределения частот), гистограмма и кумулятивная огива.

Для построения вариационной кривой на оси абсцисс откладываются значения вариант (при дискретной изменчивости) или центральные классовые значения (при непрерывном типе варьирования признака), а по оси ординат — частоты (рис. 1, а).

Для построения гистограммы на оси абсцисс откладываются границы классов, а численность каждого класса выражается площадью (и, следовательно, высотой) соответствующего прямоугольника. Соединив ломаной линией середины верхних ос-

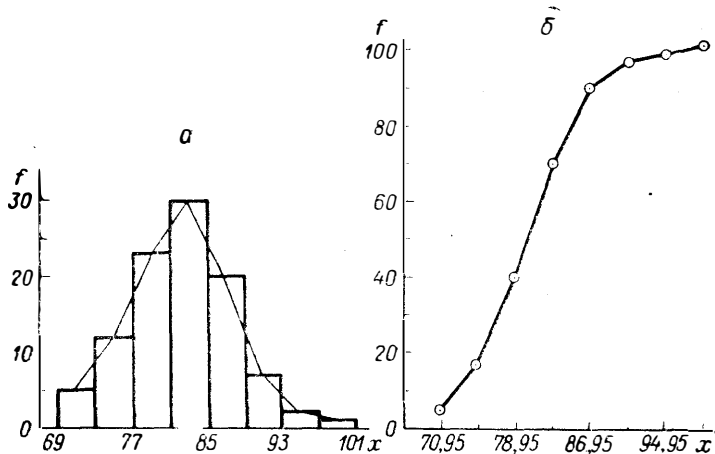


Рис. 1. Способы графического изображения вариационного ряда (по данным табл. 4).

а — вариационная кривая (полигон распределения) и гистограмма; б — кумулятивная огива.

нований всех прямоугольников, от гистограммы легко перейти к вариационной кривой.

Графическое сопоставление изменчивости одноименных признаков в сравниваемых выборках удобнее всего производить путем построения на одном графике нескольких вариационных кривых. Группировка вариант в классы для сравниваемых выборок при этом должна быть одинаковой, а при равенстве объемов выборок все частоты должны быть пересчитаны в проценты от числа наблюдений (n) в каждой выборке.

Некоторые биометрические приемы, например, вычисление критерия λ (раздел 1.5.2), основаны на составлении ряда накопленных частот: к частоте наименьшего класса прибавляется частота второго класса, к полученной сумме — частота третьего класса и т. д., причем окончательная сумма равна n , т. е. объ-

ему выборки. График ряда накопленных частот представляет собою кумулятивную огиву (рис. 1, б).

1.4.5. Основные выборочные параметры. Основные выборочные параметры можно подразделить на три группы. Показатели первой группы характеризуют некоторое среднее значение признака и называются показателями *положения* или показателями *центральной тенденции*. К их числу принадлежат различные виды средних, из которых наиболее употребительны средняя арифметическая, или медия, а также мода и медиана. Все они являются величинами именованными, сохраняющими ту размерность, в которой выражен признак. Ко второй группе относятся показатели *изменчивости*. Важнейшие из них — среднее квадратичное (или стандартное) отклонение, представляющее собой именованную величину, и коэффициент вариации — величина относительная, безразмерная. Наконец, третью группу образуют показатели *формы распределения* (показатели асимметрии и эксцесса), о которых речь пойдет несколько позже (раздел 2.4), ибо их рассмотрению должна предшествовать определенная подготовка. В данном разделе мы коснемся показателей первых двух групп, вычисление которых составляет важнейший этап обработки вариационного ряда.

Средняя арифметическая, или медия (M), есть абстрактная характеристика ряда, заменяющая собой варьирующие значения признака. Теоретически она может не совпадать ни с одной из вариантов, аккумулируя в себе их значения. Вообще говоря, для вычисления медиан достаточно суммировать все частные значения признака в выборке и разделить эту сумму на объем выборки:

$$M = \frac{\sum x}{n},$$

где $\sum x$ — сумма всех вариантов ряда; n — объем выборки.

Если же вариационный ряд сгруппирован в классы, то формула для расчета медиан приобретает вид

$$M = \frac{\sum f x_{ц}}{n},$$

где f — частота каждого класса; $x_{ц}$ — его центральное значение.

Обе формулы могут использоваться для прямого вычисления средней арифметической в малых выборках (и в любых выборках в случае применения ЭВМ), однако при ручной обработке крупных выборок следует обращаться к косвенным приемам расчета выборочных параметров (раздел 1.4.6), основанным на том, что медия обладает несколькими, присущими только ей, математическими свойствами. Отметим важнейшие из них.

1. Сумма отклонений вариантов от медиан равна нулю: $\sum (x - M) = 0$. Это свойство средней арифметической очень важно

для понимания ее как некоей равнодействующей вариационного ряда. Именно на нем основан косвенный расчет медики с помощью так называемой «условной средней» (раздел 1.4.6).

2. Сумма квадратов отклонений вариант от медики всегда меньше суммы квадратов отклонений вариант от любой другой величины (K) в пределах вариационного ряда: $\Sigma(x-M)^2 < \Sigma(x-K)^2$. Данное свойство средней арифметической также используется в схеме косвенного расчета выборочных параметров, составляет основу дисперсионного анализа (раздел 3.1) и метода наименьших квадратов (раздел 5.3.1).

3. Если к каждому значению признака прибавить или вычесть из него постоянную величину K , а также умножить или разделить каждое его значение на ту же величину K , то средняя арифметическая увеличится или уменьшится на величину K или в K раз. Это свойство медики используется для облегчения вычислительной работы. Заметив, например, что все варианты выражены числами больше 100, можно каждое из них уменьшить на 100, упростив тем самым расчеты. Полученное таким путем значение медики надо будет затем увеличить на 100.

Мода (M_0) представляет собой наиболее часто встречающееся в вариационном ряду значение признака. На графике это значение соответствует максимальной частоте, т. е. вершине вариационной кривой. Если вариационный ряд разбит на классы, то существует модальный класс, которому соответствует максимальная частота. В этом случае значение моды рассчитывается по формуле

$$M_0 = x_m + i \left(\frac{f_2 - f_1}{2f_2 - f_1 - f_3} \right),$$

где x_m — меньший предел модального класса; i — классовый интервал; f_1 — частота класса, предшествующего модальному; f_2 — частота модального класса; f_3 — частота класса, следующего за модальным. Чаще всего вариационный ряд имеет одну моду, а вариационная кривая — одну вершину. Однако иногда кривые получаются полимодальными, т. е. многовершинными (подробнее об этом см. раздел 6.1.1).

Медианой (M_e) называется такое срединное (но не среднее!) значение признака, которое делит вариационный ряд на две части, из которых одна имеет значения признака, меньшие, чем медиана, а другая — большие. Для нахождения медианы необходимо ранжировать ряд, т. е. расположить все варианты в убывающем или возрастающем порядке. При нечетном числе наблюдений середину ряда составит одна варианта, которая и будет медианой. При четном числе наблюдений в середине ряда окажутся две варианты: для получения медианы достаточно найти их среднее арифметическое.

Если же вариационный ряд сгруппирован в классы, то медиану находят по следующей формуле:

$$M_e = x_{M_e} + i \left(\frac{0,5n - \sum f}{f_{M_e}} \right),$$

где x_{M_e} — начало класса, в котором лежит медиана; n — объем выборки; $\sum f$ — сумма частот всех классов, предшествующих модальному классу; f_{M_e} — частота модального класса.

В некоторых случаях медиана дает более правильное представление о середине ряда, чем средняя арифметическая. Так бывает, например, когда при сравнительно небольшом объеме выборки наблюдаются сильно «отскакивающие» варианты, существенно влияющие на положение медианы, но никак не отражающиеся на значении медианы.

Среднее квадратичное отклонение, часто называемое просто сигмой (σ), является одним из показателей, отражающих степень изменчивости вариант. Эту задачу проще всего было бы решить, суммировав все отклонения от медианы и разделив их на объем выборки. Но согласно первому свойству средней арифметической сумма таких отклонений равна нулю. Поэтому приходится использовать квадрат отклонений вариант от медианы, который отличен от нуля и называется дисперсией или дисперсией (σ^2), широко используемой в дисперсионном анализе (глава 3) и находящей в биометрии многие другие применения:

$$\sigma^2 = \frac{\sum (x - M)^2}{n}.$$

Корень квадратный из дисперсии и есть среднее квадратичное отклонение:

$$\sigma = \sqrt{\frac{\sum (x - M)^2}{n}} \quad \text{или} \quad \sigma = \sqrt{\frac{\sum (x - M)^2}{n - 1}}.$$

Последние две формулы различаются только тем, что в знаменателе подкоренного выражения в первой из них стоит объем выборки (n), а во второй — величина $n-1$, обозначающая *число степеней свободы* в данном случае. Число степеней свободы (ν) определяет количество значений, необходимое для восстановления утерянного. Когда мы имеем дело с одним вариационным рядом, то для того, чтобы восстановить утерянное значение признака, необходимо и достаточно знать медиану и все оставшиеся, т. е. $n-1$, значения этого признака. Поэтому в случае одного ряда число степеней свободы $\nu = n-1$. При большом объеме выборки можно пользоваться первой формулой, но при $n < 100$ рекомендуется использовать вторую, т. е. учитывать число степеней свободы.

Подобно медию среднее квадратичное отклонение рассчитывается обычно косвенным путем, ибо приведенные формулы, отражающие математическую суть этого параметра, неудобны в вычислительном отношении.

Биологический смысл среднего квадратичного отклонения заключается в том, что оно показывает среднее отклонение значений признака в выборке от средней арифметической. Кроме того, оно находит применение в характеристике нормального распределения и в ряде важных биометрических методик.

Коэффициент вариации (V) представляет собою еще один параметр, характеризующий степень изменчивости признака. Дело в том, что с помощью среднего квадратичного отклонения (σ) невозможно сравнивать изменчивость признаков, имеющих разную размерность. Например, если признак высоты растения обладает $\sigma=2,8$ см, а для числа цветков $\sigma=5,4$, то из этих цифр мы не можем заключить, какой же из двух признаков более изменчив.

Коэффициент вариации, будучи величиной относительной, позволяет решать такого рода задачи. Рассчитывается он по формуле

$$V = \frac{100 \sigma}{M} \%,$$

выражаясь, таким образом, в процентном отношении сигмы к медию.

1.4.6. Косвенные способы расчета основных выборочных параметров. Существуют два способа косвенного вычисления выборочных параметров: способ произведений и способ сумм, дающие одинаковые результаты.

Рассмотрим сначала обработку ряда способом произведений (табл. 5). Порядок работы следующий. В таблице вслед за центральными классовыми значениями ($x_{ц}$) выписывают соответствующие им частоты (f). Затем выбирают так называемую *условную среднюю* (A)-варианту или центральное классовое значение, исходя при этом выборе из двух требований: условная средняя должна по возможности соответствовать максимальной частоте и в то же время находиться примерно посередине ряда. Если ряд заметно асимметричен, то условная средняя может лежать в соседнем с модальным классе, ближе к середине ряда. В нашем примере (табл. 5) условную среднюю лучше всего выбрать в модальном классе (3,0—3,9), а ее значение будет соответственно $A=3,45$.

В столбце a расчетной таблицы положение условной средней обозначается нулем, а вверх и вниз от нуля производятся отсчеты натуральными числами ($-1, -2$ и т. д.; $+1, +2$ и т. д.). Эти отсчеты характеризуют отклонения центральных классовых значений от условной средней, заменяя тем самым менее удобные натуральные отклонения.

Столбец fa включает в себе произведения частот на соответствующие отклонения классов от условной средней. В последнем столбце эти произведения еще раз умножают на величины a , получая, таким образом, произведения частот на квадраты отклонений от условной средней fa^2 .

Из того, что говорилось ранее о математической сущности выборочных параметров, должно быть ясно, что столбец fa содержит данные, необходимые для вычисления медиан, а столбец fa^2 — данные для расчета среднего квадратичного отклонения. Этим целям служат суммы Σfa и Σfa^2 , помещенные как итог в нижней строке расчетной таблицы.

Таблица 5. Обработка вариационного ряда способом произведений (x — ширина листа у *Odontites verna* (Bell.) Dum.)

x , мм	$x_{ц}$	f	a	fa	fa^2
1.0—1.9	1.45	7	—2	—14	28
2.0—2.9	2.45	42	—1	—42	42
3.0—3.9	$A=3.45$	52	0	0	0
4.0—4.9	4.45	36		+36	36
5.0—5.9	5.45	11	+2	+22	44
6.0—6.9	6.45	2	+3	+6	18
		$n=150$		$\Sigma fa=+8$	$\Sigma fa^2=168$

Для расчета параметров используются рабочие формулы, причем надо помнить, что условная средняя в нашем примере $A=3,45$, а классовый интервал $i=1$. Сначала находят среднюю арифметическую

$$M = A + i \frac{\Sigma fa}{n} = 3,45 + 1 \frac{8}{150} = 3,50 \text{ мм.}$$

Затем вычисляют среднее квадратичное отклонение

$$\sigma = i \sqrt{\frac{\Sigma fa^2}{n} - \left(\frac{\Sigma fa}{n}\right)^2} = 1 \sqrt{\frac{168}{150} - \left(\frac{8}{150}\right)^2} = \pm 1,03 \text{ мм.}$$

Коэффициент вариации определяют, пользуясь ранее приведенной формулой:

$$V = \frac{100 \sigma}{M} = \frac{106}{3,50} = 30,3\%.$$

Формулы для расчета моды (M_0) и медианы (M_e) также были даны ранее:

$$M_0 = 3,0 + 1 \left(\frac{52 - 42}{2 \cdot 52 - 42 - 36} \right) = 3,38 \text{ мм.}$$

$$M_e = 3,0 + 1 \left(\frac{0,5 \cdot 150 - 49}{52} \right) = 3,50 \text{ мм.}$$

Способ сумм отличается от способа произведений прежде всего тем, что в расчетной таблице действие умножения заменено действием сложения. Этот способ также основан на использовании условной средней. Для того чтобы убедиться в совпадении результатов обработки ряда обоими способами, возьмем в качестве примера тот же самый вариационный ряд, что и в первом случае (табл. 6).

Таблица 6. Обработка вариационного ряда способом сумм

x , мм	x_n	f	$a_1 = 56$	$a_2 = 7$	$a_3 = 0$	$a_4 = 0$
1,0 - 1,9	1,45	7	↙ 7	7	—	—
2,0 - 2,9	2,45	42	↔ 49	—	—	—
3,0 - 3,9	$A = 3,45$	52	—	—	—	—
4,0 - 4,9	4,45	36	↔ 49	—	—	—
5,0 - 5,9	5,45	11	↙ 13	↔ 15	—	—
6,0 - 6,9	6,45	2	↙ 2	↙ 2	2	—
Пояснения.		$n = 150$	$b_1 = 64$	$b_2 = 17$	$b_3 = 2$	$b_4 = 0$
$s = a + b$			$s_1 = 120$	$s_2 = 24$	$s_3 = 2$	$s_4 = 0$
$d = b - a$			$d_1 = +8$	$d_2 = +10$	$d_3 = +2$	—

Сначала по ранее изложенным правилам выбирают условную среднюю $A = 3,45$. В четырех последних столбцах против условной средней делают прочерки, исключая соответствующие клетки расчетной таблицы из дальнейшей обработки. Число таких прочерков последовательно наращивают: в 4-м столбце — 1 прочерк, в 5-м — 3, в 6-м — 5 и в 7-м — 7 прочерков. Если ряд асимметричен или слишком короток, то не все прочерки могут уместиться в таблице, а кроме того, как это получилось и в нашем примере, один или даже два последних столбца могут оказаться лишенными цифр. Это обстоятельство ничего не меняет в ходе расчетов.

Следующий этап работы состоит в написании так называемых начетных рядов отдельно в верхней и нижней половинах каждого столбца. Начетные ряды составляются путем суммирования чисел в направлении, указанном в таблице стрелками. Затем в каждой половине столбцов подсчитываются суммы a_1, a_2, a_3, a_4 и b_1, b_2, b_3, b_4 , которые и записываются в соответствующих графах. После этого в столбцах рассчитываются общие суммы по формуле $s = a + b$ и разности по формуле $d = b - a$. Разность d_4 для обычных целей не нужна.

Для расчета параметров ряда предварительно вычисляют три *статистических момента*¹: условный момент первого порядка (\bar{v}_1), условный момент второго порядка (\bar{v}_2) и центральный момент второго порядка (μ_2):

$$\bar{v}_1 = \frac{d_1}{n} = \frac{8}{150} = 0,0533;$$

$$\bar{v}_2 = \frac{s_1 + 2s_2}{n} = \frac{120 + 2 \cdot 24}{150} = \frac{168}{150} = 1,1200;$$

$$\mu_2 = \bar{v}_2 - \bar{v}_1^2 = 1,1200 - 0,0533^2 = 1,1172.$$

Затем, подставляя найденные значения в рабочие формулы, рассчитывают медию и стандартное отклонение:

$$M = A + i\bar{v}_1 = 3,45 + 1 \cdot 0,0533 = 3,50 \text{ мм};$$

$$\sigma = i\sqrt{\mu_2} = 1\sqrt{1,1172} = \pm 1,06 \text{ мм}.$$

Как уже отмечалось, группировка вариантов в классы ведет к потере некоторой части информации об изменчивости признака в выборке. Устранить этот недостаток можно путем вычитания из рассчитанного значения σ так называемой поправки Шеппарда (*III*), которая определяется по формуле

$$III = i^2:12,$$

где i — классовый интервал. Исследования статистиков показывают, однако, что введение поправки Шеппарда имеет смысл только для больших выборок ($n > 500$).

1.4.7. Ошибки репрезентативности выборочных параметров. Как уже говорилось, выборочные параметры являются приближенными оценками параметров генеральной совокупности. Величина отклонения выборочного параметра от генерального называется *ошибкой репрезентативности*.

¹ Специальное рассмотрение метода моментов не входит в нашу задачу, но некоторые его положения и случаи частного применения упомянуты далее (раздел 2.5). Желающих углубить свои знания в этом отношении мы отсылаем к книгам В. Ю. Урбаха [1964] и А. К. Митропольского [1971].

Ошибка репрезентативности меди

$$m_M = \frac{\sigma}{\sqrt{n}}.$$

Ошибка репрезентативности среднего квадратичного отклонения

$$m_\sigma = \frac{\sigma}{\sqrt{2n}}.$$

Ошибка репрезентативности коэффициента вариации

$$m_V = \frac{V}{\sqrt{2n}}.$$

В нашем примере обработки вариационного ряда (раздел 1.4.6) расчет этих ошибок дает следующие величины:

$$m_M = \frac{1,06}{\sqrt{150}} = 0,09 \text{ мм};$$

$$m_\sigma = \frac{1,06}{\sqrt{300}} = 0,06 \text{ мм};$$

$$m_V = \frac{30,3}{\sqrt{300}} = 1,75\%.$$

В разделе 2.3 показано, как, зная величину ошибки репрезентативности, определить интервал, в котором лежит генеральный параметр.

Имеется также возможность оценить точность определения средней арифметической путем расчета ее «относительной ошибки», которая не должна превышать 5%:

$$P = \frac{100m_M}{M} = \frac{V}{\sqrt{n}} \leq 5\%.$$

Значение $P > 5\%$ указывает на то, что изменчивость признака слишком высока для того, чтобы при данном числе наблюдений можно было получить достаточно точное (близкое к генеральному) значение меди: объем выборки необходимо увеличить.

Ошибка этого показателя определяется формулой

$$m_P = P \sqrt{\frac{1}{2n} + \left(\frac{P}{100}\right)^2}$$

и имеет такой же смысл, как ошибки репрезентативности других выборочных параметров.

1.4.8. Проверка принадлежности варианты к данной совокупности. На краях вариационного ряда иногда наблюдаются значения признака, сильно отличающиеся от основной массы ва-

риант. Причины появления таких «отскакивающих» вариант могут быть разными. Прежде всего надо проверить, не являются ли они результатом ошибки наблюдения, измерения или записи. Если ошибки не обнаружено, то остается решить вопрос о том, лежит ли «отскакивающая» варианта в пределах вероятного отклонения от средней арифметической (и тогда ее следует оставить в выборке) или же она выходит за эти пределы, принадлежит к другой генеральной совокупности и должна быть квалифицирована как «артефакт» (т. е. исключена из выборки).

Соответствующий критерий (τ_α) представляет собой нормированное отклонение предполагаемого «артефакта» (x_α) от меди:

$$\tau_\alpha = \frac{x_\alpha - M}{\sigma},$$

где M и σ — средняя арифметическая и среднее квадратичное отклонение, рассчитанное для всей выборки, включая возможный «артефакт». Правило оценки состоит в том, что при $\tau_\alpha < \tau_{st}$ (стандартные значения этого критерия τ_{st} приведены в Приложении II) «отскакивающее» значение варианты считается допустимым и включается в выборку, а при $\tau_\alpha > \tau_{st}$ оно признается «артефактом» и из выборки исключается.

Пусть получены следующие наблюдения: 2, 2, 3, 5, 5 и 10. Возникает вопрос: не является ли вариант 10 «артефактом»? Объем выборки $n=6$, а подсчеты показывают, что здесь $M=4,50$ и $\sigma=3,02$. Расчет критерия дает

$$\tau_\alpha = \frac{10 - 4,50}{3,02} = 1,82.$$

Стандартное значение критерия (Приложение II) при 5%-ном уровне значимости составляет $\tau_{st}=2,70$, а при 1%-ном уровне $\tau_{st}=2,16$. Таким образом, при любом из уровней существенности $\tau_\alpha < \tau_{st}$ и варианта 10 должна быть оставлена в выборке: ее отклонение лежит в пределах вероятного. Надо заметить, что данный критерий, основанный на предположении о нормальности распределения, малопригоден при заметной асимметрии вариационных кривых. В таком случае следует обращаться к непараметрическим его модификациям [Терентьев, Ростова, 1977, с. 40]. При достаточно большом объеме выборки ($n \geq 100$) исключение из ее состава отдельных «артефактов» не влияет сколько-нибудь серьезно на величину выборочных параметров. Но в меньших по объему выборках (особенно при $n \leq 20$) эту операцию, обоснованную вышеприведенным критерием, надо считать совершенно необходимой.

1.5. СРАВНЕНИЕ ОДНОИМЕННЫХ ПРИЗНАКОВ В РАЗНЫХ ВЫБОРКАХ

Сравнение признаков биологических объектов может производиться различными способами и преследовать разные цели. Нередко требуется сравнить эмпирический вариационный ряд с предполагаемой математической моделью — тем или иным теоретическим распределением. Но еще чаще возникает необходимость сопоставления двух эмпирических рядов, которое может быть выполнено как путем сравнения отдельных выборочных параметров, так и посредством сравнения рядов в целом. Задачи сравнения и его методы в этих случаях различны.

Сравнивая параметры, например, средние арифметические (M) или средние квадратичные отклонения (σ), мы судим о принадлежности выборок к одной и той же генеральной совокупности по центральной тенденции или по характеру изменчивости признака в отдельности. Сравнивая вариационные ряды в целом, мы, не конкретизируя степень различия отдельных параметров, получаем представление о сходстве или различии рядов по центральной тенденции и изменчивости признака одновременно.

Выбор метода сравнения зависит от существа поставленной биологической задачи. В ходе анализа используют понятие *нулевой гипотезы* (H_0), которая в любом случае гласит, что различия нет. Нулевой гипотезе противостоит *нон-нуль-гипотеза* (H_n), вернее, множество таких гипотез, говорящих, что та или иная степень различия имеет место. Принимая H_0 , мы одновременно отвергаем H_n , и наоборот. Решение об этом принимается на основе расчета статистических критериев различия, которые могут быть параметрическими и непараметрическими.

1.5.1. Параметрические критерии различия. Статистические критерии, с помощью которых можно установить достоверность различия между параметрами (M , σ , V) вариационных рядов одноименного признака в двух выборках (или в выборке и генеральной совокупности), называются *параметрическими*. Разработаны они в предположении, что распределения сравниваемых рядов близки к нормальному, поэтому при сравнении малых выборок ($n \leq 20$) или при подозрении, что распределение отличается от нормального, лучше пользоваться непараметрическими критериями различия (раздел 1.5.2).

Из параметрических критериев различия наиболее употребительным является критерий Стьюдента (t). В основу этого критерия положено представление о нормированном отклонении, которое подробнее будет рассмотрено при характеристике нормального распределения (раздел 2.3), но которым мы уже пользовались для проверки принадлежности варианты к данной совокупности (раздел 1.4.8).

Критерий Стьюдента (t) для сравнения одноименных параметров (P_1 и P_2) двух вариационных рядов при $n > 20$ в общей форме имеет вид

$$t = \frac{P_1 - P_2}{m_{P_1 - P_2}},$$

где в знаменателе стоит ошибка разности этих параметров, представляющая собою корень квадратный из суммы квадратов ошибок репрезентативности выборочных параметров:

$$m_{P_1 - P_2} = \sqrt{m_{P_1}^2 + m_{P_2}^2}.$$

С учетом формул ошибок репрезентативности, приведенных ранее (с. 28), критерий t приобретает окончательный вид:

а) для сравнения средних арифметических

$$t = \frac{M_1 - M_2}{\sqrt{m_{M_1}^2 + m_{M_2}^2}};$$

б) для сравнения средних квадратичных отклонений

$$t = \frac{\sigma_1 - \sigma_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}};$$

в) для сравнения коэффициентов вариации

$$t = \frac{V_1 - V_2}{\sqrt{m_{V_1}^2 + m_{V_2}^2}}.$$

Оценка достоверности разницы производится с помощью сравнения полученного значения t со стандартным (t_{st}) (Приложение III) при выбранном уровне существенности и числе степеней свободы $\nu = n_1 + n_2 - 2$. Например, изучалось число лепестков венчика у *Ficaria verna* Huds. из двух популяций в окрестностях г. Пушкина Ленинградской области (1965 г.). Получены следующие выборочные параметры:

$$M_1 = 8,66 \pm 0,06; \quad \sigma_1 = 0,78; \quad n_1 = 189;$$

$$M_2 = 8,43 \pm 0,04; \quad \sigma_2 = 0,77; \quad n_2 = 300.$$

Критерий t для определения достоверности разницы между средними

$$t = \frac{8,66 - 8,43}{\sqrt{0,06^2 + 0,04^2}} = 3,19.$$

Критерий t для определения достоверности разницы между средними квадратичными отклонениями

$$t = \frac{0,78 - 0,77}{\sqrt{\frac{0,78^2}{2 \cdot 189} + \frac{0,77^2}{2 \cdot 300}}} = 1,67.$$

По таблице значений t_{st} (Приложение III) при $v=300+189-2=487$ (практически 500) и 1%-ном уровне существенности находим, что $t_{01}=2,59$.

Отсюда следует вывод, что две популяции *Ficaria verna* Huds., достоверно отличаясь по среднему значению признака (что вызвано, очевидно, экологическими различиями местообитаний), совпадают по степени изменчивости числа лепестков венчика (что обусловлено генотипом).

Более точным способом сравнения средних квадратичных отклонений является критерий Фишера (F), широко используемый также в дисперсионном анализе (глава 3) и представляющий собой отношение двух дисперсий:

$$F = \frac{\sigma_1^2}{\sigma_2^2},$$

причем в числителе берут большую из них. Применительно к рассмотренному примеру имеем: $F = \frac{0,78^2}{0,77^2} = 1,03$. Из таблицы значений F_{st} (Приложение IV) находим, что при $v_1 = n_1 - 1 = 189 - 1 = 188$ и $v_2 = n_2 - 1 = 300 - 1 = 299$ (приходится брать $v_1 = 200$ и $v_2 = 200$) $F_{01} = 1,39$, т. е. $F < F_{st}$ при 1%-ном уровне существенности: достоверной разницы между средними квадратичными отклонениями нет и вывод остается прежним.

Все сказанное справедливо в отношении некоррелированных выборок. Если же каждому наблюдению (каждой особи) одной выборки в каком-либо отношении соответствует определенное наблюдение (конкретная особь) другой выборки, то возникает корреляция, заставляющая несколько изменить методику определения достоверности разницы между средними и сигмами. Необходимость в этом может, например, возникнуть при опытно-изучении апомиктных или вегетативно размножающихся растений, потомки которых, выращиваемые в разных условиях, являются производными одного родителя и, следовательно, теснее связаны с ним в проявлении своих признаков. Пригодные в таких случаях критерии рассматриваются в ряде пособий (например, [Урбах, 1964; Терентьев, Ростова, 1977]).

1.5.2. Непараметрические критерии различия. Статистические критерии, с помощью которых устанавливается степень сходства и различия эмпирического и теоретического или двух эмпирических рядов распределения признака в целом и которые не требуют для своего расчета знания параметров ряда, называются *непараметрическими* критериями различия. Наиболее употребителен критерий χ^2 (хи-квадрат), открытый еще в 1875—1877 гг. Хельмертом, но затем забытый и через 25 лет теоретически подробно разработанный К. Пирсоном.¹

¹ О других критериях см.: [Терентьев, Ростова, 1977].

Критерий χ^2 позволяет решать самые различные задачи из области сопоставления данных. Наиболее важные случаи применения χ^2 — установление степени соответствия эмпирически наблюдаемых и теоретически ожидаемых данных и определение степени соответствия двух эмпирических вариационных рядов. Этот критерий находит также широкое применение в некоторых вариантах корреляционного анализа (раздел 4.4). Для того чтобы понять сущность метода, необходимо рассмотреть построение и свойства данного критерия на некоторых примерах.

Наиболее общая формула критерия χ^2 имеет вид

$$\chi^2 = \sum \frac{(\mathcal{E} - T)^2}{T},$$

где \mathcal{E} — эмпирические данные, фактически наблюдаемая величина; T — теоретические данные, математическое ожидание события. Из формулы видно, что при уменьшении расхождения между эмпирическими и теоретическими данными значение χ^2 уменьшается и обращается в нуль при полном их совпадении.

Оценка достоверности различия между сравниваемыми данными производится по общему правилу оценки нуль-гипотез с помощью специальной таблицы (Приложение V) при выбранном уровне существенности: если $\chi^2 > \chi_{st}^2$, различие считается существенным и нуль-гипотеза отбрасывается. Число степеней свободы вычисляется в общем случае по формуле $\nu = k - 1$, где k — число классов сравниваемых величин, однако при сравнении эмпирических рядов с теоретическими распределениями оно будет иным, а именно: для нормального распределения $\nu = k - 3$, для биномиального и пуассонова $\nu = k - 2$.

Возьмем произвольный пример. При моногибридном скрещивании томатов с желтым и красным плодом теоретически (по закону Менделя) ожидается соотношение 3:1. В действительности из 400 плодов получено 310 красных и 90 желтых (вместо ожидаемых 300:100). Требуется оценить, имеется ли в этом случае существенное отклонение от закона Менделя? Используя приведенную выше формулу, получаем

$$\chi^2 = \frac{(310 - 300)^2}{300} + \frac{(90 - 100)^2}{100} = 1,33.$$

При $\nu = 2 - 1 = 1$ теоретическое значение $\chi_{0,5}^2 = 3,841$ (Приложение V), откуда $\chi^2 < \chi_{st}^2$, и можно заключить, что отклонение не существенно, т. е. полученное соотношение 310:90 не противоречит закону Менделя. Это утверждение справедливо для конкретного случая и не означает, что с увеличением числа наблюдений результат не может измениться. С помощью критерия χ^2 можно проверить нормальность эмпирического вариационного ряда или принадлежность его к другому типу распределения. Для этого особым образом по параметрам (M , σ) эмпирического ряда рассчитывается соответствующая теоретическая кри-

вая, с которой потом и сопоставляется эмпирический ряд (см. главу 2).

Чаще возникает необходимость сопоставления двух эмпирических рядов при сравнении одноименных признаков в двух выборках. Но прежде чем показать примеры такого сравнения, следует отметить некоторые ограничения использования χ^2 в данном случае. Одно из них состоит в том, что объем выборок должен быть достаточно велик ($n \geq 50$). Другое ограничение заключается в том, что частоты классов не должны быть менее 5, а так как на концах рядов это случается часто, то соответствующие классы приходится объединять. Наконец, кри-

Таблица 7. Сравнение вариационных рядов длины зубца чашечки (x) *Odontites verna* (Bell.) Dum. (f_1) и *O. litoralis* Fr. (f_2) посредством критерия χ^2

x , мм	f_1	f_2	$f_1 - f_2$	$(f_1 - f_2)^2$	$\frac{(f_1 - f_2)^2}{f_1 + f_2}$
1,3—1,5	1] —	—	—	—	—
1,6—1,8	2] —	13	-13	169	13,0000
1,9—2,1	44] 47	67	-20	400	3,5088
2,2—2,4	28	24	4	16	0,3077
2,5—2,7	26	34	-8	64	1,0667
2,8—3,0	38	11] 12	26	676	13,5200
3,1—3,3	7] 11	1] —	11	121	11,0000
3,4—3,6	3] —	—	—	—	—
3,7—3,9	1] —	—	—	—	—
	$n_1 = 150$	$n_2 = 150$			$\chi^2 = 42,4032$

терий χ^2 применим только к абсолютным значениям частот (нельзя использовать частоты, выраженные в процентах).

Поясним применение критерия χ^2 следующими двумя примерами. В табл. 7 приведены вариационные ряды длины зубца чашечки двух видов *Odontites* и показан алгоритм расчета критерия χ^2 , который в случае равного объема выборок ($n_1 = n_2$) определяется формулой

$$\chi^2 = \sum \frac{(f_1 - f_2)^2}{f_1 + f_2},$$

где f_1 и f_2 — частоты первого и второго рядов. В рассматриваемом примере $\chi^2 = 42,4$, а число степеней свободы (после объединения классов с частотой $f < 5$) составляет $\nu = 6 - 1 = 5$. Табличное значение $\chi_{05}^2 = 11,07$ (Приложение V) меньше получен-

ного, следовательно, нуль-гипотеза отбрасывается, и разница между сравниваемыми рядами должна быть признана существенной.

В том случае, когда объемы выборок не равны ($n_1 \neq n_2$), расчет критерия χ^2 несколько усложняется необходимостью «взвешивания» частот каждого ряда численностью другой выборки, и формула приобретает следующий вид:

$$\chi^2 = \frac{1}{n_1 \cdot n_2} \sum \frac{(f_1 n_2 - f_2 n_1)^2}{f_1 + f_2}.$$

В табл. 8 показаны необходимые для расчета χ^2 действия при сравнении числа чашелистиков у двух популяций *Trollius europaeus* L. Здесь

$$\chi^2 = \frac{538153,29}{270 \cdot 178} = 11,2,$$

а табличное значение критерия при $v=7-1=6$ составляет $\chi_{0,05}^2=12,592$ (Приложение V). Это означает, что при данном уровне существенности нуль-гипотеза отброшена быть не может и различие между рядами не доказано.

Другой критерий, получивший название «критерий лямбда» (λ), разработан А. Н. Колмогоровым и Н. В. Смирновым. Так же, как и χ^2 , он основан на сравнении частот двух распределений, но в отличие от него использует накопленные частоты (раз-

Таблица 8. Сравнение вариационных рядов числа чашелистиков у *Trollius europaeus* L. (x) в двух популяциях (f_1 и f_2) посредством критерия χ^2

x	f_1	f_2	„Взвешивание“		$f_1 n_2 - f_2 n_1$	$(f_1 n_2 - f_2 n_1)^2$	$\frac{(f_1 n_2 - f_2 n_1)^2}{f_1 + f_2}$
			$f_1 n_2$	$f_2 n_1$			
7	—	2]—	—	—	—	—	—
8	9	11]13	1602	3510	—1908	3640464	165475,63
9	38	27	6704	7290	—526	276676	4256,55
10	73	33	12994	8910	4084	16679056	157349,58
11	69	38	12282	10260	2022	4088484	38210,13
12	48	44	8544	11880	—3336	11128896	120966,26
13	28	17	4984	4590	394	155236	3449,69
	4] 5	3] 6	890	1620	—730	532900	48445,45
	1]—	1]—	—	—	—	—	—
16	—	1]—	—	—	—	—	—
17	—	1]—	—	—	—	—	—
$n_1=270$ $n_2=178$							$\Sigma=538153,29$

дел 1.4.6), применим к выборкам любого объема, не требует объединения частот при $f < 5$ и особенно хорош в случае дискретного типа варьирования признаков, так как от группировки вариант в классы мощность критерия λ , вообще говоря, снижается. Удобством этого критерия служат также простота расчетов и ненадобность в таблицах для определения уровней существенности. Дело в том, что этот критерий имеет всего три независимых от числа степеней свободы пороговых значения, соответствующих обычным уровням существенности: $\lambda_{05} = 1,36$;

Таблица 9. Сравнение вариационных рядов числа междуузлией (x) у популяций *Odontites serotina* Dum. из г. Хаапсалу ЭССР (f_1) и Ст. Петергофа (f_2) посредством критерия λ

x	Частоты		Начетные ряды		$\frac{\Sigma f_{i(1)}}{n_1}$	$\frac{\Sigma f_{i(2)}}{n_2}$	$\left \frac{\Sigma f_{i(1)}}{n_1} - \frac{\Sigma f_{i(2)}}{n_2} \right $
	f_1	f_2	$\Sigma f_{i(1)}$	$\Sigma f_{i(2)}$			
8	1	1	150	197	1,000	1,000	0
9	8	10	149	196	0,993	0,995	0,002
10	11	23	141	186	0,940	0,944	0,004
11	14	24	130	163	0,837	0,827	0,040
12	21	37	116	139	0,773	0,706	0,067
13	23	30	95	102	0,633	0,518	0,115
14	26	28	72	72	0,480	0,365	0,115
15	22	17	46	44	0,307	0,223	0,084
16	12	16	24	27	0,160	0,137	0,023
17	6	4	12	11	0,080	0,056	0,024
18	2	3	6	7	0,040	0,036	0,004
19	2	2	4	4	0,027	0,020	0,007
20	2	2	2	2	0,013	0,010	0,003
$n_1 = 150$		$n_2 = 197$					

$\lambda_{01} = 1,63$ и $\lambda_{001} = 1,95$. Если полученное эмпирическое значение λ превышает одно из выбранных пороговых значений этого критерия, различие между сравниваемыми рядами признается достоверным с соответствующей степенью вероятности ($P > 0,95$; $P > 0,99$ и $P > 0,999$).

Рассмотрим конкретный пример: сравнение числа междуузлией у двух популяций *Odontites serotina* Dum. (табл. 9). Ряды накопленных частот составляются знакомым нам способом сумм (раздел 1.4.6) снизу вверх или сверху вниз, причем последняя сумма, если расчет выполнен верно, должна соответствовать объему выборок. В последнем столбце находят максимальную

разность между накопленными частотами (D_{\max}) и подставляют ее в общую формулу для расчета критерия:

$$\lambda = D_{\max} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}.$$

В нашем примере это дает

$$\lambda = 0,115 \sqrt{\frac{150 \cdot 197}{150 + 197}} = 1,06.$$

Полученное значение $\lambda = 1,06$ не достигает даже 5%-ного порогового уровня существенности ($\lambda_{05} = 1,36$), и поэтому нуль-гипотеза об отсутствии разницы по числу междоузлий у сравниваемых популяций остается в силе. При одинаковом объеме выборок ($n_1 = n_2$) формула для расчета критерия λ упрощается:

$$\lambda = \frac{D_{\max}}{\sqrt{n}}.$$

Глава 2

ТЕОРЕТИЧЕСКИЕ И ЭМПИРИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ

В принципе любое выборочное распределение можно описать такой математической функцией, при построении графика которой кривая пройдет через все эмпирические точки. Поскольку трудно найти хотя бы два совершенно одинаковых распределения, то такого рода функций можно подобрать великое множество. Расчет в каждом отдельном случае своей специфической функции технически весьма сложен, а в биологическом отношении — формален, ибо интерпретация огромного количества различающихся функций и тем более обобщение результатов путем выведения каких-либо закономерностей изменчивости биологических признаков становятся практически невозможными.

Между тем опыт показывает, что в природе существует весьма ограниченное количество типов изменчивости признаков живых организмов, которые могут служить статистическими моделями этого явления в подавляющем большинстве случаев. Именно с ними следует сравнивать эмпирические распределения признаков в выборках.

К числу основных типов теоретических распределений принадлежат: биномиальное распределение Бернулли, распределение Пуассона, нормальное распределение Гаусса — Лапласа,

а также типы кривых Пирсона и некоторые другие. Поскольку первые три типа распределений подробно рассматриваются во всех пособиях по биометрии и математической статистике, мы ограничимся ниже лишь основными положениями и соответствующими примерами.

2.1. БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Этот тип распределения возникает тогда, когда для каждого наблюдения существуют лишь два возможных исхода (альтернативный тип изменчивости признака). Если вероятность одного исхода (значения признака) обозначить буквой p , то вероятность другого исхода составит $q=1-p$, ибо всегда $p+q=1$.

При равной вероятности двух исходов ($p=q=0,5$) распределение подчиняется коэффициентам бинома

$$(0,5 + 0,5)^n = 0,5^n + \frac{n}{1!} 0,5^n + \frac{n(n-1)}{2!} 0,5^n + \dots + 0,5^n,$$

где n — число независимых событий. Если же $p \neq q$, то формула несколько усложняется:

$$(p + q)^n = p^n q^0 + \frac{n}{1!} p q^{n-1} + \frac{n(n-1)}{2!} p^2 q^{n-2} + \dots + p^n q^0.$$

В биологии биномиальное распределение хорошо описывает, например, появление признака (события) в N_n подвыборках (сериях, пробах) одинаковой численности (n) из общего числа наблюдений (N) при альтернативном типе изменчивости. Для вычисления теоретических частот при этом требуется произвести разложение бинома:

$$N_n(p + q)^n = N_n \left(p^n q^0 + \frac{n}{1!} p q^{n-1} + \frac{n(n-1)}{2!} p^2 q^{n-2} + \dots + p^n q^0 \right),$$

где p — доля вариант, обладающих данным признаком; q — доля вариант, им не обладающих.

Рассмотрим пример. Наиболее часто встречающимся (модальным) числом листочков околоцветника у *Anemone nemorosa* L. служит число 6. Было изучено 45 пробных серий ($N_n = 45$) по 10 экземпляров в каждой ($n=10$; $N=450$), причем в каждой серии подсчитывалось количество экземпляров, обладающих модальным значением признака (т. е. 6 листочками околоцветника). Соответствует ли распределение этих экземпляров биномиальному закону? Порядок расчетов приведен в табл. 10. Столбцы 1 и 2 таблицы содержат исходные данные: x — возможное число растений с модальным значением признака в одной серии, f — число соответствующих серий (эмпирические частоты). В столбце 3 находим произведения xf и подсчитываем итоговую сумму, означающую, что 302 растения из 450 имеют модальное значение признака. Отсюда следует, что доля

Таблица 10. Вычисление теоретических частот биномиального распределения

x	f	xf	k	p^x	q^x	$p^x q^x$	$p^x q^x k$	f'	
								$p^x q^x k N$	округленно
0	0	0	1	$p^0=1,00000$	$q^0=0,00002$	0,00002	0,00002	0,0009	0
1	0	0	10	0,67100	0,00005	0,00003	0,00030	0,0135	0
2	1	2	45	0,45024	0,00014	0,00006	0,00270	0,1215	0
3	2	6	120	0,30211	0,00042	0,00013	0,01560	0,7020	1
4	3	12	210	0,20272	0,00127	0,00026	0,05460	2,4570	2
5	5	25	252	0,13602	0,00386	0,00053	0,13356	6,0102	6
6	9	54	210	0,09127	0,01172	0,00107	0,22470	10,1115	10
7	10	70	120	0,06124	0,03561	0,00218	0,26160	11,7720	12
8	6	48	45	0,04109	0,10824	0,00445	0,20025	9,0113	9
9	5	45	10	0,02757	0,32900	0,00907	0,09070	4,0815	4
10	4	40	1	$p^{10}=0,01850$	$q^0=1,00000$	0,01850	0,01850	0,8325	1
Σ	45	302	—	—	—	—	—	45,1139	45

растений с 6 листочками околоцветника составляет $p = 302 : 450 = 0,671$, а доля растений, не обладающих модальным значением этого признака, — $q = 1 - 0,671 = 0,329$.

Следовательно, для вычисления теоретических частот (f') необходимо произвести разложение бинома $45(0,671 + 0,329)^{10}$. В столбце 4 выписываем биномиальные коэффициенты (k) для данной степени бинома ($n=10$), пользуясь «арифметическим треугольником Паскаля» (Приложение VI). В столбцах 5 и 6 значения $p=0,671$ и $q=0,329$ возводятся в степени x (в противоположном порядке!), а в столбцах 7 и 8 находятся произведения $p^x q^x$ и $p^x q^x k$. Умножая цифры столбца 8 на $N_n=45$, получаем значения теоретических частот f' (столбец 9) и округляем их до целого числа (столбец 10). Сумма частот должна соответствовать величине $N_n=45$.

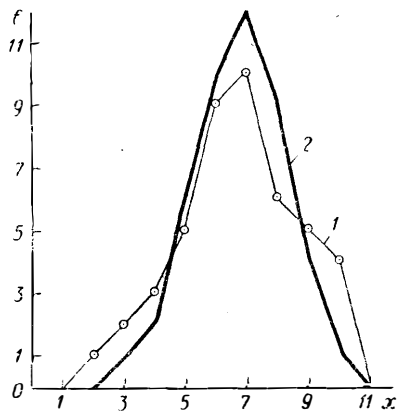


Рис. 2 Число листочков околоцветника у *Anemone nemorosa* L. (по данным табл. 10).

1 — эмпирическое распределение; 2 — соответствующее ему теоретическое биномиальное распределение.

Остается определить, соответствуют ли найденные теоретические частоты (f') эмпирическим (f), т. е. решить вопрос о том, действительно ли наши опытные данные описываются биномиальным распределением? На графике (рис. 2) видно лишь приблизительное совпадение эмпирической и теоретической кривых. Точный ответ можно получить с помощью критерия χ^2 (табл. 11). Полученное значение $\chi^2=5,077$ сравнивается с табличным при числе степеней свободы $\nu=k-2$, где k — число классов. После объединения крайних частот со значениями, меньшими 5, в табл. 11 осталось всего 5 классов, следовательно, $\nu=5-2=3$.

Таблица 11. Расчет критерия χ^2 (по данным табл. 10)

f	f'	$ f-f' $	$(f-f')^2$	$\frac{(f-f')^2}{f'}$
0	0	—	—	—
0	0	—	—	—
1]	0]	—	—	—
2]	1]	—	—	—
3]	2]	—	—	—
5] 11	6] 9	2	4	0.444
9	10	1		0.100
10	12	2	4	0.333
6	9	3	9	1.000
5] 9	4] 5	4	16	3.200
4]	1]	—	—	—
—	—	—	—	$\chi^2=5.077$

При $\nu=3$ табличное значение χ^2 (Приложение V) для 5%-ного уровня существенности составляет $\chi_{05}^2=7,815$, значит, $\chi^2 < \chi_{st}^2$, и эмпирические данные удовлетворительно моделируются биномиальным законом распределения.

Основные выборочные параметры при биномиальном распределении могут быть рассчитаны обычным путем (раздел 1.4.6), однако имеются и специфические для этого распределения формулы

$$M = np \quad \text{и} \quad \sigma = \sqrt{npq}.$$

Так, в рассмотренном выше примере $M=10 \cdot 0,671=6,71$; $\sigma=\sqrt{10 \cdot 0,671 \cdot 0,329}=1,49$, причем обе величины именованные: M — среднее число экземпляров с 6 листочками околоцветника в сериях из 10 растений; σ — изменчивость этого признака.

2.2. РАСПРЕДЕЛЕНИЕ ПУАССОНА

Когда одна из вероятностей очень мала (например, когда $p \rightarrow 0$), биномиальное распределение переходит в распределение Пуассона, которое биологически интерпретируется как распределение редких признаков (событий). Эмпирические частоты распределения Пуассона, как и в биномиальном распределении, могут представлять собой число серий (проб) одинакового объема, имеющих определенную долю особей с изучаемым признаком. Но иногда и обычный вариационный ряд подчиняется закону Пуассона.

В качестве примера возьмем вариационный ряд числа пар ветвей у береговой экологической формы *Odontites serotina* Dum.:

$x:$	0	1	2	3	
$f:$	137	8	4	1	($n=150$).

который на графике (рис. 3) представлен резко асимметричной односкатной кривой. Односкатная кривая характерна для пуассонова распределения, но не является его абсолютным признаком, ибо при увеличении частоты встречаемости признака она постепенно обретает левую ветвь и, будучи уже двускатной, некоторое время продолжает подчиняться закону Пуассона, лишь затем переходя в биномиальную кривую.

Распределение Пуассона характеризуется только одним параметром — средней арифметической (в нашем примере $M=0,13$). Среднее квадратичное отклонение и коэффициент варьирования теряют свой смысл, причем последний может принимать значения, превышающие 100% (в нашем примере $\sigma=0,45$; $V=346\%$).

В случае пуассонова распределения наблюдается примерное равенство меди и дисперсии (у нас $\sigma^2=0,20$ и $\sigma^2 \approx M$), но такое соотношение встречается и при нормальном распределении признака. Поэтому вопрос о том, подчиняется ли эмпирический ряд закону Пуассона, может быть окон-

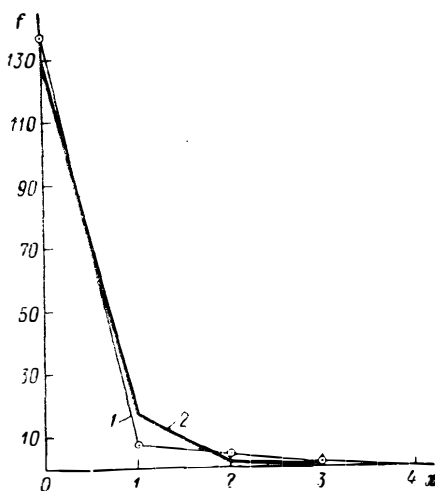


Рис. 3. Число пар ветвей у береговой экологической формы *Odontites serotina* Dum. (по данным табл. 12). 1 — эмпирическое распределение; 2 — соответствующее ему теоретическое распределение Пуассона.

чательно решен только путем расчета теоретических частот и сравнения их с опытными данными.

Вернемся к нашему примеру. Расчет теоретических частот (f') распределения Пуассона по эмпирическим данным выполняется по формуле

$$f' = \frac{M^x}{x!} n e^{-M},$$

где M — медия; x — значение признака; n — объем выборки (если проводится анализ встречаемости признака в сериях, то вместо n берут, как и в случае биномиального распределения, величину N_n , т. е. количество серий); e — основание натуральных логарифмов (функция e^{-M} табулирована: см. Приложение VII). Порядок расчетов показан в табл. 12, первые два столбца которой занимают исходные данные.

Теоретическая кривая близка к эмпирической (рис. 3). Проверку степени их соответствия для разнообразия выполним не с помощью критерия χ^2 , а с помощью критерия λ (раздел 1.5.2),

Таблица 12. Вычисление теоретических частот распределения Пуассона

x	f	M^x	$x!$	$\frac{M^x}{x!}$	f'	
					$\frac{M^x}{x!} n e^{-M}$	округленно
0	137	1,0000	1	1,0000	131,715	132
1	8	0,1300	1	0,1300	17,123	17
	4	0,0169	2	0,0085	1,120	1
3	1	0,0022	6	0,0004	0,053	0
Σ	$n=150$	—	—	—	150,011	$n=150$

Таблица 13. Расчет критерия λ
(по данным табл. 12)

f	f'	Начетные ряды		$ \Sigma f - \Sigma f' $
		Σf	$\Sigma f'$	
137	132	150	150	0
8	17	13	18	5 = D_{\max}
4		5	1	4
1	0	1	0	1
150	150	—	—	—

необходимые данные для расчета которого и порядок вычислений приведены в табл. 13:

$$\lambda = \frac{D_{\max}}{\sqrt{n}} = \frac{5}{\sqrt{150}} = 0,408.$$

Полученное значение $\lambda = 0,408$ не достигает порогового уровня этого критерия ($\lambda_{05} = 1,36$), и мы вправе заключить, что эмпирическое распределение изученного признака соответствует избранной модели — распределению Пуассона.

2.3. НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Если в бинOME $(p+q)^n$ $p=q=0,5$ и $n \rightarrow \infty$, то распределение из дискретного становится непрерывным и переходит в нормальное распределение, являющееся основной теоретической моделью в биометрии, на которой основано большинство методов. Графически (рис. 4) нормальное распределение выражается плавной симметричной куполообразной кривой с асимптотически приближающимися к оси абсцисс ветвями. При этом мода, мода и медиана совпадают ($M = M_o = M_e$).

Уравнение нормального распределения можно записать в различной форме. Приведем лишь одну формулу

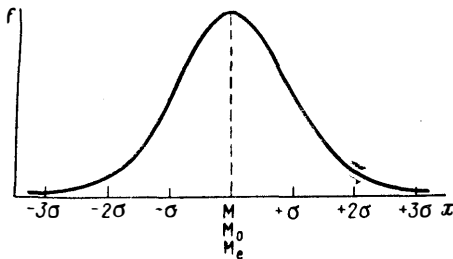


Рис. 4. Кривая нормального распределения (схема).

$$f' = \frac{ni}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-M)^2}{2\sigma^2}},$$

где f' — искомая ордината кривой (теоретическая частота); n — объем выборки; i — классовой интервал; $\pi = 3,1416$; $e = 2,71828$; x — значение признака в выборке, для которого определяется теоретическая частота; M — выборочная медиана; σ — выборочное среднее квадратичное отклонение.

В степень числа e входит величина $\frac{x-M}{\sigma} = t$, получившая название нормированного отклонения, которым мы уже пользовались ранее (раздел 1.4.8) и к которому не раз еще вернемся в дальнейшем. Пользуясь им, можно упростить правый множитель в формуле нормального распределения:

$$e^{-\frac{(x-M)^2}{2\sigma^2}} = e^{-\frac{t^2}{2}} = 2,71828^{-\frac{t^2}{2}}.$$

Легко упростить и левый множитель:

$$\frac{ni}{\sigma\sqrt{2\pi}} = \frac{ni}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} = \frac{ni}{\sigma} \cdot 0,39894.$$

В итоге рабочая формула для определения теоретических частот нормального распределения приобретает вид

$$f' = \frac{ni}{\sigma} \cdot 0,39894 \cdot 2,71828^{-\frac{t^2}{2}}$$

Значения $0,39894 \cdot 2,71828^{-\frac{t^2}{2}}$ табулированы (Приложение VIII) под названием функции

$$\theta(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

(значения этой функции ради удобства называют иногда просто «ординатами нормальной кривой»). Следовательно, задача нахождения теоретических частот заметно облегчается. Практически вычисление их удобнее всего производить так, как показано в следующем примере, где изучалось число лучей первого порядка в соцветиях *Anthriscus sylvestris* (L.) Hoffm. Объем выборки $n=1200$; $M=10,22$; $\sigma=2,31$; классовый интервал $i=1$

Таблица 14. Вычисление теоретических частот нормального распределения

x	f	x-M	t = $\frac{x-M}{\sigma}$	$\theta(t)$	f'	
					$\theta(t) \frac{ni}{\sigma}$	округленно
3	1	-7,22	-3,1255	00298	1,55	1
4	7	-6,22	-2,6926	01071	5,56	6
5	17	-5,22	-2,2597	03103	16,12	16
6	34	-4,22	-1,8268	07477	38,84	39
7	70	-3,22	-1,3939	15183	78,87	79
8	141	-2,22	-0,9610	25164	130,72	131
9	165	-1,22	-0,5281	34667	180,09	180
10	253	-0,22	-0,0952	39695	206,21	206
11	187	0,78	0,3377	37654	195,60	196
12	145	1,78	0,7706	29659	154,07	154
13	85	2,78	1,2035	19419	100,88	101
14	51	3,78	1,6364	10567	54,89	55
15	25	4,78	2,0693	04682	24,32	24
16	10	5,78	2,5022	01753	9,11	9
17	6	6,78	2,9351	00530	2,75	3
18	3	7,78	3,3680	00136	0,71	1
Σ	1200				1200,29	1200

(дискретный тип варьирования признака). Порядок расчетов показан в табл. 14.

Проверку соответствия эмпирических частот вычисленным частотам нормального распределения произведем с помощью критерия χ^2 (табл. 15). Полученное значение $\chi^2=21,184$ надо сравнить со стандартным (Приложение V) при числе степеней свободы $\nu=k-3=10$, где k — число классов сравниваемых распределений, ибо в данном случае имеются 3 ограничения: оди-

Таблица 15. Расчет критерия χ^2 (по данным табл. 14)

f	f'	$f-f'$	$(f-f')^2$	$\frac{(f-f')^2}{f}$
1]	1]	—	—	—
7] 8	6] 7	1	1	0,1429
17	16	1	1	0,0625
34	39	-5	25	0,6410
70	79	-9	81	1,0253
141	131	10	100	0,7634
165	180	-15	225	1,2500
253	206	47	2209	10,7233
187	196	-9	81	0,4133
145	154	-9	81	0,5260
85	101	-16	256	2,5347
51	55	-4	16	0,2909
25	24	1	1	0,0417
10] 19	9] 13	6	36	2,7692
6]	3]	—	—	—
3]	1]	—	—	—
				$\chi^2=21,1842$

наковый для сравниваемых рядов объем выборки, определенные значения меди и среднего квадратичного отклонения ($M=10,22$; $\sigma=2,31$), по которым строится модельное распределение.

Стандартные значения $\chi_{05}^2=18,307$, $\chi_{01}^2=23,209$, т. е. полученное значение $\chi^2=21,184$ лежит между ними. Первое стандартное значение χ^2 отвечает повышенной, а второе — обычной требовательности при сравнении эмпирических рядов с теоретическими. Поскольку рассчитанное значение χ^2 не превышает величину χ_{01}^2 , мы можем считать, что эмпирическое распре-

ление признака удовлетворительно подчиняется нормальному закону распределения (рис. 5). Это не исключает, однако, поиска и более удачной модели.

Вернемся теперь к понятию о нормированном отклонении t — величине, которая широко используется в различных биометрических методиках.

Прежде всего заметим, что в случае нормального распределения значения t колеблются в пределах $\pm 3\sigma$, т. е. возможные отклонения частных значений признака в выборке от медики лежат примерно в этом диапазоне. Точнее, в пределах $\pm 1\sigma$ лежит 68,27% всех вариантов, в пределах $\pm 2\sigma$ — 95,45, в пределах

$\pm 3\sigma$ — 99,73 и в пределах $\pm 4\sigma$ — 99,99% наблюдаемых значений признака. Это означает, что взятая наугад особь имеет следующую вероятность (P) попасть в указанные пределы: $\pm 1\sigma$ — 0,68; $\pm 2\sigma$ — 0,95; $\pm 3\sigma$ — 0,99 и $\pm 4\sigma$ — 0,999. Последним трем вероятностям соответствуют дополнительные до единицы величины (0,05, 0,01 и 0,001 или 5%, 1% и 0,1%), которые и получили название доверительных уровней, или уровней существенности. Утверждая, например, что при нормальном распределении взятая наугад особь окажется по

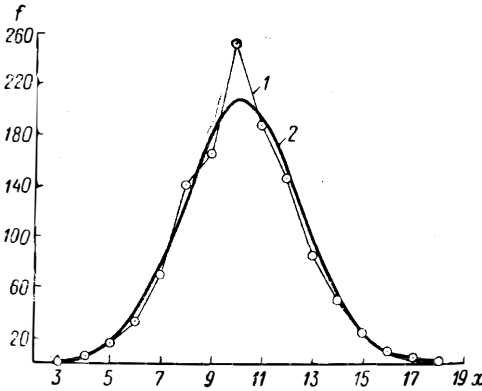


Рис. 5. Число лучей первого порядка в соцветиях *Anthriscus sylvestris* (L.) Hoffm. (по данным табл. 14).

1 — эмпирическое распределение; 2 — соответствующее ему теоретическое нормальное распределение.

данному признаку в пределах $\pm 3\sigma$, мы рискуем ошибиться лишь в одном случае из 100.

Существует возможность легкого вычисления приближенного значения среднего квадратичного отклонения по размаху варьирования признака в случае распределения, не очень сильно отличающегося от нормального:

$$\sigma \approx \frac{x_{\max} - x_{\min}}{6}.$$

Такая оценка может быть, например, полезной, если необходимо определить оптимальный объем выборки (см. с. 6).

Когда численность генеральной совокупности неизвестна, то объем выборки

$$n = \left(\frac{t\sigma}{\Delta} \right)^2,$$

где t — критерий Стьюдента, взятый в зависимости от требуемой точности исследования ($t_{05}=1,96$; $t_{01}=2,58$ и $t_{001}=3,30$); σ — среднее квадратическое отклонение; Δ — допустимая, по мнению исследователя, погрешность в определении значения меди.

С помощью нормированного отклонения t можно существенно уточнить интервал, в котором согласно выборочным данным должен находиться тот или иной генеральный параметр. Например, оценка интервала для генеральной меди выглядит так: $M_{\text{ген}}=M\pm 3m_M$. Эта оценка является излишне осторожной и дает умышленно завышенный интервал. Задавшись требуемым уровнем существенности (Приложение III), мы можем теперь уточнить этот интервал: $M_{\text{ген}}=M\pm tm_M$.

Преимущества такой оценки поясним следующим примером. Для числа междуузлий у *Odontites verna* (Bell.) Dum. были получены следующие выборочные параметры: $M=9,38$, $m_M=0,14$, $n=150$. Грубая оценка дает интервал для генеральной средней $M_{\text{ген}}=9,38\pm 3\cdot 0,14=9,38\pm 0,42$, т. е. устанавливает ее пределы от 8,96 до 9,80 междуузлия. Если же задаться достаточным для наших целей 5%-ным уровнем существенности (при $v=n-1=149$ значение $t_{05}=1,96$), то $M_{\text{ген}}=9,38\pm 1,96\cdot 0,14=9,38\pm 0,27$, и интервал для генеральной средней в действительности оказывается значительно более узким — от 9,11 до 9,65 междуузлия. При необходимости поставить более жесткое условие можно взять 1%-ный уровень существенности ($t_{01}=2,576$). Тогда интервал для генеральной средней несколько расширится, составляя 9,02—9,74 междуузлия, но все же останется более узким, чем в случае грубой оценки.

Роль нормированного отклонения t в биометрии не исчерпывается сказанным, и в нужных случаях мы будем к нему возвращаться.

Заключая этот раздел, следует сделать два дополнения.

1. В случае малого объема выборок ($n\leq 20$) нормальное распределение переходит в так называемое t -распределение Стьюдента, обоснованное позже Р. Фишером. Именно поэтому таблица значений t (Приложение III) имеет вход не по n , а по числу степеней свободы $v=n-1$: эти значения зависят от объема выборки и неодинаковы для одного и того же уровня существенности. Кривая t -распределения отличается от нормальной большей островершинностью и растянутостью своих ветвей.

2. Если эмпирическое распределение напоминает нормальное, но все же имеет заметные асимметрию или эксцесс, то можно в качестве теоретической модели испробовать функции типа A или B Шарлье [Митропольский, 1971] или перейти к вычислению теоретических частот одного из семи типов кривых Пирсона (раздел 2.5).

2.4. ЭМПИРИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ. АСИММЕТРИЯ И ЭКССЕСС

Получаемые исследователем выборочные распределения признаков в той или иной степени отличаются обычно от теоретических моделей и, в частности, от нормального распределения. Это выражается в виде асимметрии или эксцесса вариационной кривой по отношению к кривой нормального распределения (рис. 6, 1), которая всегда симметрична и имеет эксцесс, равный нулю. Асимметрия эмпирической кривой распределения обусловлена тем, что по сторонам от модального значения признака (максимума кривой) располагается неравное число вариантов. Если последних больше справа от моды, то мы имеем

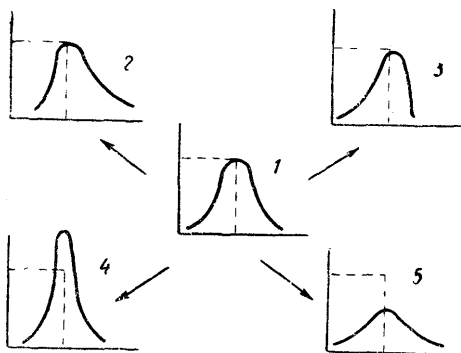


Рис. 6. Схематическое изображение наиболее обычных отклонений эмпирических кривых от нормального распределения.

Пояснения в тексте.

дело с положительной (рис. 6, 2); а если слева — с отрицательной (рис. 6, 3) асимметрией. Когда большинство значений признака группируется возле средней арифметической, возникает положительный эксцесс, и кривая называется островершинной, эксцессивной или leptokurtической (рис. 6, 4). В противоположном случае возможен отрицательный эксцесс — плосковершинная, депрессивная или platykurtическая кривая (рис. 6, 5). При сильном отрицательном эксцессе вместо вершины кривой может возникнуть выемка: кривая становится двухвершинной.

Асимметрия и эксцесс эмпирических кривых в ряде случаев характеризуют важные для исследователя биологические особенности материала, указывая, например, на изменение признака в ходе естественного отбора или в процессе географической или экологической дифференциации вида. Поэтому изучение степени и характера асимметрии и эксцесса вариационных кривых может быть специальной задачей исследования.

Показатели асимметрии (As) и эксцесса (Ex) с их ошибками репрезентативности могут быть вычислены по следующим приближенным формулам:

$$As = \frac{\sum(x - M)^3}{n \sigma^3}; \quad m_{As} = \sqrt{\frac{6}{n + 3}};$$

$$Ex = \frac{\sum(x - M)^4}{n \sigma^4} - 3; \quad m_{Ex} = 2 \sqrt{\frac{6}{n + 5}}.$$

Отклонение эмпирического ряда по асимметрии и эксцессу от нормального распределения считается существенным, если эти показатели более чем в три раза превышают свою ошибку.

2.5. МЕТОД МОМЕНТОВ И ТИПЫ КРИВЫХ ПИРСОНА

К. Пирсон установил 7 типов теоретических кривых распределения признаков у растений и животных. Выведению уравнений типов своих кривых Пирсон предпослал разработку метода статистических моментов, на основных положениях которого мы и задержим наше внимание.

Статистическим моментом i -го порядка (M_i) называется сумма произведений частот (f) на i -ю степень отклонения всех наблюдений (x) от некоторой постоянной величины (x'), отнесенная к числу наблюдений (n):

$$M_i = \frac{\sum f(x - x')^i}{n}.$$

Момент называется условным и обозначается \bar{v}_i , если за величину x' принята условная средняя (раздел 1.4.6), т. е. если $x' = A$. Если же за x' принята средняя арифметическая, т. е. $x' = M$, то момент называется центральным и обозначается μ_i .¹ Приведем соответствующие формулы условных и центральных моментов первых четырех порядков:

$$\begin{aligned} \bar{v}_1 &= \frac{\sum f(x - A)}{n}; & \mu_1 &= \frac{\sum f(x - M)}{n}; \\ \bar{v}_2 &= \frac{\sum f(x - A)^2}{n}; & \mu_2 &= \frac{\sum f(x - M)^2}{n}; \\ \bar{v}_3 &= \frac{\sum f(x - A)^3}{n}; & \mu_3 &= \frac{\sum f(x - M)^3}{n}; \\ \bar{v}_4 &= \frac{\sum f(x - A)^4}{n}; & \mu_4 &= \frac{\sum f(x - M)^4}{n}. \end{aligned}$$

Сравнивая приведенные формулы с формулами средней арифметической, среднего квадратичного отклонения, показателей асимметрии и эксцесса (разделы 1.4.6 и 2.4), можно заключить, что:

1. Условный момент первого порядка (\bar{v}_1) входит в формулы для вычисления средней арифметической как способом произведений, так и способом сумм.

2. В соответствии с первым свойством медианы (раздел 1.4.5) центральный момент первого порядка равен нулю: $\mu_1 = 0$.

¹ Существуют и другие типы моментов, на которых мы не останавливаемся.

3. Центральный момент второго порядка есть дисперсия, т. е. квадрат среднего квадратичного отклонения: $\mu_2 = \sigma^2$.

4. Формулы для вычисления показателей асимметрии (As) и эксцесса (Ex) значительно упрощаются, если их выразить в моментах:

$$As = \frac{\mu_3}{\sigma^3}; \quad Ex = \frac{\mu_4}{\sigma^4} - 3.$$

Вычисление центральных моментов по приведенным ранее формулам достаточно громоздко, тогда как условные моменты рассчитываются гораздо проще. Поэтому на практике определяют сначала условные моменты, а затем переходят к центральным с помощью следующих формул, отражающих существующую между ними связь:

$$\begin{aligned} \mu_2 &= \bar{v}_2 - \bar{v}_1^2; & \mu_3 &= \bar{v}_3 - 3\bar{v}_1\bar{v}_2 + 2\bar{v}_1^3; \\ \mu_4 &= \bar{v}_4 - 4\bar{v}_1\bar{v}_3 + 6\bar{v}_1^2\bar{v}_2 - 3\bar{v}_1^4. \end{aligned}$$

Пирсон ввел также показатели β_1 , β_2 , характеризующие отношение между центральными моментами:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}; \quad \beta_2 = \frac{\mu_4}{\mu_2^2}.$$

«Беты Пирсона» указывают на асимметрию и эксцесс кривой. Первая из них служит мерилем асимметрии: при $\beta_1 = 0$ кривая симметрична, при $\beta_1 > 0$ — асимметрична, а вторая отражает эксцесс кривой: при $\beta_2 = 3$ кривая нормальна, при $\beta_2 > 3$ — лептокуртична и при $\beta_2 < 3$ — платикуртична. Следовательно, показатель эксцесса можно выразить еще одной простой формулой: $Ex = \beta_2 - 3$.

Значение величин β_1 и β_2 далеко не исчерпывается сказанным. Именно опираясь на них, Пирсон и вывел уравнения семи типов своих кривых. Для этого им были предложены критерии «каппа-1» (κ_1) и «каппа-2» (κ_2):

$$\begin{aligned} \kappa_1 &= 2\beta_2 - 3\beta_1 - 6; \\ \kappa_2 &= \frac{\beta_1(\beta_2 + 3)^2}{4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)}. \end{aligned}$$

Принадлежность исследуемой кривой к одному из семи типов можно определить, исходя из диаграммы Эльдертона (рис. 7). Из этой диаграммы видно, что основными типами кривых Пирсона следует считать I, IV и VI, представленные отрезками, тогда как все остальные типы, изображенные на диаграмме, являются переходными.

Критерий κ_1 характеризует поведение кривой на концах: при $\kappa_1 < 0$ кривая двусторонне ограничена, а при $\kappa_1 > 0$ — дву-

сторонне асимптотична. Значению $\kappa_2=0$ на диаграмме Эльдертона соответствуют сразу два типа кривых Пирсона и нормальное распределение. Для их распознавания приходится обратиться к величине β_2 : при $\beta_2 < 3$ кривую относят ко II типу (отрицательный эксцесс), при $\beta_2 = 3$ кривая принадлежит к нормальному типу распределения и при $\beta_2 > 3$ — к VII типу кривых Пирсона (положительный эксцесс). Надо также иметь в виду, что κ_2 , как всякий показатель, имеет свою ошибку репрезентативности. Поэтому при близких к 0 значениях κ_2 ее следует приравнять к нулю, а если κ_2 достаточно велика, то относить кривую к III типу Пирсона.

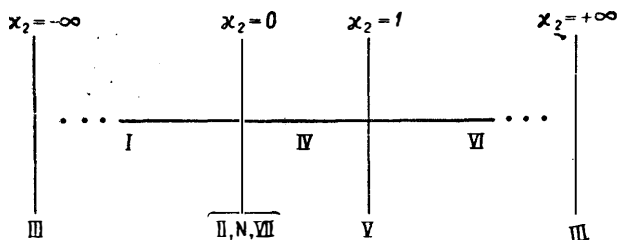


Рис. 7. Диаграмма Эльдертона.
Пояснения в тексте.

Изложенного достаточно для классификации кривых. Определим, например, к какому типу кривых Пирсона принадлежит вариационный ряд числа рылец пестика у *Papaver somniferum* L. (табл. 16). В этой таблице рассчитаны необходимые для определения моментов произведения частот на отклонения от условной средней $f(x-A)^i = fa^i$. Порядок дальнейшей работы следующий.

1. Находим условные моменты первых четырех порядков:

$$\bar{v}_1 = \frac{\sum fa}{n} = \frac{37}{468} = 0,0790; \quad \bar{v}_2 = \frac{\sum fa^2}{n} = \frac{753}{468} = 1,6090;$$

$$\bar{v}_3 = \frac{\sum fa^3}{n} = \frac{85}{468} = 0,1816; \quad \bar{v}_4 = \frac{\sum fa^4}{n} = \frac{4737}{468} = 10,1217.$$

2. Переходим от условных моментов к центральным:

$$\mu_2 = \bar{v}_2 - \bar{v}_1^2 = 1,6090 - 0,0790^2 = 1,6028;$$

$$\mu_3 = \bar{v}_3 - 3\bar{v}_1\bar{v}_2 + 2\bar{v}_1^3 = 0,1816 - 3 \cdot 0,0790 \cdot 1,6090 + 2 \cdot 0,0790^3 = -0,1987;$$

$$\mu_4 = \bar{v}_4 - 4\bar{v}_1\bar{v}_3 + 6\bar{v}_2\bar{v}_1^2 - 3\bar{v}_1^4 = 10,1217 - 4 \cdot 0,0790 \cdot 0,1816 + 6 \cdot 1,6090 \cdot 0,0790^2 - 3 \cdot 0,0790^4 = 10,1241.$$

Заметим, что при вычислении моментов надо брать не менее четырех десятичных знаков, так как в процессе округлений быстро набегают ошибки, могущая исказить окончательный результат.

3. Попутно получим медию и среднее квадратичное отклонение:

$$M = A + i\bar{v}_1 = 11 + 1 \cdot 0,0790 = 11,08;$$

$$\sigma = i\sqrt{v_1^2} = \sqrt{1,6028} = 1,27.$$

4. Находим показатели β_1 и β_2 Пирсона:

$$\beta_1 = \frac{v_3^2}{v_2^3} = \frac{-0,1987^2}{1,6028^3} = 0,0096,$$

откуда делаем вывод, что кривая практически симметрична, и

$$\beta_2 = \frac{v_4}{v_2^2} = \frac{10,1241}{1,6028^2} = 3,9408,$$

откуда ясно, что кривая имеет небольшой положительный эксцесс, т. е. слегка лептокуртична.

5. Рассчитываем показатель χ_2 Пирсона:

$$\begin{aligned} \chi^2 &= \frac{\beta_1(\beta_2 + 3)^2}{4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)} = \\ &= \frac{0,0096(3,9408 + 3)^2}{4(4 \cdot 3,9408 - 3 \cdot 0,0096)(2 \cdot 3,9408 - 3 \cdot 0,0096 - 6)} = 0,004. \end{aligned}$$

Таблица 16. Вариационный ряд числа рылец пестика (x) у *Paraver somniferum* L. и его обработка для последующего вычисления статистических моментов

x		a	fa	fa^2	fa^3	fa^4
6	1	-5	-5	25	-125	625
7	2	-4	-8	32	-128	512
8	8	-3	-24	72	-216	648
9	31	-2	-62	124	-248	496
10	94	-1	-94	94	-94	94
$A=11$	170	0	0	0	0	0
12	111	+1	+111	111	+111	111
13	37	+2	+74	148	+296	592
14	11	+3	+33	99	+297	891
15	3	+4	+12	48	+192	768
		$n=468$	$\sum fa = +37$	$\sum fa^2 = +753$	$\sum fa^3 = +85$	$\sum fa^4 = +4737$

6. Как видим, практически $\kappa_2=0$, и в соответствии с диаграммой Эльдертона рассматриваемую кривую следует отнести к VII типу кривых Пирсона.

Определив тип кривой, можно по составленным Пирсоном уравнениям рассчитать теоретические частоты. Поскольку потребность в таких расчетах возникает не столь уж часто, а сами они весьма длительны и занимают много места, мы отсылаем интересующихся этим вопросом к другому пособию [Митропольский, 1971].

Глава 3

ИЗУЧЕНИЕ ВЛИЯНИЯ ФАКТОРОВ СРЕДЫ НА ИЗМЕНЧИВОСТЬ ПРИЗНАКОВ (ДИСПЕРСИОННЫЙ АНАЛИЗ)

3.1. ЛОГИЧЕСКАЯ СХЕМА ДИСПЕРСИОННОГО АНАЛИЗА

Наблюдаемая в опытной группе растений или в выборке из природной популяции изменчивость изучаемого признака является следствием реакции особей на воздействие множества факторов внешней и внутренней среды. Одновременное выявление степени влияния нескольких факторов при помощи обычных методов практически невозможно.

Дисперсионный анализ, основы которого были разработаны Р. А. Фишером в 1920—1930 гг., позволяет устанавливать не только степень одновременного влияния на признак нескольких факторов и каждого в отдельности, но также их суммарное влияние в любых комбинациях и дополнительный эффект от сочетания разных факторов. Разумеется, и в этом случае остается значительное количество неучтенных факторов, но, во-первых, методика позволяет оценить долю их влияния на общую изменчивость признака, а во-вторых, биолог обычно имеет возможность выделить несколько ведущих факторов и исследовать именно их воздействие на изменчивость признаков.

Дисперсионный анализ позволяет решать множество задач из области теоретической и практической ботаники (систематики, анатомии, генетики, физиологии, селекции, опытного дела, сельскохозяйственной практики), когда требуется изучить воздействие природных или искусственно создаваемых факторов на интересующий исследователя признак. С помощью этого метода можно, например, установить степень влияния экологических

или географических факторов на выраженность тех или иных признаков в популяциях, т. е. проследить экологическую или географическую дифференциацию вида.

Дисперсионный анализ принадлежит к числу довольно трудоемких биометрических методов, однако правильная организация опыта или сбора данных в природных условиях существенно облегчает вычисления. Это означает, что дисперсионный анализ учит исследователя правильно планировать свою работу.

В зависимости от числа учитываемых факторов дисперсионный анализ может быть одно-, двух-, трех- и многофакторным. Техника расчетов при этом несколько различна, но логическая схема остается единой. Объем работы с увеличением числа факторов резко возрастает, поэтому уже четырехфакторный анализ следует производить с помощью ЭВМ. Однако во многих случаях даже одно-двухфакторный анализ, сравнительно легко выполняемый вручную, может принести большую пользу.

Таблица 17. Пример простейшего дисперсионного комплекса

Повторности	Урожайность, ц/га		
	A_1	A_2	A_3
1	30	35	37
2	35	36	36
3	34	37	38
n_x	3	3	3
M_x	33	36	37

$M=35,3; n=9$

Итак, задача дисперсионного анализа состоит в том, чтобы выявить ту часть общей изменчивости признака, которая обусловлена воздействием учитываемых факторов, и оценить достоверность делаемого вывода. Учитываемые факторы мы будем обозначать буквой x , неучитываемые — буквой z , а все факторы вместе — буквой y (или припиской этих букв к соответствующим символам). Неучитываемые факторы составляют «шум» — помехи, мешающие выделить степень влияния учитываемых факторов. Непременным условием дисперсионного анализа является разбивка каждого учитываемого фактора не менее чем на две качественные или количественные градации.

Для пояснения логической схемы дисперсионного анализа рассмотрим простейший произвольный пример. Предположим, что изучается влияние возрастающих доз определенного типа удобрения (фактор A) на урожайность какой-либо полевой культуры. Возрастающие дозы удобрения (градации фактора A) обозначим как A_1, A_2, A_3 . Допустим далее, что опыт был заложен в трех повторностях (на трех делянках) по каждой градации. Результаты опыта сведены в таблицу, которая называется дисперсионным комплексом (табл. 17). В этой таблице рассчитаны: n_x — число наблюдений по градациям; M_x — средние арифметические по градациям; M — общая средняя для всех ($n=9$) наблюдений, значащихся в таблице. Требуется опреде-

лить, влияет ли увеличение дозы удобрения на повышение урожайности, а если влияет, то достоверен ли этот вывод настолько, чтобы можно было дать практические рекомендации сельскому хозяйству. Казалось бы, на этот вопрос можно ответить положительно, ибо средние арифметические (M_x) возрастают в направлении от A_1 к A_3 . Учитывая, однако, небольшое число наблюдений и не такую уж большую разницу в средней урожайности (особенно при дозах удобрения A_2 и A_3), не следует топиться с выводами.

При рассмотрении табл. 17 можно отметить три типа изменчивости изучаемого признака:

1) факториальную (или групповую) изменчивость, характеризующуюся тем, что для каждой из трех градаций имеется своя средняя арифметическая (M_x). Разница в медиане зависит, очевидно, от разного действия фактора — доз удобрения A_1 , A_2 и A_3 ;

2) остаточную изменчивость, характеризующуюся различными значениями признака внутри каждой градации. Эти различия не зависят от влияния фактора, так как проявляются при одной и той же дозе удобрения. Видимо, их причина лежит вне опыта, определяется неучитываемыми в данном анализе факторами;

3) общую изменчивость, заключающуюся в том, что все 9 наблюдений дисперсионного комплекса отличаются друг от друга (или иногда совпадают).

Логично сделать вывод, что общая изменчивость представляет собой результат наложения друг на друга факториальной и остаточной изменчивости, т. е. их сумму. Этот тезис и является центральной идеей дисперсионного анализа. Остается выразить ее математически и разработать схему расчетов.

Известно, что мерилем изменчивости признака в выборке служит сумма квадратов отклонений его значений от средней арифметической $\Sigma(x-M)^2$. Эта величина, отнесенная к числу наблюдений, дает меру рассеяния, именуемую дисперсией, или вариансой (раздел 1.4.5), которая и применяется в дисперсионном анализе.

С помощью сумм квадратов отклонений удобно выразить соотношения между отмеченными выше тремя типами изменчивости.

1. Мерой факториальной изменчивости будет сумма квадратов отклонений средних значений групп (M_x) от общего среднего (M): $S_x^2 = \Sigma(M_x - M)^2$.

2. Мера остаточной изменчивости выразится суммой квадратов отклонений всех наблюдений в данной группе (градации) от среднего значения группы: $S_{yz}^2 = \Sigma(x - M_x)^2$.

3. Мерой общей изменчивости является сумма квадратов отклонений всех наблюдений в дисперсионном комплексе от общего среднего: $S^2 = \Sigma(x - M)^2$.

Тогда в соответствии с основной идеей дисперсионного анализа можно записать: $S_y^2 = S_x^2 + S_z^2$ или

$$\sum(x - M)^2 = \sum(M_x - M)^2 + \sum(x - M_x)^2.$$

Подчеркнем, что правильное применение метода дисперсионного анализа предполагает нормальное или близкое к нему распределение той совокупности, из которой по общим правилам делаются выборки. При этом выборки (группы) должны иметь примерно одинаковую дисперсию (σ^2). Желательно также обеспечить одинаковое число наблюдений (n_x) по градациям факторов.

3.2. ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

Вернемся к рассмотрению нашего произвольного примера (табл. 17). Для облегчения вычислительной работы иногда полезно произвести некоторые преобразования, а именно:

1) от всех значений признака в дисперсионной таблице можно отнять одно и то же число K . После выполнения всех расчетов поправок потребуют только средние (M_x и M): к ним придется прибавить ранее отнятое число;

2) все значения признака можно разделить на одно и то же число K , например перейти от миллиметров к сантиметрам, т. е. разделить все значения на 10. В окончательных подсчетах значения всех сумм квадратов отклонений и дисперсий придется умножить на K^2 , а M_x и M — на K ;

3) все значения признака можно умножить на одно и то же число K . Поправки при этом будут обратными тому, что сказано о делении.

Таблица 18. Преобразованный однофакторный дисперсионный комплекс (по данным табл. 17)

Повторности	Урожайность, ц/га		
	A_1	A_2	A_3
1	0	5	7
2	5	6	6
3	4	7	8
Σ	9	18	
n_x	3	3	3
M_x	3	6	7

$$M=5.3; n=9; \Sigma x=48$$

В рассматриваемом примере удобно от каждой цифры отнять число 30. Тогда после дополнительного расчета сумм значений признака по градациям (Σ) и общей суммы (Σx), необходимых для дальнейшей работы, дисперсионный комплекс приобретает следующий вид (табл. 18).

Последующие расчеты следует вести по рабочим формулам, вытекающим из логической схемы дисперсионного анализа и упрощающим технику вычислений. Общая сумма квадратов отклонений S_y^2 может быть выражена несколько иначе, чем это было сделано ранее, а именно

$$S_y^2 = \sum (x - M)^2 = \sum x^2 - nM^2,$$

но, так как $M = \frac{\sum x}{n}$, то в итоге имеем

$$S_y^2 = \sum x^2 - \frac{(\sum x)^2}{n},$$

а по этой формуле уже значительно проще рассчитать искомую величину.

Проследим порядок расчетов, исходя из данных табл. 18.

1. Прежде всего находят общий поправочный член:

$$H = \frac{(\sum x)^2}{n} = \frac{(0 + 5 + 4 + 5 + 6 + 7 + 7 + 6 + 8)^2}{9} = 256.$$

2. Затем определяют общую сумму квадратов отклонений:

$$S_y^2 = \sum x^2 - H = (5^2 + 4^2 + 5^2 + 6^2 + 7^2 + 7^2 + 6^2 + 8^2) - 256 = 44.$$

3. Находят факториальную сумму квадратов отклонений:

$$\begin{aligned} S_x^2 &= \frac{(\sum A_1)^2}{n_{A_1}} + \frac{(\sum A_2)^2}{n_{A_2}} + \dots + \frac{(\sum A_i)^2}{n_{A_i}} - H = \\ &= \frac{9^2}{3} + \frac{18^2}{3} + \frac{21^2}{3} - 256 = 26. \end{aligned}$$

4. Рассчитывают остаточную сумму квадратов отклонений:

$$S_z^2 = S_y^2 - S_x^2 = 44 - 26 = 18.$$

5. Вычисляют факториальную и остаточную дисперсии, как меры соответствующих типов изменчивости признака в дисперсионном комплексе:

$$\sigma_x^2 = \frac{S_x^2}{v_x}; \quad \sigma_z^2 = \frac{S_z^2}{v_z}.$$

В этих формулах фигурируют степени свободы (v_x , v_z), ибо дисперсия (σ^2) и есть сумма квадратов отклонений в расчете на одну степень свободы. Число степеней свободы для факториальной дисперсии равно числу градаций (групп) фактора без

единицы, так как все группы связаны друг с другом лишь одним общим условием — значением средней арифметической всего дисперсионного комплекса (M). В нашем примере имеются три градации фактора A , поэтому $v_x = 3 - 1 = 2$. Число степеней свободы для остаточной дисперсии равно числу наблюдений в комплексе минус число групп (градаций фактора), ибо все n наблюдений связаны наличием в каждой группе своей средней арифметической (M_x). В рассматриваемом примере $v_z = 9 - 3 = 6$. Число степеней свободы для вычисления общей дисперсии всего комплекса равно числу наблюдений в комплексе без единицы, ибо в этом случае все наблюдения связаны только одним общим условием — наличием общей средней (M): $v_y = 9 - 1 = 8$.

Вообще, если обозначить через n число наблюдений в однофакторном дисперсионном комплексе, а через k — число градаций учитываемого фактора, то

$$v_x = k - 1; \quad v_z = n - k; \quad v_y = n - 1.$$

Обращаясь к разбираемому примеру, находим

$$\sigma_x^2 = \frac{26}{2} = 13; \quad \sigma_z^2 = \frac{18}{6} = 3.$$

6. Рассчитывают доли влияния учтенного и неучтенных в анализе факторов (η_x^2 и η_z^2) как отношения соответствующих сумм квадратов отклонений:

$$\eta_x^2 = \frac{S_x^2}{S_y^2} \quad \text{и} \quad \eta_z^2 = \frac{S_z^2}{S_y^2}.$$

Эти величины представляют собой не что иное, как квадраты корреляционных отношений (раздел 4.3.1) — показателей нелинейной связи между явлениями (признаками). Здесь дисперсионный анализ соприкасается с корреляционным. Применительно к нашему примеру имеем

$$\eta_x^2 = \frac{26}{44} = 0,59 \text{ (или 59\%); } \eta_z^2 = \frac{18}{44} = 0,41 \text{ (или 41\%).}$$

Заметим, что в сумме эти показатели всегда должны составлять единицу (100%).

Теперь уже можно ответить на один из интересующих нас вопросов: учитываемый фактор (разные дозы удобрения) ответствен за 59% изменчивости результативного признака (урожайности), тогда как на долю неучтенных в анализе факторов падает 41%. Но мы еще не знаем, насколько достоверен этот вывод. Ответ на этот вопрос дает заключительный этап дисперсионного анализа.

7. Определяют значение критерия Фишера (F), представляющего собой отношение двух дисперсий (раздел 1.5.1) — факторальной и остаточной:

$$F = \frac{\sigma_x^2}{\sigma_z^2} = \frac{13}{3} = 4,33,$$

и сравнивают его с табличным (Приложение IV). Напомним, что в эту таблицу входят по степеням свободы в зависимости от величины дисперсий (в таблице v_1 соответствует большей, а v_2 — меньшей дисперсии). На пересечении столбца $v_1 = v_x = 2$ и строки $v_2 = v_z = 6$ находим два числа, верхнее из которых соответствует стандартному значению $F_{05} = 5,14$, а нижнее — $F_{01} = 10,92$.

Поскольку $F < F_{05}$, мы не имеем права отбросить нуль-гипотезу: несмотря на то, что 59% изменчивости урожайности вызвано различием в дозах удобрения, это справедливо лишь для данного опыта и не может быть распространено на другие подобные случаи.

Если бы критерий Фишера подтвердил достоверность влияния повышения дозы удобрения на урожайность, то вопрос о том, какую именно дозу (A_2 или A_3) следует рекомендовать, решался бы сравнением средних арифметических по этим градациям при помощи критерия Тьюки (см.: [Лакин, 1980, с. 227]).

Рассмотренный пример был очень простым: по каждой градации фактора мы имеем всего лишь три наблюдения. На практике значения признака часто бывают представлены множеством наблюдений или вариационными рядами. В этих случаях обработка дисперсионных комплексов ведется в табличной форме несколько иным путем, хотя логическая схема анализа остается прежней. Покажем это на следующем примере.

При изучении естественного отбора в популяциях *Odontites serotina* Dum. в условиях проникновения их в пределы береговой полосы в западной Эстонии было подмечено уменьшение числа междуузлий по мере приближения к воде. Число междуузлий является важным диагностическим признаком данного вида, а его уменьшение ведет к конвергенции с другим видом — *O. litoralis* Fr., типичным обитателем береговой полосы (отселектированные отбором популяции первого вида нередко ошибочно относились ботаниками ко второму виду). Очень важно было поэтому строго доказать влияние близости моря на число междуузлий у *O. serotina* Dum.

В описываемых условиях факторами отбора служат засоление субстрата и раннее осеннее затопление береговой полосы, прерывающее цветение и воспроизведение позднеспелых особей *O. serotina* Dum., обладающих большим числом междуузлий. Поначалу не было возможности учесть влияние каждого фактора в отдельности. Поэтому оба они условно были объеди-

нены в один (затопление+засоление) фактор, градациями которого служила степень удаления популяций от воды.

Дисперсионному анализу были подвергнуты 5 популяций, произрастающих на следующем удалении от берега моря: A_1 —300 м, A_2 —250, A_3 —200, A_4 —150 и A_5 —30. Проследим порядок обработки соответствующего однофакторного дисперсионного комплекса (табл. 19).

Таблица 19. Однофакторный дисперсионный анализ влияния удаленности от берега моря на число междуузлий (x) у *Odontites serotina* Dum.

x	a	A_1	A_2	A_3	A_4	A_5	n_y	$a_1=459$	$a_2=687$
20	+9	1	—	—	—	—		1	1
19	+8	2	—	—	—	—	2	3	4
18	+7	3	1	—	—	—	4	7	11
17	+6	5	3	—	—	—	8	15	26
16	+5	7	8	—	—	—	15	30	56
	+4	8	11	—	—	—	19	49	105
14	+3	12	12	—	—	1	25	74	179
13	+2	21	16	11	2	2	52	126	303
12	+1	16	21	21	5	5	68	194	—
$A=11$	0	12	10	35	10	10	77	—	—
10	—1	8	8	16	15	15	62	229	—
9	—2	4	6	11	31	21	73	167	291
8	—3	1	3	5	23	32	64	94	124
7	—4	—	1	1	14	14	30	30	30
n_x		100	100	100	100	100	500	$b_1=520$	$b_2=445$
$\sum fa$		+218	+165	—14	—193	—197	$\sum \sum fa = -21$		
$(\sum fa)^2$		47524	27225	196	37249	38809			
$h = (\sum fa)^2 : n_x$		475,24	272,25	1,96	372,49	388,09	$\sum h = 1510,03$		

1. Складывая частоты по градациям, рассчитывают суммарный ряд n_y и, выбрав условную среднюю (в примере $A=11$), делают отсчеты от нее (a) во втором столбце таблицы.

2. В нижней части таблицы производят неполную обработку каждого из рядов способом произведений, получая $\sum fa$, $(\sum fa)^2$ и вспомогательную величину h . Суммируя значения $\sum fa$, получают и записывают правее величину $\sum \sum fa$. Аналогично поступают с величинами h , получая $\sum h$.

3. В правой части таблицы производят неполную обработку суммарного ряда n_y способом сумм, ограничиваясь двумя колонками, т. е. вычислением сумм a_1 , a_2 , b_1 и b_2 .

4. Рассчитывают три вспомогательные величины:

$$s_1 = a_1 - b_1 = 499 - 520 = -21 \quad (\text{для проверки: } s_1 = \sum \sum fa);$$

$$s_2 = a_1 + b_1 + 2(a_2 + b_2) = 499 + 520 + 2(687 + 445) = 3283;$$

$$H = \frac{s_1^2}{n} = \frac{(-21)^2}{500} = 0,88.$$

5. По рабочим формулам получают необходимые суммы квадратов отклонений:

$$S_y^2 = s_2 - H = 3283 - 0,88 = 3282,12;$$

$$S_z^2 = s_2 - \sum h = 3283 - 1510,03 = 1772,97;$$

$$S_x^2 = \sum h - H = 1510,03 - 0,88 = 1509,15.$$

6. Полученные данные сводят в итоговую таблицу (табл. 20), одновременно производя в ней оставшиеся расчеты. Из таблицы видно, что влияние учтенного фактора (степень удаления популяций от воды) на результативный признак (число междуузлий) составляет 46 % и высоко достоверно (значение $F=105,39$ во много раз превышает стандартные значения $F_{05}=2,39$ и $F_{01}=3,36$), несмотря на то, что влияние неучтенных в анализе факторов сильнее (54 %).

Таблица 20. Результаты однофакторного дисперсионного анализа (по данным табл. 19)

Показатели	Формулы	Учитываемый фактор, x	Неучитываемые факторы, z	Все факторы, y
Суммы квадратов отклонений	S_i^2	1509,15	1772,97	3282,12
Степень влияния фактора . . .	$\gamma_{iy}^2 = \frac{S_i^2}{S_y^2}$	0,46	0,54	1,00
Число степеней свободы	ν_i	4	495	499
Дисперсии	$\sigma_i^2 = \frac{S_i^2}{\nu_i}$	377,29	3,58	ν_x 4
Критерий Фишера	$F = \frac{\sigma_x^2}{\sigma_z^2}$	105,39	—	ν_y 2,39 (400) 3,36

Примечание. В правом нижнем углу помещена выписка из таблицы стандартных значений критерия Фишера (Приложение IV). Формулы для определения числа степеней свободы даны на с. 58.

3.3. ДВУХФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

Переходя к двухфакторному анализу (тем более это касается многофакторного анализа), введем некоторые ограничения. Основное из них состоит в том, что включаемые в дисперсионный анализ факторы должны быть независимы друг от друга, корреляция между ними недопустима. Нельзя, например, изучать одновременно влияние температуры и влажности воздуха на урожайность какой-либо культуры, ибо температура и влажность воздуха обычно сильно коррелируют.

Крайне желательно, чтобы число наблюдений по градациям было одинаковым или хотя бы пропорциональным. В противном случае расчеты сильно усложняются. В соответствии с этим требованием, которое учит правильно планировать эксперимент или наблюдения в природе, дисперсионные комплексы подразделяются на равномерные, пропорциональные и неравномерные (табл. 21).

Таблица 21. Типы дисперсионных комплексов
(числа наблюдений произвольные)

Дисперсионный комплекс	Сочетания градаций факторов				Соотношение численностей			
	A ₁		A ₂		натуральное		относительное	
	B ₁	B ₂	B ₁	B ₂	A ₁ : A ₂	B ₁ : B ₂	A ₁ : A ₂	B ₁ : B ₂
Равномерный	40	40	40	40	80 : 80	80 : 80	1 : 1	1 : 1
Пропорциональный . . .	30	120	120	480	150 : 600	150 : 600	1 : 4	
Неравномерный	20	180	80	220	200 : 300	100 : 400	2 : 3	1 : 4

В равномерных и пропорциональных комплексах действует основное соотношение дисперсионного анализа $S_y^2 = S_x^2 + S_z^2$, но в неравномерных комплексах это соотношение нарушается и возникает необходимость введения поправок на неравномерность [Плохинский, 1970, с. 234], расчет которых отнимает много времени.

Расчеты в двухфакторном анализе ведутся в основном аналогично тому, как это делалось в однофакторном анализе, однако добавляются некоторые новые действия. В двухфакторном анализе изучается раздельное влияние на признак фактора *A* и фактора *B*, их суммарное влияние (*A+B*) и влияние их сочетания (*AB*), которое характеризует дополнительный эффект от совместного одновременного воздействия факторов. В связи со сказанным в основную формулу дисперсионного анализа придется внести некоторые изменения.

Факториальная сумма квадратов отклонений (S_x^2) распадается на три составные части:

$$S_x^2 = S_A^2 + S_B^2 + S_{AB}^2,$$

а сама основная формула приобретает вид

$$S_y^2 = S_A^2 + S_B^2 + S_{AB}^2 + S_z^2.$$

Таким образом, по сравнению с однофакторным анализом задача усложняется необходимостью нахождения величин S_A^2 , S_B^2 и S_{AB}^2 .

Рассмотрим конкретный пример. Ранее (табл. 19) мы анализировали влияние удаленности от берега моря на число междуузлий у *Odontites serotina* Dum. Дополнительный сбор материала позволил расчленить этот обобщенный фактор на его естественные составляющие — затопление и засоление. Теперь каждый фактор представлен двумя градациями:

- Затопление (A) ← [_____ A₁—территория затопляется.
 _____ A₂—территория не затопляется.
- Засоление (B) ← [_____ B₁—субстрат засолен.
 _____ B₂—субстрат не засолен.

Четырем возможным сочетаниям градаций двух факторов соответствовали выборки из популяций этого вида, сделанные в следующих местообитаниях: A₁B₁ — берег моря у г. Хаапсалу ЭССР, A₁B₂ — пойма р. Луги Ленинградской области, A₂B₁ — солончаки Приазовья и A₂B₂ — близ ст. Сосново Ленинградской области. Предварительно пришлось убедиться в отсутствии географической изменчивости числа междуузлий у рассматриваемого вида в пределах заданной территории. Полученный дисперсионный комплекс представлен в табл. 22.

Порядок расчетов вначале не отличается от описанного для однофакторного дисперсионного комплекса: в нижней части таблицы производится неполная обработка рядов способом произведений с целью получения величин $\Sigma \Sigma f_a$ и Σh , а суммарный ряд n_y обрабатывается способом сумм для получения величин a_1 , a_2 , b_1 и b_2 . Но затем для вычисления S_A^2 и S_B^2 приходится прибегнуть к дополнительным расчетам (правая половина таблицы).

1. Подсчитывают численности наблюдений по градациям (A₁, A₂, B₁ и B₂) факторов, суммируя соответствующие численности n_x . Например, градация B₁ выступает в двух сочетаниях: A₁B₁ и A₂B₁, поэтому ее суммарная численность равна 100 + 52 = 152.

2. Аналогично этому суммируют соответствующие значения Σf_a . Например, для той же градации B₁ $\Sigma f_a = -197 + 186 = -11$.

3. Возводят полученные значения Σf_a в квадрат и делят их на численность каждой градации, получая величины h_i .

Таблица 22. Двухфакторный дисперсионный анализ влияния затопления (фактор А) и засоления (фактор В) территории на число междуузлий (x) у *Odontites serotina* Dum.

x	a	A ₁		A ₂		n _y	a ₁ =309	a ₂ =456	Градации	n _i	Σfa	(Σfa) ²	h _i = $\frac{(\Sigma fa)^2}{n_i}$
		B ₁	B ₂	B ₁	B ₂								
20	+9	—	—	1	—	1	1	1	A ₁	199	-273	74529	374,52
19	+8	—	—	1	—	1	2	3	A ₂	102	+250	62500	612,75
18	+7	—	—	3	—	3	5	8	$\Sigma h_A = 987,27$				
17	+6	—	—	3	1	4	9	17					
16	+5	—	—	8	4	12	21	38					
15	+4	—	1	8	1	10	31	69					
14	+3	1	3	9	7	20	51	120	B ₁	152	-11	121	0,80
13	+2	2	5	14	8	29	80	200	B ₂	149	-12	144	0,97
12	+1	5	11	3	10	29	109	—	$\Sigma h_B = 1,77$				
11	0	10	18	2	9	39	—	—					
10	-1	15	29	—	7	51	153	—					
9	-2	21	18	—	3	42	102	179					
8	-3	32	11	—	—	43	60	77					
7	-4	14	3	—	—	17	17	17					
n _x		100	99	52	59	301	b ₁ =332	b ₂ =273					
Σfa		-197	-76	+183	+64	Σ Σ fa = -23							
(Σfa) ²		38809	5776	34596	4096								
h = (Σfa) ² : n _x		388,09	58,34	665,31	81,92	Σ h = 1193,66							

4. Суммируя значения h_i , получают величины Σh_A и Σh_B , характеризующие доли влияния факторов A и B , необходимые для дальнейших расчетов.

5. Проверяют полученные цифры: а) сложение численностей по градациям (n_i) каждого фактора в отдельности должно дать величину n : $199+102=301$ и $152+149=301$; б) сложение значений Σfa по градациям должно дать величину $\Sigma \Sigma fa$: $-273+250=-23$ и $-11-12=-23$.

6. Рассчитывают знакомые по однофакторному анализу промежуточные величины:

$$s_1 = a_1 - b_1 = 309 - 332 = -23;$$

$$s_2 = a_1 + b_1 + 2(a_2 + b_2) = 309 + 332 + 2(456 + 273) = 2099;$$

$$H = s_1^2 : n = 23^2 : 301 = 1,76.$$

7. Вычисляют суммы квадратов отклонений:

$$S_y^2 = s_2 - H = 2099 - 1,76 = 2097,24;$$

$$S_x^2 = \Sigma h - H = 1193,66 - 1,76 = 1191,90;$$

$$S_z^2 = s_2 - \Sigma h = 2099 - 1193,66 = 905,34.$$

8. Факториальную сумму квадратов отклонений (S_x^2) по следующим рабочим формулам расчленяют на три части, соот-

Таблица 23. Результаты двухфакторного дисперсионного анализа (по данным табл. 22)

Показатели	Формулы	Учитываемые воздействия факторов				Не учитываемые факторы, z	Все факторы, y						
		A	B	AB	x								
Суммы квадратов отклонений	S_i^2	985,51	0,01	206,38	1191,90	905,34	2097,24						
Степень влияния факторов	$r_{iy}^2 = \frac{S_i^2}{S_y^2}$	0,47	0,00	0,10	0,57	0,43	1,00						
Число степеней свободы	ν_i	1	1	1	3	297	300						
Дисперсии	$\sigma_i^2 = \frac{S_i^2}{\nu_i}$	985,51	0,01	206,38	397,3	3,05	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>ν_x</td><td>1</td><td>3</td></tr><tr><td>ν_z</td><td></td><td></td></tr></table>	ν_x	1	3	ν_z		
ν_x	1	3											
ν_z													
Критерий Фишера	$F = \frac{\sigma_i^2}{\sigma_z^2}$	323,12	0,00	67,67	130,26	—	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>297</td><td>3,86</td><td>2,62</td></tr><tr><td>(400)</td><td>6,70</td><td>3,83</td></tr></table>	297	3,86	2,62	(400)	6,70	3,83
297	3,86	2,62											
(400)	6,70	3,83											

Примечание. Число степеней свободы рассчитывается здесь по следующим формулам: $\nu_A = r_A - 1$; $\nu_B = r_B - 1$; $\nu_{AB} = \nu_A \cdot \nu_B$; $\nu_x = (r_A \cdot r_B) - 1$; $\nu_z = n - (r_A \cdot r_B)$; $\nu_y = n - 1$, где r_A и r_B — число градаций факторов A и B соответственно, n — общий объем дисперсионного комплекса.

ветствующие влиянию фактора A , фактора B и их сочетания AB :

$$S_A^2 = \sum h_A - H = 987,27 - 1,76 = 985,51;$$

$$S_B^2 = \sum h_B - H = 1,77 - 1,76 = 0,01;$$

$$S_{AB}^2 = S_x^2 - S_A^2 - S_B^2 = 1191,90 - 985,51 - 0,01 = 206,38.$$

Нетрудно заметить, что при этом сохраняется соотношение $S_A^2 + S_B^2 + S_{AB}^2 = S_x^2$ (т. е. $985,51 + 0,01 + 206,38 = 1191,90$).

После этого остается составить итоговую таблицу (табл. 23) и произвести в ней некоторые подсчеты, аналогичные таковым для однофакторного дисперсионного анализа.

В результате видим, что затопление (фактор A) само по себе довольно сильно (47 %) и в высшей степени достоверно ($P = 0,999$) влияет на число междоузлий у *O. serotina* Dum., чего никак нельзя сказать о засолении (фактор B), влияние которого практически равно нулю. В то же время сочетание воздействия обоих факторов (AB) еще на 10 % усиливает наблюдаемый результат. Это значит, что эффект уменьшения числа междоузлий выражен сильнее в одновременно затопляемых и засоленных местообитаниях по сравнению с только затопляемыми.

В заключение отметим, что существуют и некоторые особые случаи дисперсионного анализа (например, иерархический и ранговый дисперсионный анализ, анализ с использованием качественных признаков и др.), с которыми можно ознакомиться в других руководствах [Плохинский, 1970; Терентьев, Ростова, 1977; Лакин, 1980].

Глава 4

ИЗУЧЕНИЕ ВЗАИМНОЙ СВЯЗИ ПРИЗНАКОВ (КОРРЕЛЯЦИОННЫЙ АНАЛИЗ)

4.1. ПОНЯТИЕ О КОРРЕЛЯЦИОННОЙ СВЯЗИ

Представление о корреляциях как взаимозависимостях органов живых существ восходит к глубокой древности. Еще Аристотель утверждал, что природа дает одной части организма то, что забирает у другой его части. Впоследствии эта мысль нашла отражение в «законе равновесия органов», сформулированном Жоффруа Сент-Илером, но первое крупное обобщение в этой

области принадлежит Ж. Кювье, выдвинувшему в 1815 г. свои знаменитые принципы: принцип координации и принцип корреляции. Оба принципа имеют прямое отношение к явлению биологических корреляций, отражая соответственно его филогенетический и онтогенетический аспекты [Шмальгаузен, 1969; Шмидт, 1979, 1981a]. Согласно второму принципу все органы живого существа составляют систему, причем части ее зависят друг от друга таким образом, что изменение одной части влечет за собой изменение других. Кювье не мог знать генетической основы явления морфологических корреляций, но это не мешало ему восстанавливать примерный облик ископаемых животных по нескольким найденным костям.

Теперь мы знаем, что корреляция признаков организма обусловлена плейотропным действием генов и некоторыми другими особенностями функционирования наследственного аппарата. Известно и то, что, в сущности, все явления природы, на каком бы уровне организации живого вещества мы их ни рассматривали, прямо или косвенно связаны друг с другом огромным количеством корреляционных зависимостей.

При изучении корреляций надо иметь в виду, что любая математически доказанная связь должна находить удовлетворительное биологическое истолкование или хотя бы не противоречить смыслу явления. Дело в том, что существуют так называемые «бессмысленные» корреляции. Например, в Европе при продвижении из Скандинавии на юг уменьшение среднего роста человека сопровождается возрастанием доли католиков среди населения. Математически эта связь достоверна, но было бы несерьезно утверждать, что религиозные убеждения людей зависят от их роста!

Корреляцию не следует путать с причинностью. Каждая корреляция, разумеется, имеет свою причину, и если два явления постоянно следуют одно за другим, то причинно-следственный характер их связи может быть вполне очевиден. Например, погодные условия определяют урожай, а не наоборот. Однако даже в этом простом примере первопричина связи лежит на другом уровне, ибо состояние погоды зависит от ряда других факторов.

Что же касается морфологических корреляций, то их причины могут быть вскрыты только путем специальных генетико-физиологических исследований. Вот почему надо остерегаться объяснять, скажем, изменение одного признака или органа изменением другого: как правило, мы можем лишь констатировать сам факт наличия корреляционной связи. Сказанное ничуть не умаляет большого значения корреляционного анализа в биологии: только биометрическим путем можно обнаружить и изучить множество корреляций, не поддающихся визуальному наблюдению, только так можно исследовать сложную внутреннюю структуру связей между признаками живого объекта или явления.

Термин «корреляция» (co-relation), предложенный Ф. Гальтоном в 1888 г. и буквально означающий «соотношение» или «взаимосвязь», должен как всякое строгое понятие употребляться в присущем ему смысле: *корреляция есть наличие взаимной согласованности в изменчивости двух или нескольких признаков (явлений)*. Установление меры этой согласованности составляет начальную задачу корреляционного анализа, на основе выполнения которой возможны дальнейшие построения. Если исследователя интересует вопрос о том, насколько именно изменяется один признак при определенном изменении другого, следует перейти к регрессионному анализу (глава 5).

Закономерности неживой природы чаще всего определяются функциональными связями, при которых одному значению признака (аргумента) соответствует одно определенное значение (реже несколько определенных значений) другого признака. Так, ребро куба однозначно определяет его объем. В биологии полная (равная единице) связь практически не встречается, хотя близкие к ней значения нередко имеют место. Живым объектам свойственна корреляционная связь, при которой каждому значению одного признака соответствует «облако рассеивания» множества значений другого признака, варьирующих вокруг средней величины. Поэтому для установления корреляционной связи недостаточно одного или нескольких наблюдений, а требуется обработка статистической выборки.

По своей форме корреляционная связь может быть линейной и нелинейной, по направлению — прямой (положительной) и обратной (отрицательной), по величине — от нулевой до полной ($0 \div \pm 1$). Существует ряд показателей связи, отличающихся не только техникой расчета, но также своим смысловым содержанием и мощностью. Одни из них улавливают только линейную связь (точнее линейную составляющую любой корреляционной связи), другие пригодны и при нелинейных зависимостях. Первые рассчитываются проще и используются значительно чаще, но не дают правильного представления о степени связи в случае заметной ее нелинейности (в разделе 4.3.2 будет дан критерий нелинейности связи). Таким образом, в зависимости от характера биологического материала и целей исследования следует применять разные показатели связи, среди которых есть и непараметрические. Любой показатель связи есть выборочная величина и тем самым служит приближенной оценкой соответствующего показателя генеральной совокупности.

Изучая отдельные признаки, мы до сих пор имели дело с их моновариантным распределением, т. е. изучали отдельные вариационные ряды. Корреляционный анализ изучает сопряженную изменчивость двух или нескольких признаков, совместное распределение которых оказывается бивариантным или мультивариантным. С составлением таких распределений (корреляцион-

ных таблиц или «корреляционных решеток») и начинается работа.

В общем виде корреляционная решетка (табл. 24) представляет собой многопольную таблицу из $k_x \times k_y$ клеток, где k_x и k_y — число значений (или классов) признаков x и y . Значения признаков (или классы) удобно располагать в возрастающем порядке: для x — слева направо, для y — сверху вниз. В клетках таблицы производят разность сопряженных частот (f_{xy}) в зависимости от значений двух признаков одновременно.

Таблица 24. Схема корреляционной решетки

y	x	x_1	x_2	x_3	...	x_i	f_y
y_1		$f_{x_1 y_1}$	$f_{x_2 y_1}$	$f_{x_3 y_1}$...	$f_{x_i y_1}$	f_{y_1}
y_2		$f_{x_1 y_2}$	$f_{x_2 y_2}$	$f_{x_3 y_2}$...	$f_{x_i y_2}$	f_{y_2}
y_3		$f_{x_1 y_3}$	$f_{x_2 y_3}$	$f_{x_3 y_3}$...	$f_{x_i y_3}$	f_{y_3}
...	
y_i		$f_{x_1 y_i}$	$f_{x_2 y_i}$	$f_{x_3 y_i}$...	$f_{x_i y_i}$	f_{y_i}
	f_x	f_{x_1}	f_{x_2}	f_{x_3}	...	f_{x_i}	n

4.2. ЛИНЕЙНАЯ КОРРЕЛЯЦИЯ

При линейной корреляционной связи равномерным изменениям одного признака соответствуют в среднем равномерные же изменения другого признака. Линейная или близкая к ней по форме связь возникает тогда, когда распределения обоих признаков не слишком отличаются от нормального. Внешним ее проявлением служит концентрация максимальных сопряженных частот вдоль одной из диагоналей корреляционной решетки (рис. 8). Степень рассеяния частот говорит о силе связи: чем рассеяние меньше, тем связь сильнее, и наоборот. Можно определить и знак связи: если ряды x и y расположены в возрастающем порядке, как показано в табл. 24, то при положительной связи длинная ось эллипса рассеивания сопряженных частот расположится по диагонали из левого верхнего в правый нижний угол таблицы, а при отрицательной связи — по другой диагонали.

4.2.1. Коэффициент линейной корреляции. Для получения меры линейной корреляции обратимся к понятию о нормированном отклонении (раздел 2.3). Поскольку нормированное отклонение t есть отклонение варианты от средней арифметической, отнесенное к величине среднего квадратичного отклонения,

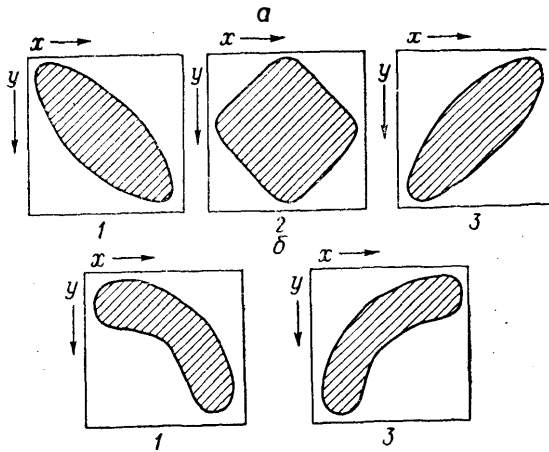


Рис. 8. Характер распределения сопряженных частот в корреляционной решетке при линейной (а) и нелинейной связи (б).

1 — положительная корреляция; 2 — отсутствие связи; 3 — отрицательная корреляция.

а в корреляционном анализе мы имеем дело не менее чем с двумя признаками (x и y), то соответственно получаем два нормированных отклонения

$$t_x = \frac{x_i - M_x}{\sigma_x} \quad \text{и} \quad t_y = \frac{y_i - M_y}{\sigma_y}.$$

Замечательным свойством нормированных отклонений двух вариационных рядов, позволяющим использовать их для получения меры связи, является то, что сумма их произведений, отнесенная к числу наблюдений, колеблется в пределах от 0 до ± 1 :

$$-1 \leq \frac{\sum t_x t_y}{n} \leq +1.$$

Эта величина и получила название коэффициента линейной корреляции (r):

$$r = \frac{\sum t_x t_y}{n} = \frac{\sum (x_i - M_x)(y_i - M_y)}{n \sigma_x \sigma_y}.$$

В вычислительном отношении приведенная формула неудобна, поэтому было разработано несколько ее модификаций, из которых наиболее употребительной является формула Бравэ—Пирсона, основанная на способе условных средних:

$$r = \frac{\sum f_{xy} a_x a_y - \frac{\sum f_x a_x \cdot \sum f_y a_y}{n}}{n \sigma_x^* \sigma_y^*},$$

где \hat{f}_{xy} — сопряженные частоты рядов x и y ; a_x и a_y — отклонения от условных средних; f_x и f_y — частоты вариационных рядов; n — объем выборки; σ_x^* и σ_y^* — средние квадратичные отклонения, взятые без учета классового интервала («неполные сигмы»).

Рабочая формула для расчета коэффициента корреляции включает в себя некоторые выборочные параметры, которые

Таблица 25. Корреляция между длиной соцветия (y) и числом цветков в нем (x) у *Odontites verna* (Bell.) Dum.

y	x	a_x	3-4	5-6	7-8	9-10	11-12	13-14	15-16	17-18	19-20	21-22	f_y
		a_y	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	
10-29	-3	5	1										6
30-49	-2		14	6	2								22
50-69	-1	2	5	13	11	9							40
70-89	0			1	12	7	2	1					23
90-109	+1					4	10	4					18
110-129	+2					4	9	6	4				23
130-149	+3						2	4	3				9
150-169	+4								5	1	1		7
170-189	+5								1	1			2
f_x			7	20	20	25	24	23	15	13	2	1	$n=150$

необходимо предварительно рассчитать путем обработки каждого ряда в отдельности. Помимо объема выборки (n) и «неполных сигм» (σ_x^* и σ_y^*) надо знать величину условных моментов первой степени $\frac{\sum f_x a_x}{n}$ и $\frac{\sum f_y a_y}{n}$ при обработке рядов способом произведений или $(\bar{v}_1)_x$ и $(\bar{v}_1)_y$ при обработке их способом сумм. Кстати, при использовании способа сумм правая часть числителя в формуле коэффициента корреляции становится проще, так как

$$\frac{\sum f_x a_x \cdot \sum f_y a_y}{n} = (\bar{v}_1)_x \cdot (\bar{v}_1)_y \cdot n.$$

Рассмотрим конкретный пример. При изучении связи между числом цветков в соцветии и его длиной у *Odontites verna* (Bell.) Dum. была составлена корреляционная решетка (табл. 25), по расположению сопряженных частот в которой видно, что связь положительна, достаточно сильна и не слишком отли-

чается от линейной. Необходимые для расчета коэффициента корреляции выборочные параметры следующие: $(\bar{v}_1)_x = -0,32$; $(\bar{v}_1)_y = 0,18$; $\sigma_x^* = 2,043$; $\sigma_y^* = 1,857$. Проследим порядок дальнейшей работы.

1. Столбец и строку корреляционной решетки (табл. 25), соответствующие условным средним, выбранным при обработке каждого признака в отдельности, выделяют жирными линиями. Получающийся «крест» делит корреляционную решетку на четыре квадранта: I — левый верхний, II — правый верхний, III — левый нижний и IV — правый нижний. Сопряженные частоты, попавшие внутрь «креста», из дальнейших расчетов исключаются.

2. Как и при обработке каждого ряда в отдельности, от условных средних, положение которых отмечено нулем, вправо и влево делают отсчеты числами натурального ряда с соответствующим знаком.

3. В каждом квадранте производят подсчет суммы произведений сопряженных частот (f_{xy}) на оба соответствующих отклонения (a_x и a_y) от условных средних. Например, для I квадранта имеем: $5(-3)(-4) + 1(-3)(-3) + 14(-2)(-3) + 6(-2)(-2) + 2(-2)(-1) + 2(-1)(-4) + 5(-1)(-3) + 13(-1)(-2) + 11(-1)(-1) = +241$. Аналогичным образом получают суммы и для других квадрантов (в нашем примере 0 для II и III, +272 для IV). Складывая их, определяют общую для корреляционной решетки сумму: $\sum f_{xy} a_x a_y = +241 + 0 + 0 + 272 = +513$.

4. Подставляя эту сумму и указанные выше значения выборочных параметров в формулу коэффициента корреляции, рассчитываем его величину:

$$r = \frac{513 - (-0,32 \cdot 0,18 \cdot 150)}{150 \cdot 2,043 \cdot 1,857} = +0,92.$$

Для оценки достоверности коэффициента корреляции можно воспользоваться критерием t , вычислив предварительно ошибку репрезентативности самого коэффициента:

$$m_r = \frac{1 - r^2}{\sqrt{n}} = \frac{1 - 0,92^2}{\sqrt{150}} = 0,013;$$

$$t = \frac{r}{m_r} = \frac{0,92}{0,013} = 70,77.$$

Полученное значение t сравнивают со стандартным (Приложение III) при числе степеней свободы $\nu = n - 2$. В нашем примере $\nu = 150 - 2 = 148$ и величина t во много раз превосходит значение t_{st} при любом уровне существенности ($t_{001} = 3,37$), т. е. коэффициент корреляции в высшей степени достоверен.

Более строгим способом оценки достоверности коэффициента корреляции служит z -преобразование Фишера, использовать которое рекомендуется при объеме выборки $n < 100$ и при низ-

ких значениях коэффициента корреляции ($r \leq 0,20$). Дело в том, что распределение коэффициентов корреляции в генеральной совокупности асимметрично из-за преобладания слабых и средних связей, тогда как критерий t рассчитан на нормальное или близкое к нему распределение. Поэтому значение r предварительно переводят в величину z , которая распределена нормально: $z = 0,5 \{ \ln(1+r) - \ln(1-r) \}$.

Практически этот перевод осуществляется по готовой таблице (Приложение IX), в которую входят по рассчитанному значению r . В разбираемом примере значению $r = 0,92$ соответствует $z = 1,589$. После этого вычисляют ошибку величины z и критерий t :

$$m_z = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{147}} = 0,083;$$

$$t = \frac{z}{m_z} = \frac{1,589}{0,083} = 19,14.$$

При $v = n - 2 = 148$ по-прежнему $t \gg t_{001}$, и вывод о высокой достоверности полученного коэффициента корреляции остается в силе, хотя меньшее значение t указывает на повышенную строгость оценки с помощью z -преобразования.

В заключение этого раздела следует сделать несколько дополнений. Если исследователя интересует не столько степень связи, сколько мера независимости двух признаков, то следует рассчитать коэффициент алиенации, определяемый формулой $r_a = 1 - |r|$, т. е. величиной, дополняющей коэффициент корреляции до 1. В рассмотренном примере коэффициент алиенации очень мал: $r_a = 1 - 0,92 = 0,08$.

Иногда возникает вопрос: какие значения r следует считать высокими, средними или низкими? На первый взгляд может показаться, что при $r = 0,5$ 50% изменчивости признаков носит согласованный характер. Но это не так. Коэффициент корреляции — средняя выборочная величина, и если, скажем, $r = 0$, то это вовсе не значит, что в выборке нет особей с положительной или отрицательной связью между признаками. Наоборот, при высоком значении r у некоторых особей связь между признаками может быть близка к нулю. Математическая теория корреляции доказывает, что при $r = 0,5$ не половина, а только 25% изменчивости одного признака действительно обусловлено согласованным изменением другого признака, а не случайными комбинациями. Поэтому, вообще говоря, более точной мерой корреляционной связи служит величина r^2 , получившая название коэффициента детерминации. Отсюда следует, что значения $r \geq 0,70$, при которых истинная взаимообусловленность признаков составляет около 50%, можно считать высокими; значения $r = 0,50 - 0,70$ — средними и значения $r \leq 0,50$ — низкими. Можно добавить, что у растений, как объектов менее интегрированных

по сравнению с животными, сильных корреляционных связей в целом меньше. Например, из 577 коэффициентов корреляции, рассчитанных для трех видов *Odontites*, лишь 49 (т. е. 8,5 %) превысили значение $r=0,70$.

В предвидении корреляционного анализа следует избегать слишком малых выборок, так как при этом достоверными оказываются лишь сильные связи. Так, на 5%-ном уровне существенности при $n=5$ достоверны только значения $r \geq 0,878$; при $n=10$ — $r \geq 0,632$, но уже при $n=20$ пороговый уровень снижается до $r \geq 0,444$; при $n=30$ достоверны значения $r \geq 0,361$, а при $n=100$ — значения $r \geq 0,196$.

4.2.2. Частная и множественная корреляция. Корреляционная связь между двумя признаками может осуществляться не непосредственно, а косвенно — за счет связи каждого из них в отдельности с каким-либо третьим (четвертым и т. д.) признаком. Например, размеры вегетативных органов обычно сильно коррелируют с высотой растения, и для изучения связи между ними в «чистом» виде необходимо найти способ исключить влияние на эту связь высоты растений.

Если рассчитаны парные коэффициенты корреляции r_{xy} , r_{xz} и r_{yz} между тремя признаками (x , y , z), то исключить влияние признака z на связь между признаками x и y можно, рассчитав по следующей формуле коэффициент частной корреляции:

$$zr_{xy} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}.$$

Ошибка этого коэффициента вычисляется по формуле

$$m_z r_{xy} = \sqrt{\frac{1 - zr_{xy}^2}{n - 2}},$$

а достоверность связи оценивается обычным путем (раздел 4.2.1) с помощью критерия t (лучше производить z -преобразование) при числе степеней свободы $\nu = n - 3$.

Возьмем конкретный пример. При изучении связей между длиной соцветия (x), длиной листа (y) и высотой растения (z) в выборке ($n=150$) *Odontites serotina* Dum. были получены достоверные значения парных коэффициентов корреляции: $r_{xy} = 0,46$, $r_{xz} = 0,61$ и $r_{yz} = 0,70$. Требуется установить, какова связь между первыми двумя признаками в «чистом» виде, т. е. не влияет ли высота растения на полученную величину $r_{xy} = 0,46$?

Подставляя требуемые значения в формулу коэффициента частной корреляции, имеем

$$zr_{xy} = \frac{0,46 - 0,61 \cdot 0,70}{\sqrt{(1 - 0,61^2)(1 - 0,70^2)}} = 0,06.$$

Ошибка коэффициента

$$m_z r_{xy} = \sqrt{\frac{1 - 0,06^2}{148}} = 0,08$$

превышает его значение, и поэтому без вычисления t ясно, что он недостоверен. Как видим, достоверное значение связи между длиной соцветия и длиной листа $r_{xy} = 0,46$ при исключении влияния высоты растения оказалось недостоверным, практически равным нулю, т. е. в действительности эти признаки независимы друг от друга, а в выборке связаны косвенно, через высоту растения. Надо заметить, что при частной корреляции первоначальный знак связи может даже изменяться на противоположный.

Существуют формулы, позволяющие исключить влияние одновременно двух [Плохинский, 1970, с. 158] и большего числа признаков, но необходимость таких операций возникает редко.

Задача множественной корреляции по своему смыслу противоположна цели частной корреляции: на основе имеющихся парных коэффициентов корреляции можно установить степень связи одного признака с двумя другими, вместе взятыми. Формула коэффициента множественной корреляции имеет вид

$$r_{x \cdot yz} = \sqrt{\frac{r_{xy}^2 + r_{xz}^2 - 2r_{xy}r_{xz}r_{yz}}{1 - r_{yz}^2}},$$

где точка в обозначении $r_{x \cdot yz}$ означает, что изучается взаимосвязь признака x с признаками y и z , вместе взятыми. Оценка достоверности коэффициента множественной корреляции производится по общим правилам при $\nu = n - 3$.

Вернемся к предыдущему примеру, но поставим теперь иную задачу: установить связь длины соцветия (x) с длиной листа (y) и высотой растения (z), вместе взятыми. Подставляя приведенные ранее значения парных коэффициентов корреляции в формулу коэффициента множественной корреляции, имеем

$$r_{x \cdot yz} = \sqrt{\frac{0,46^2 + 0,61^2 - 2 \cdot 0,46 \cdot 0,61 \cdot 0,70}{1 - 0,70^2}} = 0,61.$$

Полученное значение достоверно, что видно из следующего: величине $r_{x \cdot yz} = 0,61$ соответствует $z = 0,7089$ (Приложение IX),

$$m_z = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{147}} = 0,083; \quad t = \frac{z}{m_z} = \frac{0,7089}{0,083} = 8,54$$

и при $\nu = 150 - 3 = 147$ (Приложение III) $t > t_{st}$ при любом уровне существенности.

В заключение данного раздела следует отметить, что методы частной и множественной корреляции, изложенные выше,

основаны на использовании коэффициента корреляции r и, следовательно, справедливы только для линейной и близкой к ней связи.

4.3. НЕЛИНЕЙНАЯ КОРРЕЛЯЦИЯ

При нелинейной корреляционной связи равномерным изменениям одного признака соответствуют в среднем неравномерные, но подчиняющиеся определенной закономерности изменения другого признака. Нелинейная связь возникает обычно при заметном отклонении одного или обоих распределений признаков от нормального, а ее внешним проявлением служит расположение максимальных сопряженных частот в корреляционной решетке по изогнутой линии (см. рис. 8, б). Однако далеко не всегда по внешнему виду корреляционной решетки можно уверенно судить о нелинейности связи, поэтому в разделе 4.3.2 будут описаны специальные критерии для определения типа связи.

4.3.1. Корреляционное отношение (η) измеряет как линейную, так и нелинейную зависимость. В первом случае $\eta = r$, но чем сильнее выражена нелинейность связи, тем больше значение корреляционного отношения превышает величину коэффициента корреляции r . Один из способов вычисления корреляционного отношения был показан в связи с изложением техники дисперсионного анализа (глава 3), в котором квадрат этого показателя (η^2) используется в качестве меры влияния факторов на изучаемый признак.

Ниже мы рассмотрим другой способ вычисления корреляционного отношения, связанный с техникой регрессионного анализа (глава 5). Дело в том, что в отличие от коэффициента корреляции, являющегося мерой обоюдной связи между признаками, корреляционное отношение способно отражать как зависимость признака y от признака x ($\eta_{y/x}$), так и зависимость признака x от признака y ($\eta_{x/y}$). Таким образом, для пары признаков могут быть рассчитаны два корреляционных отношения, первое из которых условно назовем прямым, а второе — обратным. В общем случае прямое и обратное корреляционные отношения не совпадают, но чем сильнее связь и чем ближе она к линейной, тем больше (вплоть до совпадения) сближаются между собой их значения. На практике, исходя из биологической значимости того или иного признака, рассчитывают обычно одно из двух корреляционных отношений.

Другое отличие от коэффициента корреляции состоит в том, что корреляционное отношение принимает значения не от -1 до $+1$, а от 0 до $+1$, и направление связи следует определять по внешнему виду корреляционной решетки (рис. 8).

Корреляционное отношение есть не что иное, как отношение двух средних квадратичных отклонений, одно из которых характеризует часть изменчивости первого признака, обусловлен-

ную его зависимость от второго признака, а другое является обычной мерой общей изменчивости первого признака:

$$\eta_{y/x} = \frac{\sigma_{y/x}}{\sigma_y}; \quad \eta_{x/y} = \frac{\sigma_{x/y}}{\sigma_x}.$$

Следовательно, новизна задачи состоит в вычислении значений $\sigma_{y/x}$ и $\sigma_{x/y}$ способ расчета которых будет показан ниже. Что касается оценки достоверности корреляционного отношения, то рекомендуется вычислять ошибку квадрата этого показателя по формуле

$$m_{\eta^2} = (1 - \eta^2) \frac{k-1}{n-k},$$

где k — число классов вариационного ряда; n — объем выборки, после чего при $v_1 = k-1$ и $v_2 = n-k$ можно использовать критерий Фишера

$$F = \frac{\eta^2}{m_{\eta^2}},$$

сравнивая его со стандартным значением (Приложение IV).

Рассмотрим конкретный пример. Изучается связь между высотой растения и числом пар цветков в осевом соцветии *Odontites serotina* Dum. Корреляционная решетка (табл. 26) строится без разбивки на квадранты; не нужны в ней и графы, в которых при расчете коэффициента корреляции проставляются отклонения от условных средних. Зато к решетке добавляются строка и столбец для записи точек эмпирических линий регрессии y/x и x/y (читается: « y по x » и « x по y »).

Таблица 26. Корреляция между высотой растения (y) и числом пар цветков в осевом соцветии (x) у *Odontites serotina* Dum.

y	x	2,5	6,5	10,5	14,5	18,5	22,5	f_y	x/y
49,5	48	2						50	2,66
89,5	58	53	4					115	4,62
129,5	12	52	54	5				123	8,19
169,5	2	16	42	21	2			83	10,74
209,5	3	11	12	18	7			51	11,68
249,5	1	1	8	9	3			22	12,68
289,5		2	2			1	1	6	12,50
f_x		124	137	122	53	13	1	$n=450$	
y/x		83,37	127,16	160,32	192,90	218,73	289,50		

Примечание. Для обоих признаков в таблице даны центральные классовые значения вариационных рядов (x и y).

Эти точки рассчитываются для каждой строки и каждого столбца как взвешенные средние арифметические. Например, для строки с центральным классовым значением $y=89,5$ точка $x/y = (58 \cdot 2,5 + 53 \cdot 6,5 + 4 \cdot 10,5) : 115 = 4,62$, а для столбца с центральным классовым значением $x=10,5$ точка $y/x = (4 \cdot 89,5 + 54 \cdot 129,5 + 42 \cdot 169,5 + 12 \cdot 209,5 + 8 \cdot 249,5 + 2 \cdot 289,5) : 122 = 160,32$.

Для дальнейших расчетов необходимо знать медию и средние квадратичные отклонения обоих рядов, т. е. предварительно каждый из них должен быть обработан в отдельности. В разбираемом примере $M_y = 134,7$, $\sigma_y = 56,0$; $M_x = 7,8$, $\sigma_x = 4,4$. Расчеты ведутся в табличной форме в соответствии с алгоритмом, содержащимся в заголовках столбцов (табл. 27).

Таблица 27. Расчет прямого корреляционного отношения ($\eta_{y/x}$) (по данным табл. 26)

x	y/x	$y/x - M_y$	$(y/x - M_y)^2$	f_x	$(y/x - M_y)^2 \cdot f_x$
2,5	83,37	-51,33	2634,77	124	326711,48
6,5	127,16	-7,54	56,85	137	7788,45
10,5	160,32	+25,62	656,38	122	80078,36
14,5	192,90	+58,20	3387,24	53	179523,72
18,5	218,73	+84,03	7061,04	13	91793,52
22,5	289,50	+154,80	23953,04	1	23963,04

$$M_y = 134,7; \sigma_y = 56,0$$

$$n = 450 \quad \Sigma = 709858,57$$

В последнем столбце этой таблицы содержится ряд взвешенных квадратов отклонений точек эмпирической линии регрессии y/x от M_y . Величиной, характеризующей долю изменчивости признака y , обусловленную его зависимостью от признака x , будет сигма этого ряда отклонений:

$$\sigma_{y/x} = \sqrt{\frac{\Sigma(y/x - M_y)^2 \cdot f_x}{n}} = \sqrt{\frac{709858,57}{450}} = 39,72.$$

Искомое прямое корреляционное отношение составляет

$$\eta_{y/x} = \frac{\sigma_{y/x}}{\sigma_y} = \frac{39,72}{56,0} = 0,71;$$

ошибка квадрата этого показателя

$$m_{\eta_{y/x}}^2 = (1 - 0,71^2) \frac{6 - 1}{450 - 6} = 0,0056,$$

откуда

$$F = \frac{0,71^2}{0,0056} = 90,02.$$

Из Приложения IV находим, что при $\nu_1 = 6 - 1 = 5$ и $\nu_2 = 450 - 6 = 444$ стандартное значение $F_{01} = 3,06$, и поскольку $F \gg F_{st}$, полученное корреляционное отношение в высшей степени достоверно.

Аналогичным образом рассчитывается и обратное корреляционное отношение ($\eta_{x/y}$), только теперь в основу расчета бе-

Таблица 28. Расчет обратного корреляционного отношения ($\eta_{x/y}$) (по данным табл. 26)

y	x/y	$x/y - M_x$	$(x/y - M_x)^2$	f_y	$(x/y - M_x)^2 \cdot f_y$
49,5	2,66	-5,14	26,42	50	1321,00
89,5	4,62	-3,18	10,11	115	1162,65
129,5	8,19	+0,39	0,15	123	18,45
169,5	10,74	+2,94	8,64	83	717,12
209,5	11,68	+3,88	15,05	51	767,55
249,5	12,68	+4,88	23,81	22	523,82
289,5	12,50	+4,70	22,09	6	132,54

$$M_x = 7,8; \sigma_x = 4,4$$

$$|n = 450| \Sigma = 4643,13$$

рется отклонения точек эмпирической линии регрессии x/y от M_x (табл. 28). Последовательно получаем

$$\sigma_{x/y} = \sqrt{\frac{(x/y - M_x)^2 \cdot f_y}{n}} = \sqrt{\frac{4643,13}{450}} = 3,21;$$

$$\eta_{x/y} = \frac{\sigma_{x/y}}{\sigma_x} = \frac{3,21}{4,4} = 0,73;$$

$$m_{x/y}^2 = (1 - 0,73^2) \frac{7-1}{450-7} = 0,0063;$$

$$F = \frac{0,73^2}{0,0063} = 84,59$$

и при $\nu_1 = 7 - 1 = 6$ и $\nu_2 = 450 - 7 = 443$ стандартное значение $F_{01} = 2,85$, т. е. $F \gg F_{st}$ и вывод тот же, что в первом случае.

Как видно, прямое ($\eta_{y/x}$) и обратное ($\eta_{x/y}$) корреляционные отношения в рассмотренном примере практически совпадают. Для сравнения укажем, что рассчитанный по данным

табл. 26 коэффициент линейной корреляции составляет $r=0,70$, т. е. незначительно уступает по своей величине корреляционным отношениям. Это обстоятельство наводит на мысль о том, что в разобранном примере корреляционная связь между признаками близка к линейной. Не будем, однако, торопиться с выводом, ибо существует возможность проверить это более строгим способом.

4.3.2. Критерии нелинейности связи. Точно линейная связь в биологии такая же редкость, как и строго нормальное распределение признака. На практике небольшие отклонения связи от линейной можно не принимать в расчет и использовать линейные методы, облегчающие исследование корреляции и регрессии. Важно, однако, не переступать тот порог, по достижении которого отклонение связи от линейной становится существенным.

Известно несколько критериев нелинейности связи, в том или ином виде использующих разность $\eta^2 - r^2$, т. е. требующих предварительного расчета корреляционного отношения и коэффициента корреляции и основанных на их сравнении. Наиболее употребительны из них критерии Блэкмана и Фишера. Критерий Блэкмана более прост и может быть записан в виде следующей формулы:

$$n(\eta^2 - r^2) \geq 11,37,$$

где n — объем выборки; η — большее из двух корреляционных отношений; r — коэффициент корреляции. Связь признается нелинейной, если рассчитанное значение критерия превышает величину 11,37.

Критерий Фишера считается более точным:

$$F = \frac{(\eta_{y/x}^2 - r^2)(n - k_x)}{(1 - \eta_{y/x}^2)(k_x - 2)},$$

где k_x — число классов в ряду x (если расчет ведется для прямого корреляционного отношения $\eta_{y/x}$), а остальные обозначения те же, что в предыдущей формуле. Рассчитанное значение F сравнивается по обычным правилам со стандартным (Приложение IV) при числе степеней свободы: $\nu_1 = k_x - 2$; $\nu_2 = n - k_x$.

В предыдущем разделе при рассмотрении примера (табл. 26) были получены следующие показатели: $\eta_{x/y} = 0,73$; $r = 0,70$ при $n = 450$ и $k_y = 7$.

Критерий Блэкмана $450(0,73^2 - 0,70^2) = 19,31$ превышает величину 11,37, что указывает на нелинейный характер зависимости.

Критерий Фишера для данного корреляционного отношения

$$F = \frac{(0,73^2 - 0,70^2)(450 - 7)}{(1 - 0,73^2)(7 - 2)} = 8,14$$

подтверждает этот вывод, ибо при $\nu_1 = 7 - 5 = 4$ и $\nu_2 = 450 - 7 =$

=443 стандартное значение $F_{01}=3,06$ (Приложение IV), т. е. $F > F_{st}$.

Вообще говоря, этого достаточно для признания существенности нелинейной связи, так как расчет строится на большем из двух корреляционных отношений. Если же использовать меньшее из них ($\eta_{y/x}=0,71$; $k_x=6$), то критерий Блэкмана $450(0,71^2-0,70^2)=6,3$ позволяет считать связь линейной, а значение критерия Фишера

$$F = \frac{(0,71^2 - 0,70^2)(450 - 6)}{(1 - 0,71^2)(6 - 2)} = 3,16$$

при $v_1=4$ и $v_2=444$ лежит между стандартными значениями $F_{05}=2,39$ и $F_{01}=3,36$ (Приложение IV), т. е. в зависимости от нашей требовательности связь может быть признана нелинейной или линейной. В целом следует сделать вывод, что нелинейность связи в рассматриваемом примере незначительна.

4.4. НЕПАРАМЕТРИЧЕСКИЕ ПОКАЗАТЕЛИ СВЯЗИ, ОСНОВАННЫЕ НА МЕТОДЕ χ^2

Коэффициент линейной корреляции (r) и корреляционное отношение (η) для своего расчета требуют знания выборочных параметров, используются для измерения связи только между

Таблица 29. Схема четырехпольной корреляционной решетки

y	x	x_1	x_2	Σ
y_1		a	b	$a+b$
y_2		c	d	$c+d$
	Σ	$a+c$	$b+d$	$a+b+c+d=n$

количественными признаками и зависят от расположения строк и столбцов в корреляционной решетке (оба вариационных ряда должны быть расположены в своем естественном порядке, перестановка строк и столбцов недопустима).

Между тем на практике нередко возникает потребность в измерении связи между качественными признаками, не имеющими таких параметров, как медиана и среднее квадратичное отклонение, или рядом которых мы вправе придать любой порядок написания классов (например, белый — синий — красный или белый — красный — синий). В подобных случаях следует использовать непараметрические показатели связи, называемые также показателями сопряженности, основанные на использовании метода χ^2 .

В простейшем случае оба качественных признака варьируют по альтернативному типу (т. е. имеют по два класса каждый), и корреляционная решетка оказывается четырехпольной (табл. 29). Критерий χ^2 для такой решетки рассчитывается по формуле

$$\chi^2 = \frac{(ad - bc)^2 \cdot n}{(a + b)(b + d)(c + d)(a + c)},$$

где a , b , c и d — сопряженные частоты в соответствующих клетках табл. 29; n — объем выборки.

Если в клетках корреляционной решетки содержатся очень малые частоты (меньше 5), то в формулу необходимо ввести поправку

$$\chi^2 = \frac{(ad - bc - 0.5n)^2 \cdot n}{(a + b)(b + d)(c + d)(a + c)}.$$

Достоверность связи оценивается обычным путем (Приложение V) при $v=1$, так как число степеней свободы для корреляционных решеток $v = (k_x - 1)(k_y - 1)$, где k_x и k_y — число классов ряда x и ряда y .

Силу связи можно измерить с помощью коэффициента сопряженности Шарлье, основываясь на полученном значении χ^2 или минуя его вычисление:

$$r_{\beta} = \sqrt{\frac{\chi^2}{n}} = \frac{ad - bc}{\sqrt{(a + b)(b + d)(c + d)(a + c)}}.$$

Если χ^2 не вычислялся, то достоверность этого коэффициента проверяется по критерию t (Приложение III), для чего предварительно рассчитывают ошибку самого показателя

$$m_{r_{\beta}} = \frac{1 - r_{\beta}^2}{\sqrt{n}},$$

а затем значение $t = r_{\beta} : m_{r_{\beta}}$.

Расчет коэффициента Шарлье настолько прост, что для предварительной «прикидки», когда исследователя интересует не столько точная мера связи, сколько вопрос о достоверности ее наличия, допустимо искусственно сводить многопольную решетку (количественные признаки!) к четырехпольной, разбивая каждый ряд на два класса по положению медианы или по какому-либо иному принципу. При этом, разумеется, происходит потеря части информации и получаемая мера связи оказывается неточной, но на вопрос о достоверности связи таким путем обычно удается получить правильный ответ.

Разберем один пример такого рода. Изучалась связь между высотой растения и числом пар ветвей у *Odontites verna* (Bell.) Dum. Оба признака количественные, варьирующие первый по

непрерывному, а второй по дискретному типу. Для предварительной прикидки вариационный ряд высоты растений был разбит на два класса по положению медии ($M=208$ мм), т. е. по принципу «высокие — низкие», а другой ряд поделен на «ветвистые — неветвистые» без учета числа пар ветвей. Исходя из полученной корреляционной решетки (табл. 30), имеем

$$\chi^2 = \frac{(63 \cdot 46 - 28 \cdot 13)^2 \cdot 150}{(63 + 28)(28 + 46)(13 + 46)(63 + 13)} = 31,898,$$

и связь высокодостоверна, ибо при $v=1$ $\chi_{01}^2 = 6,635$ (Приложение V).

Таблица 30. Корреляция между степенью ветвистости (у) и высотой растений (х) у *Odontites verna* (Bell.) Dum.

у	х	208 мм и выше	До 208 мм	Σ
Ветви есть		63	28	91
Ветвей нет		13	46	59
Σ		76	74	150

Расчет коэффициента сопряженности Шарлье по обеим формулам дает одинаковый результат:

$$r_{\beta} = \sqrt{\frac{\chi^2}{n}} = \sqrt{\frac{31,898}{150}} = 0,461;$$

$$r_{\beta} = \frac{ad - bc}{\sqrt{(a+b)(b+d)(c+d)(a+c)}} = \frac{63 \cdot 46 - 28 \cdot 13}{91 \cdot 74 \cdot 59 \cdot 76} = 0,461.$$

Для сравнения укажем, что коэффициент линейной корреляции, вычисленный для полной корреляционной решетки (8×8 классов), составляет $r=0,562$. Таким образом, потеря части информации при переходе от многопольной к четырехпольной корреляционной решетке привела в данном примере к уменьшению истинной силы связи.

В случае многопольных корреляционных решеток (качественные признаки с числом классов более двух; количественные признаки) расчет критерия χ^2 затрудняется тем, что для каждой (даже пустой!) клетки корреляционной решетки требуется вычислить теоретическую частоту и сравнить ее с эмпирической частотой, записанной в данной клетке.

Вычисление теоретических частот основано на теореме умножения вероятностей: вероятность одновременного появления двух независимых событий равна произведению вероятностей появления каждого из них в отдельности. Например, если корреля-

ционная решетка (табл. 31) заполнена эмпирическими частотами $f_{Aa}, f_{Ba}, \dots, f_{Dd}$, то теоретическая частота для клетки Aa составит

$$f'_{Aa} = \frac{f_A}{n} \cdot \frac{f_a}{n} \cdot n = \frac{f_A f_a}{n}.$$

В обобщенном виде можно записать, что для любой клетки корреляционной решетки, независимо от того, содержит ли она эмпирическую частоту (f_{xy}) или осталась пустой, теоретическая частота (f'_{xy}) вычисляется по формуле

$$f'_{xy} = \frac{f_x f_y}{n},$$

где f_x и f_y — суммарные частоты классов для признаков x и y , а n — объем выборки.

Таблица 31. Схема корреляционной решетки для вычисления χ^2

y	x	a	b	c	d	f_y	$\frac{f_y}{n}$
A		f_{Aa}	f_{Ab}	f_{Ac}	f_{Ad}	f_A	$\frac{f_A}{n}$
B		f_{Ba}	f_{Bb}	f_{Bc}	f_{Bd}	f_B	$\frac{f_B}{n}$
C		f_{Ca}	f_{Cb}	f_{Cc}	f_{Cd}	f_C	$\frac{f_C}{n}$
D		f_{Da}	f_{Db}	f_{Dc}	f_{Dd}	f_D	$\frac{f_D}{n}$
	f_x	f_a	f_b	f_c	f_d	n	—

Практически удобно сначала рассчитать величины $\frac{f_y}{n}$ (правый столбец табл. 31), а уже потом умножить их на эмпирическую суммарную частоту f_x соответствующего столбца. После этого для всех клеток корреляционной решетки последовательно должны быть вычислены значения $f_{xy} - f'_{xy}$, $(f_{xy} - f'_{xy})^2$ и $(f_{xy} - f'_{xy})^2 : n$. Сумма последних величин, отнесенных к значениям теоретических частот, и подставляется в формулу критерия χ^2 :

$$\chi^2 = \sum \frac{(f_{xy} - f'_{xy})^2}{f'_{xy}}.$$

Достоверность связи оценивается (Приложение V) при $v = (k_x - 1)(k_y - 1)$, где k_x и k_y — число классов признаков x и y .

При полном совпадении эмпирических частот с теоретическими, что означает полное отсутствие связи, $\chi^2=0$, однако этот критерий не имеет верхнего предела и потому не может быть использован в качестве непосредственного показателя силы связи.

Учитывая это обстоятельство, К. Пирсон в 1904 г. ввел понятие о *контингенции* или *средней сопряженности* признаков (φ^2), представляющее собой величину χ^2 в расчете на одну особь: $\varphi^2 = \frac{\chi^2}{n}$ (этой величиной мы уже пользовались в одной из формул коэффициента сопряженности Шарлье для случая четырехпольной решетки). Затем Пирсон построил свой коэффициент контингенции:

$$C = \sqrt{\frac{\varphi^2}{1 + \varphi^2}} = \sqrt{\frac{\chi^2}{n + \chi^2}}.$$

При отсутствии связи $C=0$; при слабых связях этот показатель дает почти такие же значения, как коэффициент корреляции (r) или корреляционное отношение (η). Однако его недостатком является то, что при усилении связи он по сравнению с этими показателями начинает давать все более заниженные значения и никогда не достигает единицы. Было показано, что коэффициент Пирсона зависит от числа классов в корреляционной решетке и реагирует на неравенство числа классов по горизонтали и вертикали: чем больше число классов, тем ближе к единице величина C при полной в действительности связи; равенство числа классов усиливает этот эффект. Но даже в случае решетки, содержащей 10×10 классов, максимальное значение этого показателя при полной связи составляет 0,949. При меньшем числе классов оно значительно ниже, например для решетки 5×5 равно 0,894, а для решетки 2×2 — всего 0,707.

Значительно лучшим показателем сопряженности следует считать коэффициент, предложенный А. А. Чупровым:

$$T = \sqrt{\frac{\chi^2}{n\sqrt{v}}} = \sqrt{\frac{\varphi^2}{\sqrt{v}}} = \sqrt{\frac{C^2}{(1 - C^2)\sqrt{v}}},$$

который имеет пределы от 0 до 1 и, хотя также не совпадает полностью со значениями r и η , все же гораздо более сопоставим с ними, чем коэффициент Пирсона.

Как уже говорилось, для расчета χ^2 , строго говоря, требуется вычислять теоретические частоты для всех, в том числе и для пустых, клеток корреляционной решетки. При большом числе классов эта процедура становится весьма трудоемкой. Существует способ косвенного расчета контингенции, минуя вычисление χ^2 , при котором на пустые клетки можно не обращать внимания. Получаемый результат почти не отличается от истинного.

Покажем это на примере изучения связи между длиной и шириной листа у *Melampyrum nemorosum* L. (табл. 32), за-

метив еще раз, что показатели сопряженности могут использоваться и в случае количественных признаков. Порядок работы будет следующим:

1. Эмпирические сопряженные частоты (f_{xy}) помещают в верхней части соответствующих клеток корреляционной ре-

Таблица 32. Расчет средней сопряженности (φ^2) при изучении связи между длиной (y , мм) и шириной (x , мм) листа у *Melampyrum nemorosum* L.

y	x	0-2	3-5	6-8	9-11	12-14	15-17	f_y	$\sum \frac{(f_{xy})^2}{f_x}$	$\sum \frac{(f_{xy})^2}{f_x} : f_y$
15-24	1		2					3	—	—
	(1)		(4)					—	—	—
		1	0,286					—	1,286	0,429
25-34			9	4				—	—	—
			(81)	(16)				—	—	—
			5,786	0,727				—	6,513	0,501
35-44			3	10	1			14	—	—
			(9)	(100)	(1)			—	—	—
			0,643	4,545	0,100			—	5,288	0,378
45-54			7	6	6			19	—	—
			(49)	(36)	(36)			—	—	—
			2,227	3,600	5,143			—	10,970	0,577
55-64			1	3		1		5	—	—
			(1)	(9)		(1)		—	—	—
			0,045	0,900		0,500		—	1,445	0,289
65-74					1	1		2	—	—
					(1)	(1)		—	—	—
					0,143	0,500		—	0,643	0,322
f_x		1	14	22	10	7	2	56	—	$\varphi^2+1=2,495$

сетки, возводят их в квадрат (средние цифры в скобках) и делят на суммарные частоты f_x , записывая результат в нижней части каждой клетки. Например, для клетки, соответствующей классам $x=6-8$ и $y=45-54$, этот результат находим как $(f_{xy})^2 : f_x = 7^2 : 22 = 2,227$.

2. Полученные цифры суммируют по каждой строке, записывая суммы в предпоследнем столбце таблицы. Затем делят их на суммарные частоты f_y и получают данные последнего столбца.

3. Суммируя числа последнего столбца, получают величину, которая превышает φ^2 ровно на единицу. Следовательно, в нашем примере $\varphi^2 = 1,496$.

Подставляя значение φ^2 в соответствующие формулы, имеем:

$$\text{коэффициент Пирсона } C = \sqrt{\frac{1,496}{2,496}} = 0,774;$$

$$\text{коэффициент Чупрова } T = \sqrt{\frac{1,496}{\sqrt{(6-1)(6-1)}}} = 0,547.$$

Достоверность связи легко определить, рассчитав значение $\chi^2 = n\varphi^2 = 56 \cdot 1,496 = 83,776$. При $\nu = (6-1)(6-1) = 25$ стандартное значение $\chi^2 = 44,314$ (Приложение V), и, следовательно, связь высоко достоверна.

Для сравнения укажем, что если в нашем примере (табл. 32), следуя более строгой методике, рассчитать теоретические частоты для всех, в том числе и пустых, клеток, то получаются следующие цифры: $\varphi^2 = 1,423$; $\chi^2 = 79,68$; $C = 0,766$ и $T = 0,533$. Как видим, точный способ, требующий затраты гораздо большего количества времени, приводит к получению практически того же самого результата.

Глава 5

ИЗУЧЕНИЕ ЗАВИСИМОСТИ ИЗМЕНЕНИЯ ОДНОГО ПРИЗНАКА ОТ ИЗМЕНЕНИЯ ДРУГОГО (РЕГРЕССИОННЫЙ АНАЛИЗ)

5.1. ПОНЯТИЕ О РЕГРЕССИИ

Как уже говорилось, корреляционный анализ, изучающий силу обоюдной связи между признаками, не дает ответа на вопрос о том, как именно (на сколько в среднем) изменяется один признак при определенном изменении другого. Однако при рассмотрении такого показателя, как корреляционное отношение η (раздел 4.3.1), мы фактически уже перешли к изучению зависимости изменения одного признака от изменения другого.

Анализ регрессии тесно соприкасается с корреляционным анализом, является его логическим продолжением и, используя особые методы, развивает и углубляет представление о корреляционной связи в следующих важных направлениях. Во-первых, приемы регрессионного анализа позволяют выявить и гра-

фически отобразить зависимость изменения одного признака от изменения другого (регрессию y по x , обозначаемую y/x , и регрессию x по y , обозначаемую x/y). Во-вторых, на основе составления и решения уравнений регрессии становится возможным выравнивание эмпирических линий регрессии, т. е. моделирование наблюдаемой зависимости путем подбора соответствующей функции, график которой и представляет собой теоретическую линию регрессии. В-третьих, если подобранная функция не только формально описывает связь в интервале интерполяции эмпирических данных, но отражает биологическую сущность явления, то открывается перспектива прогнозирования значений признака в зоне экстраполяции, т. е. за пределами ряда фактически сделанных наблюдений.

Итак, под регрессией подразумевается зависимость изменений одного признака от изменений другого или нескольких признаков. В соответствии с этим регрессия, подобно корреляции, может быть парной (простой) или множественной, а в зависимости от формы связи — линейной или нелинейной.

В отличие от корреляционного анализа, требующего достаточно большого объема выборки, анализ регрессии возможен и при наличии всего нескольких пар сопряженных наблюдений, однако его имеет смысл производить лишь при обнаружении достоверных и достаточно сильных (порядка $r \geq 0,7$) связей между признаками.

Начинать регрессионный анализ целесообразно с построения эмпирических линий регрессии.

5.2. ЭМПИРИЧЕСКИЕ ЛИНИИ РЕГРЕССИИ

Точки эмпирических линий регрессии вычисляются как взвешенные средние арифметические по строкам и столбцам корреляционной решетки. Техника этой работы была изложена в связи с расчетом корреляционного отношения (раздел 4.3.1), поэтому на ней мы задерживаться не будем.

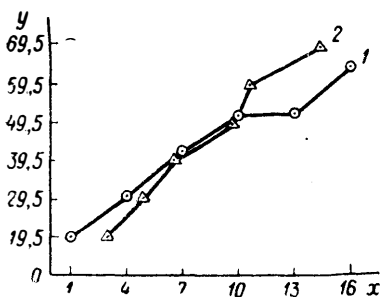


Рис. 9. Эмпирические линии регрессии длины (y) и ширины (x) листа *Melampyrum polonicum* (Beauv.) Соб.

1 — регрессия y/x ; 2 — регрессия x/y .

Возьмем, однако, новый пример. Изучалась зависимость между длиной и шириной листа у *Melampyrum polonicum* (Beauv.) Соб. Корреляционная решетка и рассчитанные точки эмпирических линий регрессии y/x и x/y даны в табл. 33, а сами эмпирические линии регрессии нанесены на график (рис. 9).

Графическое изображение эмпирических линий регрессии надо считать обязательным,

так как по их внешнему виду можно сделать некоторые предварительные заключения о характере связи. При полном отсутствии связи эмпирические линии регрессии, пересекаясь под прямым углом, располагаются параллельно осям графика. Чем сильнее связь, тем меньше угол между линиями: при полной связи они параллельны друг другу. Направление эмпирических линий регрессии говорит о знаке связи, а их конфигурация ориентировочно указывает на степень линейности связи (при этом надо оценивать основную тенденцию хода кривых на графике, мысленно сглаживая их изломы). Обладая определенными навыками, исследователь по форме эмпирической линии регрессии может достаточно уверенно судить о том,

Таблица 33. Корреляция между длиной (y) и шириной (x) листа у *Metapyrum polonicum* (Beauv.) Sod и точки эмпирических линий регрессии y/x и x/y

$y \quad x$	1	4	7	10	13	16	f_y	x/y
19,5	1	2					3	3,0
29,5		9	4				13	4,9
39,5		3	10	1			14	6,6
49,5			7	6	6		19	9,8
59,5			1	3		1	5	10,6
69,5						1	2	14,5
f_x	1	14	22	10	7	2	$n=56$	—
y/x	19,5	30,2	41,8	51,5	52,4	64,5	—	—

какая теоретическая функция окажется пригодной для ее выравнивания.

Из рассматриваемого примера (рис. 9) видно, что связь между длиной и шириной листа *M. polonicum* положительна (с увеличением значений одного признака возрастают и значения другого) и достаточно сильна (угол между линиями невелик). Форма кривых наводит на мысль о некоторой нелинейности связи, однако для начала целесообразно попытаться описать наблюдаемую регрессию более простыми линейными методами.

5.3. ЛИНЕЙНАЯ РЕГРЕССИЯ

Определение линейной регрессии совпадает с определением линейной корреляции: равномерным изменениям одного признака соответствуют равномерные в среднем изменения другого

признака. Указанием на линейность служит возможность проведения на графике от руки прямой линии таким образом, чтобы точки эмпирической линии регрессии располагались по обе стороны от прямой и по возможности ближе к ней.

5.3.1. Уравнение линейной регрессии и метод наименьших квадратов. Построение теоретической линии регрессии. Уравнением линейной регрессии служит уравнение прямой линии: $y = a + bx$, где y — значение зависимой переменной (признака); x — значение независимой переменной (признака или фактора, влияющего на первый признак); a — начальное значение y при $x = 0$; b — угловой коэффициент (тангенс угла наклона линии регрессии к оси абсцисс, отражающий пропорциональную зависимость y от x).

Задача состоит в нахождении неизвестных параметров a и b . Существует несколько способов решения этой задачи, из кото-

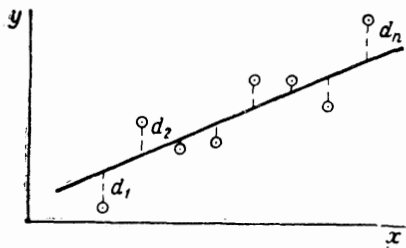


Рис. 10. Сущность метода наименьших квадратов (схема).

Пояснения в тексте.

рых лучшим является метод наименьших квадратов. Суть метода состоит в том, что находят такую прямую линию, которой соответствует наименьшая сумма квадратов отклонений от нее до точек эмпирической линии регрессии.

Иначе говоря, сумма $d_1^2 + d_2^2 + \dots + d_n^2$ (рис. 10) должна иметь наименьшее возможное значение. Для этого составляется и решается система

стольких так называемых нормальных уравнений, сколько неизвестных требуется определить. В случае линейной регрессии y/x система состоит из двух уравнений и выглядит следующим образом:

$$\begin{cases} na + b \sum x = \sum y, \\ a \sum x + b \sum x^2 = \sum xy, \end{cases}$$

где y — точки эмпирической линии регрессии y/x ; n — число пар сопряженных наблюдений; x — значения признака.

Необходимые для подстановки в нормальные уравнения суммы ($\sum x$, $\sum y$, $\sum x^2$, $\sum xy$) удобно рассчитывать в табличной форме. Применительно к нашему примеру регрессии длины листа *M. polonicum* на его ширину (табл. 33) соответствующие расчеты выполнены в левой части табл. 34, причем требуемые суммы указаны в нижней ее строке. Подставляем эти суммы в систему нормальных уравнений:

$$\begin{cases} 6a + 51b = 259,9, \\ 51a + 591b = 2661,1 \end{cases}$$

и решаем ее обычным алгебраическим путем. Для этого, разде-

лив первое уравнение на 6, а второе на 51, освободимся от коэффициентов при неизвестном a :

$$\begin{cases} a + 8,5b = 43,32, \\ a + 11,59b = 52,18 \end{cases}$$

и после вычитания первого уравнения из второго получим $3,09b = 8,86$, откуда следует, что $b = 2,87$. Подставив значение $b = 2,87$ в любое из ранее полученных уравнений, находим, что $a = 18,92$. Теперь можно записать искомое уравнение: $y = 18,92 + 2,87x$.

Таблица 34. Выравнивание эмпирической линии регрессии длины листа (y) на его ширину (x) у *Melanopygia colonicum* (Beauv.) с помощью уравнения прямой линии (y')

Расчеты для определения параметров уравнения				Построение теоретической линии регрессии		Расчет критерия χ^2		
x	y	x^2	xy	bx	$a+bx=y'$	$y-y'$	$(y-y')^2$	$\frac{(y-y')^2}{y'}$
1	19,5	1	19,5	2,87	21,79	-2,29	5,2441	0,2407
4	30,2	16	120,8	11,48	30,40	-0,20	0,0400	0,0013
7	41,8	49	292,6	20,09	39,01	+2,79	7,7841	0,1995
10	51,5	100	515,0	28,70	47,62	+3,88	15,0544	0,3161
13	52,4	169	681,2	37,31	56,23	-3,83	14,6689	0,2609
16	64,5	256	1032,0	45,92	64,84	-0,34	0,1156	0,0018
51	259,9	591	2661,1	-	259,89	+0,01	42,9071	1,0203

Посредством обратного расчисления, последовательно подставляя в найденное уравнение значения $x=1$, $x=4$, $x=7$ и т. д. (табл. 34), рассчитываем точки (y') теоретической линии регрессии. Теоретическую линию регрессии для ее визуального сравнения с эмпирической полезно нанести на график (рис. 11, а), а степень их совпадения можно проверить посредством расчета критерия χ^2 (см. табл. 34). Полученное значение $\chi^2 = 1,0203$ далеко не достигает стандартных значений этого критерия, составляющих при $v = n - 1 = 5$ $\chi_{05}^2 = 11,1$ и $\chi_{01}^2 = 15,1$ (Приложение V), указывая на достаточно хорошее соответствие теоретической линии регрессии эмпирическому ряду.

Ошибка уравнения регрессии рассчитывается по формуле

$$m_{y/x} = \sqrt{\frac{\sum(y - y')^2}{n - k}},$$

где n — число точек линии регрессии; k — число коэффициентов в уравнении регрессии, включая свободный член.

В нашем примере $\Sigma(y-y')^2=42,9071$ (см. табл. 34); $n=6$ и $k=2$, поэтому

$$m_{y/x} = \sqrt{\frac{42,9071}{6-2}} = \pm 3,28 \text{ мм.}$$

Это — средний показатель точности, с которым «работает» выведенное нами уравнение регрессии.

По аналогии с вышеизложенным рассчитывается и вторая теоретическая линия регрессии (x/y), для чего в системе нормальных уравнений признаки x и y следует поменять местами:

$$\begin{cases} na + b \Sigma y = \Sigma x, \\ a \Sigma y + b \Sigma y^2 = \Sigma xy. \end{cases}$$

Исходя из данных табл. 33, читатель может самостоятельно выполнить необходимые расчеты. Поэтому мы приведем лишь итоговое уравнение теоретической линии регрессии x/y : $x = -1,66 + 0,22y$.

Удовлетворительно интерполируя эмпирические данные, уравнение прямой линии в нашем примере не в состоянии, однако, обеспечить экстраполяцию за пределы эмпирического ряда наблюдений. В этом легко убедиться, если, например, в уравне-

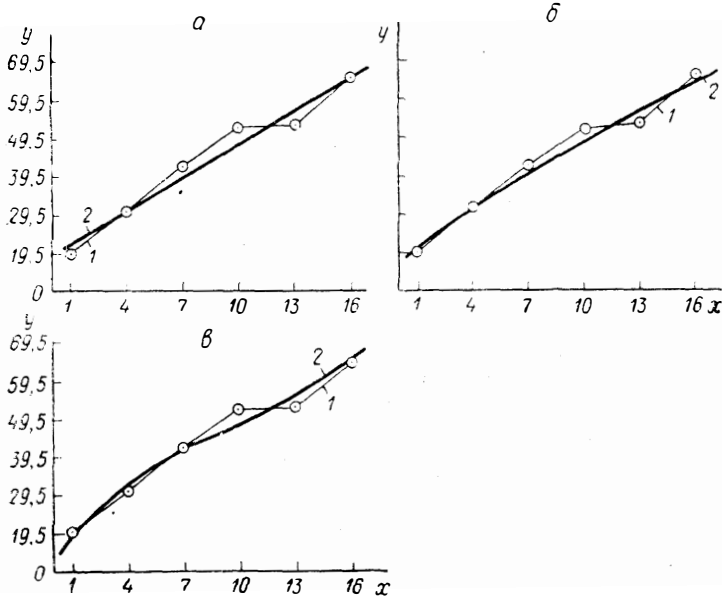


Рис. 11. Выравнивание эмпирической линии регрессии y/x (1) уравнением прямой линии (2) по данным табл. 34 (а), уравнением параболы второй степени (2) по данным табл. 35 (б) и уравнением параболы третьей степени (2) по данным табл. 36 (в).

нии $y = 18,92 + 2,87x$ придать ширине листа (x) нулевое значение: длина листа (y) при этом окажется равной 18,92 мм, что лишено биологического смысла. Таким образом, линейная функция не отражает полностью биологическую сущность связи между длиной и шириной листа. Если нас не удовлетворяет формальная интерполяция, мы должны продолжить поиск и найти такую функцию, которая наряду с интерполяцией позволяла бы производить экстраполяцию (и обеспечить нулевое значение длины листа при нулевой его ширине). В дальнейшем мы еще вернемся к этому вопросу.

5.3.2. Коэффициенты регрессии. Угловым коэффициентом (b) в уравнении линейной регрессии, отражающий пропорциональную зависимость между признаками, называется коэффициентом регрессии (R). Для признаков длины и ширины листа у *M. polonicum* мы получили в предыдущем разделе два уравнения регрессии (y/x и x/y), из которых следует, что $R_{y/x} = 2,87$, а $R_{x/y} = 0,22$.

Биологический смысл коэффициентов регрессии состоит в том, что они представляют собой меру изменения одного признака от определенного изменения другого. В нашем примере значение $R_{y/x} = 2,87$ говорит о том, что с увеличением ширины листа (x) на 1 мм (принятая точность измерения) его длина (y) увеличивается в среднем на 2,87 мм. Второй коэффициент ($R_{x/y} = 0,22$) свидетельствует о том, что при увеличении длины листа (y) на 1 мм его ширина (x) возрастает в среднем на 0,22 мм.

Коэффициенты регрессии могут быть вычислены и без составления уравнения регрессии. Один из способов основан на предварительном вычислении тех сумм, которые применительно к регрессии y/x указаны в нижней строке табл. 34 (для определения величины $R_{x/y}$ читателю придется составить аналогичную таблицу, поменяв местами x и y , но он уже выполнил это, если последовал нашему совету рассчитать уравнение регрессии x/y). Соответствующие формулы дают следующие результаты:

$$R_{y/x} = \frac{n \sum xy - \sum x \cdot \sum y}{n \sum x^2 - (\sum x)^2} = \frac{6 \cdot 2661,1 - 51 \cdot 259,9}{6 \cdot 591 - 51^2} = 2,87;$$

$$R_{x/y} = \frac{n \sum xy - \sum x \cdot \sum y}{n \sum y^2 - (\sum y)^2} = \frac{6 \cdot 2661,1 - 51 \cdot 259,9}{6 \cdot 13631,5 - 276^2} = 0,26.$$

Значение первого коэффициента полностью совпадает с ранее рассчитанным, а небольшое отклонение второго (против величины 0,22, полученной ранее) объясняется потерей десятичных знаков при промежуточных округлениях.

Другой способ определения коэффициентов регрессии может быть использован тогда, когда предварительно производился корреляционный анализ и известно значение коэффициента

корреляции (r) между признаками, а также их средние квадратичные отклонения:

$$R_{y/x} = r \frac{\sigma_y}{\sigma_x}; \quad R_{x/y} = r \frac{\sigma_x}{\sigma_y},$$

где σ_y и σ_x — «полные» сигмы, взятые с учетом классового интервала.

Эти формулы наглядно показывают тесную связь корреляционного анализа с регрессионным: перемножение их левых и правых частей приводит к выражению

$$r = \sqrt{R_{y/x} \cdot R_{x/y}},$$

из которого следует вывод, что коэффициент корреляции есть среднее геометрическое из двух коэффициентов регрессии. Сказанное соответствует биологическому смыслу показателей: коэффициент корреляции является относительной величиной, показывающей обоюдную степень связи между признаками, а коэффициенты регрессии — величины, конкретизирующие зависимость каждого из признаков от другого.

В примере с *M. polonicum* коэффициент корреляции $r = 0,799$, а средние квадратичные отклонения признаков $\sigma_y = 11,8$ мм, $\sigma_x = 3,3$ мм. Отсюда

$$R_{y/x} = 0,799 \frac{11,8}{3,3} = 2,86 \quad \text{и} \quad R_{x/y} = 0,799 \frac{3,3}{11,8} = 0,22,$$

и, как видим, результат практически тождествен тому, который ранее был получен иными способами.

Как всякая выборочная величина коэффициент регрессии имеет свою ошибку репрезентативности и с ее помощью может быть оценена его достоверность. Покажем это применительно к коэффициенту регрессии $R_{y/x}$ в нашем примере. Ошибка показателя вычисляется следующим образом:

$$m_{R_{y/x}} = \frac{\sigma_y}{\sigma_x} \sqrt{\frac{1-r^2}{n-2}} = \frac{11,8}{3,3} \sqrt{\frac{1-0,799^2}{56-2}} = 0,29.$$

Далее можно использовать критерий t :

$$t = \frac{R_{y/x}}{m_{R_{y/x}}} = \frac{2,87}{0,29} = 9,9,$$

и при $v = n - 2 = 56 - 2 = 54$ (здесь n — объем выборки по табл. 33) полученное значение t высоко достоверно, так как значительно превышает $t_{01} = 2,68$ (Приложение III).

5.4. НЕЛИНЕЙНАЯ РЕГРЕССИЯ

Когда одинаковым приращениям одного признака сопутствуют неодинаковые, но изменяющиеся по определенному закону приращения другого признака, регрессия, так же как корреля-

ция, оказывается нелинейной. Внешним признаком нелинейной регрессии служит то, что эмпирические линии регрессии на графике выглядят кривыми различной конфигурации. При небольших отклонениях от линейности допустимо использование более простых приемов линейной регрессии, но в сомнительных случаях необходимо проверка линейности связи (раздел 4.3.2).

5.4.1. Параболическое интерполирование. Уравнение параболы n -й степени: $y = a + bx + cx^2 + dx^3 + \dots + mx^n$ представляет собой очень гибкую и удобную для расчетов функцию, широко используемую для интерполяции эмпирических данных. Ее можно «оборвать» на любом члене и получить последовательно выражения: $y = a$ (отсутствие зависимости), $y = a + bx$ (уравнение линейной зависимости), $y = a + bx + cx^2$ (парабола второй степени) и т. д. Если имеется n точек, то парабола n -й степени имеет $n - 1$ перегибов: при этом концы кривой уходят в бесконечность при четном n в одну сторону, при нечетном n — в разные стороны.

Нарращивая число членов уравнения (используя при n точках параболу степени $n - 1$), в принципе можно описать почти все случаи нелинейной регрессии (т. е. добиться совпадения теоретической линии регрессии с эмпирическими точками). Вычисление парабол высоких степеней сопровождается резким увеличением громоздкости расчетов, но еще хуже то, что полученные при этом уравнения чаще всего не поддаются биологической интерпретации: имеет место лишь формальная «подгонка» эмпирических данных под достаточно гибкую математическую функцию.

Критерием правильности выбора функции может служить только удовлетворительное биологическое истолкование полученного результата. Если функция выбрана правильно и соответствует биологической сути описываемой зависимости, то она должна не только хорошо интерполировать эмпирические данные, но и допускать экстраполяцию за пределы наблюдаемого ряда. Добиться удовлетворительной экстраполяции удается далеко не всегда. Но нельзя выдавать интерполяционную формулу, хорошо выравнивающую эмпирический ряд, за модель всего биологического явления в целом.

Общая схема расчета одина для парабол любой степени. Параболы степени выше третьей практически используются редко, требуют расчетов на ЭВМ, и мы не будем их рассматривать.

Обратимся к уравнению параболы второй степени $y = a + bx + cx^2$, используя данные прежнего примера (табл. 33) регрессии длины на ширину листа у *M. polonicum*. Уже говорилось, что уравнение прямой (т. е. парабола первой степени) довольно хорошо интерполирует эту зависимость, но не допускает экстраполяцию в меньшую сторону (при ширине листа $x = 0$ его длина оказывается равной $y = 18,92$ мм!). Посмотрим, что дает использование параболы второй степени.

Расчет ведется методом наименьших квадратов, но в отличие от линейной регрессии придется решать систему не из двух, а уже из трех нормальных уравнений с тремя (a , b и c) неизвестными:¹

$$\begin{cases} na + b \sum x + c \sum x^2 = \sum y, \\ a \sum x + b \sum x^2 + c \sum x^3 = \sum xy, \\ a \sum x^2 + b \sum x^3 + c \sum x^4 = \sum x^2y. \end{cases}$$

Как видно, по сравнению с уравнением прямой линия задача усложняется нахождением сумм $\sum x^3$, $\sum x^4$ и $\sum x^2y$. Расчет нужных сумм показан в левой части табл. 35.

Таблица 35. Вывравнивание эмпирической линии регрессии длины листа (y) на его ширину (x) у *Melampyrum polonicum* (Beauv.) Soó уравнение параболы второй степени (y')

Расчеты для определения параметров уравнения			Построение теоретической линии регрессии			Расчет критерия χ^2		
x^3	x^4	x^2y	bx	cx^2	$a+bx+cx^2=y'$	$y-y'$	$(y-y')^2$	$\frac{(y-y')^2}{y'}$
1	1	19,5	3,55	-0,04	20,58	-1,08	1,1664	0,0567
64	256	483,2	14,20	-0,64	30,63	-0,43	0,1849	0,0060
343	2401	2048,2	24,85	-1,96	39,96	+1,84	3,3856	0,0847
1000	10000	5150,0	35,50	-4,00	48,57	+2,93	8,5849	0,1768
2197	28561	8855,6	46,15	-6,76	56,46	-4,06	16,4836	0,2920
4096	65536	16512,0	56,80	-10,24	63,63	+0,87	0,7569	0,0119
7701	106755	33068,5	—	—	259,83	+0,07	30,5623	0,6281

Примечание. Значения x , y , x^2 , xy см. в табл. 34.

Исходя из данных этой таблицы, составляем систему нормальных уравнений

$$\begin{cases} 6a + 51b + 591c = 259,9, \\ 51a + 591b + 7701c = 2661,1, \\ 591a + 7701b + 106755c = 33068,5, \end{cases}$$

решая которую обычным алгебраическим путем, находим: $c = -0,04$; $b = 3,55$; $a = 17,07$.

Следовательно, уравнение параболы второй степени выглядит так: $y = 17,07 + 3,55x - 0,04x^2$.

¹ Аналогичным образом, путем наращивания степеней x и числа неизвестных, составляются системы нормальных уравнений и для парабол более высокого порядка, причем количество уравнений в системе составляет $n+1$, если n — степень параболы.

Путем обратного расчисления (табл. 35) можно высчитать точки (y') теоретической линии регрессии и построить соответствующий график (рис. 11, б). Парабола второй степени вполне удовлетворительно выравнивает эмпирический ряд, что подтверждается также расчетом критерия χ^2 (табл. 35): при $\nu=5$ значение $\chi^2=0,6281$ значительно меньше стандартных значений $\chi_{0,05}^2=11,1$ и $\chi_{0,01}^2=15,1$ (Приложение V).

Средняя ошибка найденного уравнения регрессии составляет:

$$m_{y/x} = \sqrt{\frac{(y - y')^2}{n - k}} = \sqrt{\frac{30,5623}{6 - 3}} = \pm 3,19 \text{ мм.}$$

Она чуть меньше, чем в случае применения линейной модели (с. 92), но парабола второй степени по-прежнему не допускает экстраполяции за пределы эмпирического ряда: при $x=0$ длина листа $y=17,07$ мм. Таким образом, с помощью этой функции невозможно добиться естественного с биологической точки зрения соотношения $y=0$ при $x=0$.

Попробуем повысить порядок параболы, взяв уравнение третьей степени: $y=a+bx+cx^2+dx^3$.

В этом случае придется решать систему, состоящую из четырех нормальных уравнений:

$$\begin{cases} an + b \sum x + c \sum x^2 + d \sum x^3 = \sum y, \\ a \sum x + b \sum x^2 + c \sum x^3 + d \sum x^4 = \sum xy, \\ a \sum x^2 + b \sum x^3 + c \sum x^4 + d \sum x^5 = \sum x^2y, \\ a \sum x^3 + b \sum x^4 + c \sum x^5 + d \sum x^6 = \sum x^3y. \end{cases}$$

Применительно к нашему примеру (табл. 36) это означает:

$$\begin{cases} 6a + 51b + 591c + 7701d = 259,9, \\ 51a + 591b + 7701c + 106755d = 2661,1, \\ 591a + 7701b + 106755c + 1537701d = 33068,5, \\ 7701a + 106755b + 1537701c + 22725771d = 447104,5. \end{cases}$$

Решение этой системы дает значения неизвестных: $d=0,0122$; $c=-0,3804$; $b=6,1357$; $a=12,9739$.

Итак, искомое уравнение параболы третьей степени: $y = 12,9739 + 6,1357x - 0,3804x^2 + 0,0122x^3$.

Обратное расчисление (табл. 36) позволяет получить точки (y') теоретической линии регрессии и построить по ним соответствующий график (рис. 11, в). Вполне удовлетворительно выполненная интерполяция подтверждается также значением $\chi^2=0,5019$ (табл. 36), далеко не достигающим при $\nu=5$ стандартных значений $\chi_{0,05}^2=11,1$ и $\chi_{0,01}^2=15,1$ (Приложение V).

Средняя ошибка уравнения регрессии составляет

$$m_{y/x} = \sqrt{\frac{\sum(y - y')^2}{n - k}} = \sqrt{\frac{22,5021}{6 - 4}} = \pm 3,35 \text{ мм,}$$

т. е. мало отличается от ошибок уравнений прямой линии (3,28) и параболы второго порядка (3,19). Однако парабола третьей степени по-прежнему не допускает экстраполяции, ибо при нулевой ширине листа ($x=0$) его длина $y=12,97$ мм, что противоречит здравому смыслу.

Таблица 36. Вывравнивание эмпирической (y) на его ширину (x) у *Melat* уравнением параболы

Расчеты для определения параметров уравнения			Построение теоретической	
x^1	x^2	x^3y	bx	cx^2
1	1	19,5	6,1357	-0,3804
1624	4096	1932,8	24,5428	-6,0864
16807	117649	14337,4	42,9499	-18,6396
100000	1000000	51500,0	61,3570	-38,0400
371293	4826809	115122,8	79,7641	-64,2876
1048576	16777216	264192,0	98,1712	-97,3824
1537701	22725771	447104,5	—	—

Примечание. Значения x , y , xu , x^2 , x^2u , x^3 , x^4 см. в табл. 34 и 35.

Дальнейшее повышение степени параболы, сопровождаемое увеличением громоздкости расчетов, вряд ли способно существенно улучшить интерполяцию, а вопрос о соответствии параболической функции внутренней сущности изучаемой зависимости, скорее всего, таким путем решить не удастся.

В определенной мере здесь может помочь так называемый прием конечных разностей. Дело в том, что у параболы n -й степени конечная разность ее членов постоянна: $\Delta n_y = \text{const}$, а следующая разность равна нулю: $\Delta n_{y+1} = 0$. Поясним сказанное простым примером параболы четвертой степени $y=x^4$, где для удобства отброшены члены низшего порядка (x , x^2 и x^3), а также коэффициенты (табл. 37). Из таблицы видно, что разности (Δ) возрастающих порядков стремятся сблизиться и становятся одинаковыми на той разности, которая соответствует степени параболы.

Отсюда следует, что если мы хотим установить соответствие параболы внутренней сущности биологического явления, то эмпирическую линию регрессии надо «разложить» на последовательные разности ($\Delta_1, \Delta_2 \dots \Delta_n$) и посмотреть, какие из них более или менее стабилизируются (полного совпадения, разумеется, ожидать нельзя). Если окажутся слабо «пульсирующие» разности, то можно воспользоваться параболой соответствующей степени.

В нашем примере с длиной и шириной листа *M. polonicum*

линии регрессии длины листа,
pyrum polonicum (Beauv.) Soo
третьей степени (y')

линии регрессии		Расчет критерия χ^2		
dx^3	$a+bx+cx^2+dx^3=y'$	$y-y'$	$(y-y')^2$	$\frac{(y-y')^2}{y'}$
0,0122	18,74	+0,76	0,5776	0,0308
0,7808	32,21	-2,01	4,0401	0,1254
4,1846	41,47	+0,33	0,1089	0,0026
12,2000	48,49	+3,01	9,0601	0,1868
26,8034	55,25	-2,85	8,1225	0,1470
49,9712	63,73	+0,77	0,5929	0,0093
—	259,89	+0,01	22,5021	0,5019

стабилизировать разности точек эмпирической линии регрессии (y/x) не удается (табл. 38), и, следовательно, параболическое интерполирование в этом случае описывает закономерность формально, не допуская экстраполяции.

Между тем существует эмпирически найденный сравнительно небольшой набор функций, удачно моделирующих основные биологические закономерности. Ориентировочные приемы, облегчающие подбор нужной формулы, показаны в следующем разделе. Забегая вперед, укажем, что регрессия размерных признаков органов растений и животных (а именно такой пример и рассматривался выше) обычно хорошо моделируется аллометрической функцией (см. раздел 6.3).

5.4.2. Другие интерполяционные формулы. При подборе интерполяционных формул следует обращаться к графическому анализу, сравнивая эмпирические линии регрессии с графиками

Таблица 37. Разложение параболы $y=x^4$ на последовательные разности (Δ)

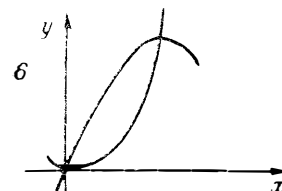
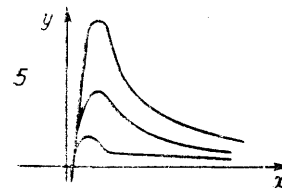
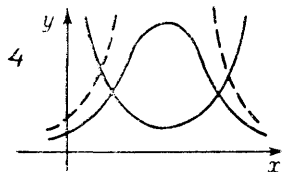
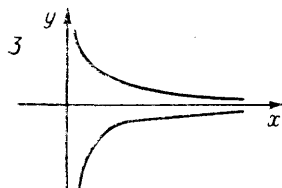
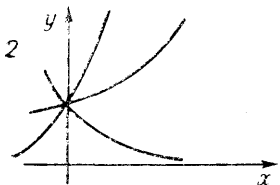
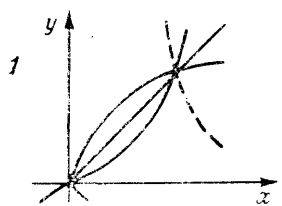
x	x^4	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5
1	1					
2	16	15				
3	81	65	50			
4	256	175	110	60		
5	625	369	194	84	24	0
6	1296	671	302	108	24	0
7	2401	1105	434	132	24	0
8	4096	1695	590	156		

Примечание. Разложение параболы ограничено значением $x=8$, но этот процесс можно и продолжить.

Таблица 38. Разложение эмпирической линии регрессии длины листа (y) на его ширину (x) у *Melampyrum polonicum* (Beauv.) Соо на последовательные разности (Δ)

y, x	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5
19,5					
30,2	10,7				
41,8	11,6	0,9			
51,5	9,7	1,9	1,0	5,9	
52,4	0,9	8,8	6,9	4,5	1,4
64,5	12,1	11,2	2,4		

встречающихся в биологии функций (рис. 12). В затруднительных случаях рекомендуется поиск выравнивающей функции посредством трансформации переменных. Этот прием заключается в построении нового графика, оси которого градуируются уже не натуральными значениями x и y , а такими значениями переменных (см. рис. 12), чтобы кривая на графике превратилась в прямую линию. Критерием соответствия данной формулы эмпирической линии регрессии служит рассеяние точек последней приблизительно вдоль прямой, которая на графике может располагаться как угодно. Если этого не происходит, надо переходить к испытанию другой формулы.



Функции	Преобразования для линеаризации	
	x	y
$y = ax^b$	$\lg x$	$\lg y$
$y = ae^{\delta x}$	x	$\lg y$
$y = a + \frac{b}{x}$	$\frac{1}{x}$	y
$y = \frac{1}{a + bx + cx^2}$	x	$\left(\frac{1}{y}\right)'$
$y = \frac{x}{a + bx + cx^2}$	x	$\left(\frac{x}{y}\right)'$
$y = ax + bx^3$	x^2	$\frac{y}{x}$

Рис. 12. Подбор интерполяционной формулы посредством линеаризации уравнения.

Пояснения в тексте

Представляется полезным прокомментировать биологический смысл некоторых формул.

Формула 1 (рис. 12) обычно хорошо описывает зависимость между размерами органов растений и животных, а также распределение организмов внутри экологических стаций. Она носит название аллометрической функции и специально рассматривается в дальнейшем (раздел 6.3).

Формулы 4 и 5 (рис. 12) характеризуют взаимоотношения

Таблица 39. Вычисление параметров уравнения линейной Фг. на высоту растения (x)

Расчеты для определения параметров уравнения							
y	x	z	y^2	x^2	z^2	xy	xz
4,5	41,7	7,6	20,25	1738,89	57,76	187,65	316,92
14,5	55,3	10,3	210,25	3058,09	106,09	801,85	569,59
24,5	73,2	12,5	600,25	5358,24	156,25	1793,40	915,00
34,5	99,1	16,0	1190,25	9820,81	256,00	3418,95	1585,60
44,5	110,5	13,0	1980,25	12210,25	169,00	4917,25	1436,50
54,5	110,5	18,0	2970,25	12210,25	324,00	6022,25	1989,00
177,0	490,3	77,4	6971,50	44396,53	1069,10	17141,35	6812,61

ареалов и количества видов в родах, но значение их этим не исчерпывается. Эти формулы (особенно формула 5) хорошо описывают известное в зоологии «правило оптимума», отражающее зависимость размеров тела животных от средней температуры данного пояса и, следовательно, от географической широты.

Помимо изображенных на рис. 12 функций для описания биологических явлений могут применяться и другие формулы. Так, рост органов растений и животных во времени описывается S-образными кривыми, которые входят в обобщенный класс кривых роста и рассматриваются в разделе 6.2.

Циклические явления (временные изменения численности популяций, так называемые «волны жизни», а также временные ряды фенологических наблюдений, связывающие сроки наступления фенофаз растений с периодическими колебаниями климата, и др.) могут быть описаны периодическими функциями типа

$$y = a + b \sin\left(\frac{360}{r} x\right) + c \cos\left(\frac{360}{r} x\right),$$

где r — число наблюдений за цикл через равные промежутки

времени; x — порядковый номер наблюдения от $x_1=0$ до $x_i=r-1$.

В некоторых случаях оказываются полезными «гибридные» формулы, например сочетание прямой линии с параболой $y = a + bx + c \lg x$ или функция типа $\lg y = a + b \lg x + c \lg x^2$.

Применение перечисленных функций связано с использованием метода наименьших квадратов, для чего любая функция предварительно должна быть приведена к линейному виду.

множественной регрессии длины соцветия (y) у *Odontites litoralis* и длину наиболее крупного листа (z)

Обратное расчисление				Расчет критерия χ^2		
yz	bx	cz	$a + bx + cz = y'$	$y - y'$	$(y - y')^2$	$\frac{(y - y')^2}{y'}$
34,20	20,8625	7,9602	3,93	+0,57	0,3249	0,0827
149,35	27,6666	10,7882	13,56	+0,94	0,8836	0,0652
306,25	36,6220	13,0925	24,82	-0,32	0,1024	0,0041
552,00	49,5797	16,7584	41,44	-6,94	48,1636	1,1623
578,50	55,2832	13,6162	44,01	+0,49	0,2401	0,0055
981,00	55,2832	18,8532	49,24	+5,26	27,6676	0,5619
2601,30	—	—	177,00	0,00	77,3822	1,8817

5.5. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Зависимость изменения одного признака от одновременного изменения двух или нескольких других признаков изучается методами множественной регрессии. С увеличением числа признаков и с переходом к нелинейной множественной регрессии сложность и громоздкость вычислений быстро нарастают, требуя использования ЭВМ. Поэтому ниже рассматривается только простейший, но чаще всего встречающийся в биометрической практике случай линейной множественной регрессии одного признака по двум другим.

Изучалась зависимость длины соцветия *Odontites litoralis* Fr. от высоты растения и длины самого крупного листа. Задачу сформулируем следующим образом: можно ли, зная высоту растения и длину листа, достаточно точно определить длину соцветия?

В левой части табл. 39 представлены исходные данные: y — центральные классовые значения признака длины соцветия; x — точки эмпирической линии регрессии высоты растений на длину соцветия (x/y); z — точки эмпирической линии регрессии длины

листа на длину соцветия (z/y). Там же произведен расчет необходимых сумм, указанных, как обычно, в нижней строке таблицы.

Уравнение линейной множественной регрессии имеет вид: $y = a + bx + cz$, а система нормальных уравнений в этом случае выглядит следующим образом:

$$\begin{cases} na + b \sum x + c \sum z = \sum y, \\ a \sum x + b \sum x^2 + c \sum xz = \sum xy, \\ a \sum z + b \sum xz + c \sum z^2 = \sum yz. \end{cases}$$

Учитывая, что $n=6$ (количество сопряженных значений признаков) и беря из табл. 39 необходимые суммы, подставляем эти величины в систему нормальных уравнений:

$$\begin{cases} 6a + 490,3b + 77,40c = 177,00, \\ 490,3a + 44396,53b + 6812,61c = 17141,35, \\ 77,4a + 6812,61b + 1069,10c = 2601,30. \end{cases}$$

Решая эту систему, получаем искомые значения неизвестных: $a = -24,8944$, $b = 0,5003$ и $c = 1,0474$.

Следовательно, уравнение, отражающее зависимость длины соцветия у *O. litoralis* от высоты растения и длины самого крупного листа, можно записать так: $y = -24,8944 + 0,5003x + 1,0474z$.

Путем обратного расчисления (табл. 39) находим теоретические значения (y'), которые на глаз достаточно согласуются с соответствующими эмпирическими значениями признака (y). Подтвердить это соответствие помогает критерий χ^2 (табл. 39): полученная величина $\chi^2 = 1,88$ далеко не достигает стандартных (при $\nu = n - 1 = 5$) значений $\chi_{05}^2 = 11,1$ и $\chi_{01}^2 = 15,1$ (Приложение V).

Таким образом, подставляя в полученное выше уравнение фактически наблюдаемые у той или иной особи значения x (высота растения) и z (длина самого крупного листа), можно надеяться на достаточно точное определение y (длины соцветия). Имеется возможность определить ошибку такого предположения (ошибку уравнения линейной множественной регрессии). Для этого надо по данным табл. 39 рассчитать средние арифметические:

$$M_y = \frac{\sum y}{n} = \frac{177}{6} = 29,5;$$

$$M_x = \frac{\sum x}{n} = \frac{490,3}{6} = 81,7;$$

$$M_z = \frac{\sum z}{n} = \frac{77,4}{6} = 12,9$$

и вычислить следующие величины:

$$\sum a_y^2 = \sum y^2 - nM_y^2 = 6971,5 - 6 \cdot 29,5^2 = 1750;$$

$$\sum a_y a_x = \sum xy - nM_y M_x = 17141,35 - 6 \cdot 29,5 \cdot 81,7 = 2680,45;$$

$$\sum a_y a_z = \sum yz - nM_y M_z = 2601,3 - 6 \cdot 29,5 \cdot 12,9 = 318.$$

Найденные значения подставляются в формулу ошибки уравнения линейной множественной регрессии

$$m_{y \cdot xz} = \sqrt{\frac{\sum a_y^2 - (b \sum a_y a_x + c \sum a_y a_z)}{n - k}},$$

где b и c — значения параметров найденного уравнения множественной регрессии; n — число сопряженных значений признаков; k — число коэффициентов уравнения регрессии, включая свободный член.

В рассматриваемом примере

$$m_{y \cdot xz} = \sqrt{\frac{1750 - (0,5003 \cdot 2680,45 + 1,0474 \cdot 318)}{6 - 3}} = \pm 5,03.$$

Следовательно, прогнозируя длину соцветия *O.litoralis* по высоте особи и длине ее самого крупного листа, мы рискуем ошибиться в среднем на 5 мм.

Глава 6

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ИЗУЧЕНИЯ РОСТА И ОРГАНОГЕНЕЗА У РАСТЕНИЙ

6.1. КВАНТИРОВАННОСТЬ РОСТА И ОРГАНОГЕНЕЗА

Множество фактов прямо или косвенно свидетельствует о том, что процессы роста и органогенеза у растений носят дискретный характер. Рост органов или растения в целом, а также число развиваемых им одноименных органов или их частей не могут остановиться на любой произвольной величине. Соседние значения этих величин разделены некоторым интервалом, в пределах которого ни рост, ни число органов не могут обрести конечного значения.

Эта закономерность, имеющая, как оказалось впоследствии, всеобщий характер, впервые была подмечена в 1906 г. русским ботаником В. М. Арциховским. Серьезное исследование данной проблемы было выполнено д'Арси Томпсоном [Thompson, 1942], первое издание книги которого вышло в свет в 1917 г. Для обозначения низшей дискретной единицы роста Томпсон ввел термин «квант роста».

Независимо от него явление квантированности числа метамеров у древесных пород было открыто А. В. Гурским [1948]. Попытки обобщения данных о квантированности роста и органогенеза у растений [Шмидт, 1961, 1968] получили дальнейшее развитие в работах М. М. Магомедмирзаева [1968, 1970а, б, 1972, и др.].

Квантированность роста и органогенеза у растений нередко имеет замаскированный характер и обнаруживается только лишь в процессе статистической обработки данных. Свое внешнее проявление она находит в появлении многовершинных кривых размера или числа органов, а также в соблюдении определенных числовых соотношений между последними. Рассмотрим последовательно эти вопросы.

6.1.1. Многовершинные кривые и их биологическая интерпретация. Получение многовершинной (полимодальной) кривой неизбежно вызывает у исследователя вопрос о причине такой не-

обычной ее формы. К сожалению, в существующих пособиях по биометрии этому вопросу почти не уделяется никакого внимания. Обычно ограничиваются указанием на то, что многовершинная кривая состоит из нескольких наслаивающихся друг на друга одновершинных кривых, являясь следствием гетерогенности выборки, результатом включения в нее особей, принадлежащих к разным генеральным совокупностям.

Многовершинная кривая, действительно, может возникнуть в указанных условиях, например при смешении в выборке особей разных видов, а также экземпляров одного вида, взятых из разных частей его ареала или разных типов местообитаний. Именно так иногда удается уловить географическую или экологическую дифференциацию вида. Однако надо иметь в виду, что даже заведомо гетерогенный материал способен в некоторых случаях давать одновершинные и даже близкие к нормальным кривые за счет «удачного» совмещения рядов с образованием общей для них моды.

Гетерогенность выборки — возможная, но далеко не единственная причина возникновения многовершинных кривых. Другие причины этого явления могут быть как чисто техническими (устраняемыми), вызывающими ложную многовершинность, так и естественными, характеризующими настоящую многовершинность кривых, в основе которой лежат биологические особенности варьирования признака.

Техническими причинами полимодальности могут быть недостаточность объема выборки при наличии сильной изменчивости признака и нарушение правил группировки вариантов в классы.

Иное дело — настоящая многовершинность, обусловленная биологическими причинами. С ней мы сталкиваемся тогда, когда наличие нескольких мод кривой не может быть объяснено техническими погрешностями или гетерогенностью выборки. В таком случае не остается ничего иного, как искать причину явления в самом характере варьирования признака, чему и посвящены следующие разделы.

Вершины кривых нередко выражены столь четко, что сомнений в их реальности не возникает. Но бывает и так, что вершины невысоки или между ними видны менее выдающиеся пики. Можно ли считать их также модами или же такие зубцы кривой лежат в пределах вероятных отклонений?

Этот вопрос решается путем сравнения ординат вершин с соседними ординатами кривой. Соответствующий данному случаю критерий имеет вид

$$t = \frac{f_i - f_k}{\sqrt{\frac{f_i(n - f_i)}{n}}} > 3,$$

где f_i — испытываемая модальная ордината; f_k — соседняя с ней ордината кривой; n — объем выборки, а подкоренное выраже-

ние есть не что иное, как ошибка испытуемой ординаты. При $t > 3$ модальная ордината достоверно отличается от соседней (т. е. вершина по отношению к ней вполне реальна); уверенность в правильности такого вывода возрастает, если аналогичный результат дает сравнение модальной ординаты с обеими соседними ординатами кривой.

Возьмем пример. Выборка из 97 экз. ($n=97$) *Leucanthemum vulgare* Lam. дала следующий вариационный ряд числа краевых цветков в корзинках (x):

x :	12	13	14	15	16	17	18	19	20	21	22
f :	1	14	13	4	6	9	7	10	12	20	1

Как видно, полученная кривая имеет три вершины с ординатами $f_1=14$, $f_2=9$ и $f_3=20$, соответствующими числу краевых цветков 13, 17 и 21. Реальность вершин f_1 и f_3 сомнений не вызывает, так как по крайней мере по отношению к одной из соседних (равных единице) они заведомо достоверны. Проверим ординату $f_2=9$, слабо отличающуюся от соседних (6 и 7). По отношению к «соседу слева» имеем

$$t = \frac{9 - 6}{\sqrt{\frac{9(97 - 9)}{97}}} = 1,05,$$

по отношению к «соседу справа»

$$t = \frac{9 - 7}{\sqrt{\frac{9(97 - 9)}{97}}} = 0,70.$$

В обоих случаях $t < 3$, следовательно, мода $f_2=9$, соответствующая 17 краевым цветкам, недостоверна, и рассматриваемая кривая в действительности имеет не три, а только две реальные вершины, ординаты которых ($f_1=14$ и $f_3=20$) соответствуют значениям $x=13$ и $x=21$ — числам так называемого ряда Фибоначчи.

6.1.2. Многовершинные кривые числа органов и ряд Фибоначчи. В конце прошлого века первый немецкий ботаник-биометрик Ф. Людвиг обнаружил, что кривые числа краевых цветков в корзинках многих видов сем. Asteraceae часто оказываются многовершинными, причем основные максимумы (моды) этих кривых соответствуют числу цветков 3, 5, 8, 13, 21, 34 ..., образуя ряд Фибоначчи,¹ каждый последующий член которого получается суммированием двух предыдущих.

У некоторых видов сложноцветных число краевых цветков варьирует очень сильно, и хотя указанная закономерность

¹ Итальянский математик, известный также под именем Леонардо из Пизы (1170—1250 гг.).

нередко проявляется и в относительно небольших по своему объему выборках, Людвиг для большей убедительности производил огромное количество подсчетов. Так, в одной из своих работ он [Ludwig, 1897] приводит сводный вариационный ряд ($n = 18573$) числа краевых цветков у *L. vulgare*:

x :	7	8	9	10	11	12	13	14	15	16	17	18	
f :	2	19	25	47	79	177	521	451	531	547	602	711	
	19	20	21	22	23	24	25	26	27	28	29	30	31
	952	1728	3888	1971	1252	894	653	664	424	431	322	235	202
	32	33	34	35	36	37	38	39	40	41	42	43	
	209	329	357	193	66	29	16	17	15	0	5	3	

Основные максимумы этого ряда падают на число цветков 13, 21 и 34, т. е. соответствуют членам ряда Фибоначчи, причем главный из них приходится на $x = 21$: такое количество краевых цветков имеют 3888 (или 20,9%) растений в этой громадной выборке. Кроме основных максимумов видны менее выдающиеся пики кривой на 26, 28 и 39 цветках, проверка которых по формуле, приведенной в предыдущем разделе, убеждает в их достоверности. На первый взгляд кажется, что наличие этих второстепенных пиков противоречит рассматриваемой закономерности, ибо числа 26, 28 и 39 не входят в ряд Фибоначчи. Оказывается, однако, что, исходя именно из ряда Фибоначчи, теоретически можно показать неизбежность появления второстепенных пиков многовершинной кривой, соответствующих этим (и другим) числам.

Дело заключается в том, что ряд Фибоначчи обладает замечательным свойством: последовательность частных, образованных делением каждого его члена на последующий, т. е. производный от него ряд $\frac{1}{2}; \frac{2}{3}; \frac{3}{5}; \frac{5}{8}; \frac{8}{13}; \frac{13}{21}; \frac{21}{34} \dots$, в пределе стремится к величине 0,61803..., характеризующей в геометрии знаменитое «золотое сечение».

Если взять любой отрезок между соседними числами из ряда Фибоначчи и поделить его в пропорции «золотого сечения», то величина «дочерних» отрезков вновь будет соответствовать числам из этого ряда. Например, если взять отрезок между числами Фибоначчи 34 и 55 и поделить «золотым сечением» его длину (21) на две части, то большая часть составит $21 \times 0,61803 = 12,9786$, а меньшая — $21 \times 0,38197 = 8,0214$, что округленно дает числа того же ряда 13 и 8. Граница между этими отрезками по отношению к исходному ряду Фибоначчи даст значение $x = 34 + 13 = 47$ (если большой отрезок откладывать на оси x слева, а меньший справа) или $x = 34 + 8 = 42$ (если поступать наоборот). Числа 42 и 47 и будут абсциссами вершин

следующего порядка, характеризующими теоретическое положение второстепенных максимумов кривых (рис. 13). Указанную процедуру можно продолжить: разделив отрезок между числами 34 и 42 (равный 8) «золотым сечением», получим: $8 \times 0,61803 = 4,9442$ и $8 \times 0,38197 = 3,0558$, т. е. округленно 5 и 3 — цифры из ряда Фибоначчи, откуда следует, что пики следующего порядка должны соответствовать числу цветков $x = 34 + 5 = 39$ и $x = 34 + 3 = 37$.

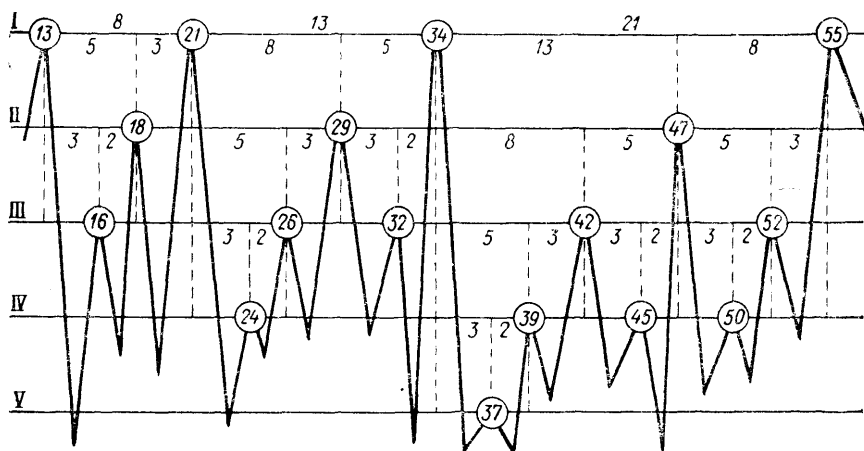


Рис. 13. Схема возникновения второстепенных максимумов в результате «золотого сечения» интервалов ряда Фибоначчи.

Римскими цифрами обозначены уровни, на которых возникают максимумы I—V порядков в интервале основных чисел ряда от 13 до 55.

Из схемы, приведенной на рис. 13, следует, что кроме основных пиков кривой, соответствующих членам ряда Фибоначчи, теоретически следует ожидать появления второстепенных максимумов на числах 16, 18, 23 (или 24), 26, 29, 31 (или 32), 37, 39, 42, 44 (или 45), 47, 50, 52 и т. д.

Людвигом было также указано, что главная (наиболее высокая) вершина кривых числа краевых цветков у разных видов сложноцветных падает на разные числа ряда Фибоначчи (это справедливо и для тех случаев, когда варьирование признака невелико и кривая имеет лишь одну вершину). Соответствующие числа выглядят, например, следующим образом: для *Senecio fuchsii* C. C. Gmel. — 5; для *Achillea ptarmica* L., *Anthemis arvensis* L. и *Centaurea cyanus* L. — 8; для *Anthemis cotula* L., *Senecio viscosus* L., *S. jacobea* L. и *Chrysanthemum segetum* L. — 13; для *Anthemis tinctoria* L., *Leucanthemum vulgare* Lam. и *Tripleurospermum inodorum* (L.) Sch. Bip. — 21; для *Bellis perennis* L. — 34.

Последователи Людвиг возвели обнаруженное им явление в ранг закона, распространяющегося на варьирование других органов растений. Многовершинность вариационных кривых была даже объявлена отличительной чертой изменчивости признаков растений по сравнению с животными.

Между тем вскоре выяснилось, что абсолютизировать «закон Людвиг» нет оснований: не только варьирование других органов, но даже изменчивость числа краевых цветков в соцветиях сложноцветных не всегда характеризуется многовершинными кривыми с модами, соответствующими числам ряда Фибоначчи. Особенно убедительно это было показано работами П. Фоглера [Vogler, 1910] и В. Н. Сукачева [1918], после которых интерес к проблеме резко снизился.

Однако если нельзя согласиться с абсолютизацией «закон Людвиг», то нельзя и сбрасывать со счета те факты, которые свидетельствуют о существовании особого типа варьирования признаков растений, находящего свое внешнее выражение в подчинении ряду Фибоначчи. Например, недавно Н. Г. Рытовой [1971] было показано, что число боковых побегов и общее число побегов всех возрастов и порядков у *Festuca rubra* L. образуют ряды Фибоначчи.

За последние 20—30 лет интерес к обсуждаемой проблеме снова возрос в связи с дальнейшей разработкой теории филлотаксиса (симметрии листорасположения), начало которой было положено еще работами Бонне и Шимпера, но особенно развито д'Арси Томпсоном [Thompson, 1942]. Специальное рассмотрение этой теории не входит в нашу задачу, но мы должны остановиться на том ее аспекте, который связан с проявлением в листорасположении ряда Фибоначчи.

Существуют два основных способа характеристики спирального листорасположения. Первый из них основан на допущении того, что листья побега располагаются строго один над другим, образуя вертикальные ряды — ортостихи. Мысленная спираль, последовательно соединяющая листья, называется *генетической* или *основной* спиралью и делится ортостихами на ряд листовых циклов. Угол расхождения между листьями выражают в долях от 360° . Для определения такого угла и написания формулы листорасположения (в виде дроби) находят на побеге два листа, расположенных друг над другом на одной ортостихе и подсчитывают количество оборотов генетической спирали между этими листьями (числитель дроби) и число листьев в данном листовом цикле (знаменатель дроби).

В простейшем случае генетическая спираль делает один оборот между листьями, расположенными друг над другом, и включает в себя два листа, что можно выразить формулой $1/2$. Эта формула характеризует двурядное (настоящее очередное) расположение листьев под углом 180° друг от друга. В других случаях на пути одного оборота спирали встречаются три листа:

формула будет $1/3$, а угол между листьями 120° . Часто два оборота спирали включают в себя пять листьев (формула $2/5$, угол 144°). Наблюдается расположение листьев по формулам $3/8$ (угол 135°), реже — $5/13$ (угол примерно 138°). Далее эстафету принимает расположение чешуй в шишках хвойных, следующее формулам $8/21$, $13/34$, $21/55$ и т. д. Некоторые типы листорасположения схематически показаны на рис. 14.

Итак, ряд встречающихся у растений формул листорасположения имеет вид: $\frac{1}{2}, \frac{1}{3}, \frac{2}{5}, \frac{3}{8}, \frac{5}{13}, \frac{8}{21}, \frac{13}{34}, \frac{21}{55} \dots$, где числитель показывает количество оборотов основной спирали до ли-

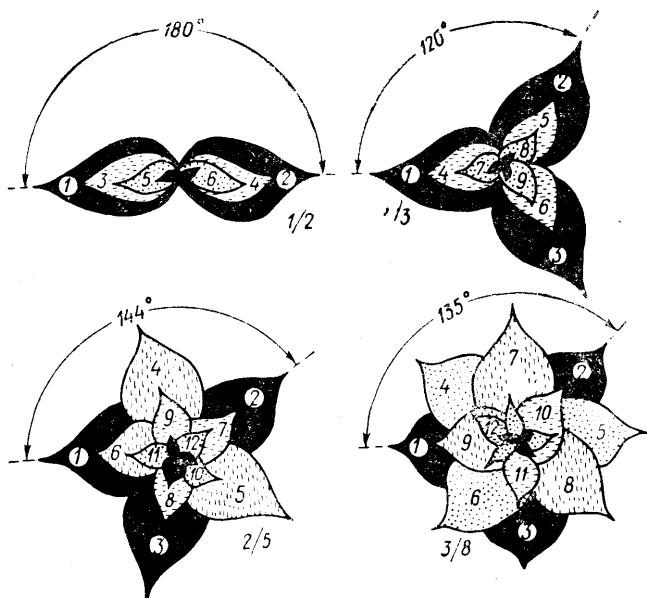


Рис. 14. Схематическое изображение некоторых типов листорасположения.

ста, лежащего выше исходного на одной ортостихе с ним, а знаменатель — порядковый номер листа, на котором заканчивается полный виток спирали.

Нетрудно видеть, что числители и знаменатели этого ряда, взятые в отдельности, составляют ряд Фибоначчи, который, следовательно, отражает внутренние закономерности образования листьев на побеге. Высшие члены этого ряда стремятся к величине $0,38197 \dots$, характеризующей меньший из двух отрезков «золотого сечения». Интересно, что при этом листья расположились бы под углом $137^\circ 30'$ друг к другу, образовав идеальную листовую мозаику: ни один лист не располагался бы точно над

другим. Заметим, наконец, что формула $1/3$ наиболее свойственна однодольным, а формула $2/5$ — двудольным, т. е. что знаменатели этих дробей символизируют наиболее характерный для каждого из классов покрытосеменных план строения цветка.

Сказанное достаточно хорошо описывает листорасположение у взрослых растений. Однако на ряде объектов было показано, что в ходе онтогенеза характер листовой спирали может изменяться, и формула листорасположения у взрослого растения не соответствует расположению листовых зачатков на поперечном срезе почки. Поэтому возник другой подход, предполагающий изучение филлотаксиса на ранних стадиях онтогенеза — морфологическое исследование строения почки или даже порядка заложения листовых примордиев на конусе нарастания.

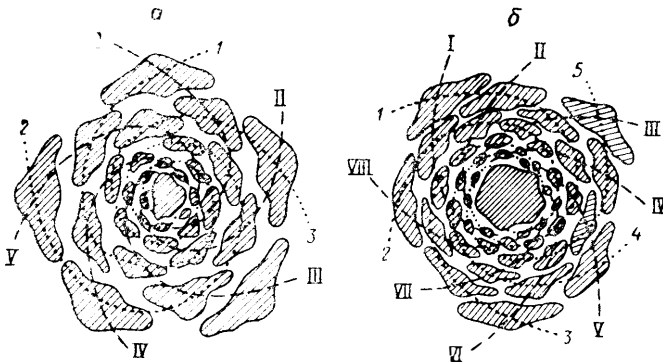


Рис. 15. Контактные парастихии на поперечном срезе верхушки побега (по: Эсау К., 1969).

Листорасположение: а — $1/3$; б — $5/8$. Арабскими и римскими цифрами обозначены левые и правые спирали.

Исходя из этого Л. Плантефоль [Plantefol, 1946, 1948] разработал другой вариант теории филлотаксиса, согласно которому генетической (основной) спирали реально не существует, но имеется несколько спиралей листовых зачатков, положение которых определяется деятельностью гипотетических «генеративных центров листообразования», расположенных в инициальном кольце апикальной меристемы. Оставляя в стороне это недоказанное предположение, не раз подвергавшееся критике, отметим, что при рассмотрении поперечного среза почки (а также корзинок сложноцветных или семенной шишки хвойных при наблюдении сверху) генетическую спираль, действительно, обнаружить чаще всего не удастся. Ни один листовой зачаток (семянка, чешуя) не располагается точно над другим, а угол расхождения между последовательными зачатками близок к «идеальному»

углу $137^{\circ}30'$. Зато всегда отчетливо вырисовываются (рис. 15) два типа спиралей — левые и правые, получившие название *парастихи* (или контактные парастихи).

Число парастих, идущих по часовой стрелке и против нее, у разных видов растений может быть различным. Изображая это соотношение в виде дроби, числитель которой показывает количество левых, а знаменатель — количество правых парастих, можно получить следующий ряд встречающихся в природе формул листорасположения: $\frac{2}{3}$, $\frac{3}{5}$, $\frac{5}{8}$, $\frac{8}{13}$, $\frac{13}{21}$, $\frac{21}{34}$, $\frac{34}{55}$... члены которого отличаются от членов предыдущего ряда. Однако числители и знаменатели нового ряда по-прежнему составляют ряд Фибоначчи, а предел, к которому стремятся высшие члены ряда, опять-таки определяется «золотым сечением» (только теперь величиной не меньшего, а большего его отрезка, а именно 0,61803...).

Несмотря на то, что характер листорасположения иногда не укладывается в приведенные схемы, мы можем констатировать существование у растений такого типа изменчивости числа и расположения органов, который математически описывается рядом Фибоначчи, содержащим алгоритм закономерно изменяющегося шага дискретности — кванта числа органов.

6.1.3. Многовершинные кривые размера органов. Многовершинность кривых проявляется не только у дискретно варьирующих признаков. Нередко такие кривые характеризуют изменчивость размерных признаков органов растений и рост самих растений. В. М. Арциховский [1906], изучив высоту растений *Capsella bursa-pastoris* (L.) Medic., *Artemisia absinthium* L., *Achillea millefolium* L. и *Stachys annua* L. в нескольких экологических вариантах, обнаружил, что как выборочные, так и суммарные для каждого вида кривые нередко оказывались многовершинными. Он пришел к выводу, что способность растений развить определенные размеры изменяется не плавно, а скачками, и что их рост может остановиться не на любых, а на определенных величинах.

Многовершинные кривые характеризуют изменчивость длины венчика и длины чашечки у видов рода *Odontites* (рис. 16 и 17). Удивительное совпадение максимумов кривых, лишаящее эти признаки таксономического значения, свидетельствует о проявлении квантированности роста. Следует добавить, что аналогичные кривые получаются вне зависимости от того, измеряются ли признаки в выборке из популяции или же берутся цветки с одного растения. На характер кривых не оказывают влияния ни географические, ни экологические факторы, что говорит о генетической обусловленности явления. Многовершинные кривые были получены также для длины венчика *Rhinanthus serotinus* (Schoench.) Oborny и для длины хвоинок у *Larix rossica* (Endl.) Sabine ex Trautv.

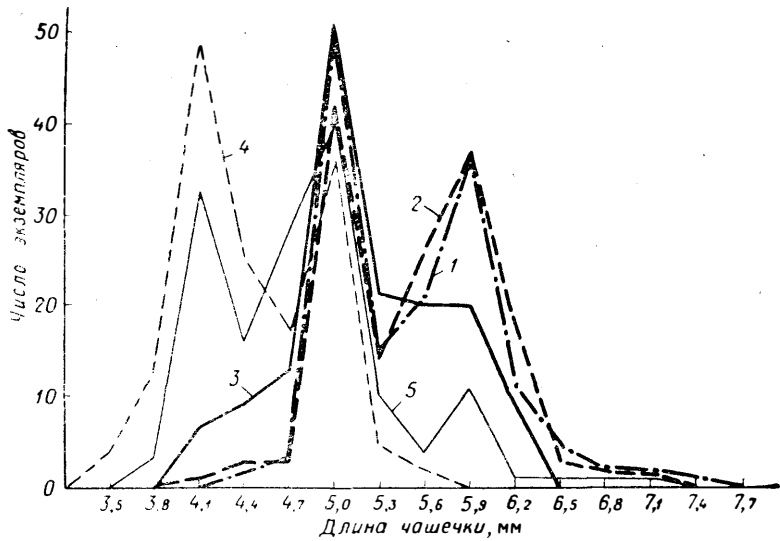


Рис. 16. Многовершинные кривые длины венчика у видов и экологических форм *Odontites*.
 1 — *O. litoralis* Fr., 2 — *O. verna* (Bell.) Dum., 3 — *O. serotina* Dum., 4 — береговая и 5 — сорно-полевая экологические формы последнего вида.

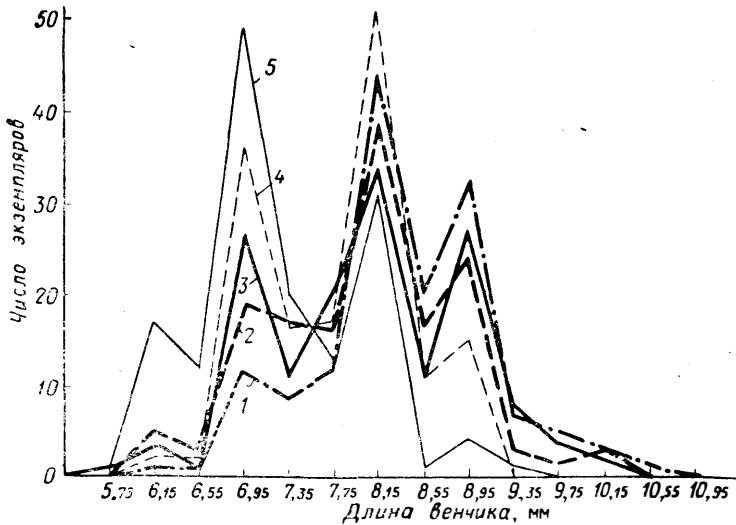


Рис. 17. Многовершинные кривые длины чашечки у видов и экологических форм *Odontites*.
 Обозначения те же, что на рис. 16.

Явление квантированности роста неоднократно наблюдалось М. М. Магомедмирзаевым [1968, 1970а, б, 1972, и др] при изучении длины хвоинок у *Pinus sosnowskyi* Nakai (рис. 18) и ряда линейных параметров листовой пластинки видов *Betula* и *Acer*, причем им также была подчеркнута независимость величины кванта роста от условий существования. Многовершинные кривые роста междоузлий у грецкого ореха отмечены И. Н. Коноваловым и др. [1960].

Совершенно очевидно, что во всех перечисленных случаях мы имеем дело с внешним проявлением определенных физиологических закономерностей роста. Среди физиологов одним из наиболее тонких исследователей ростовых процессов у растений

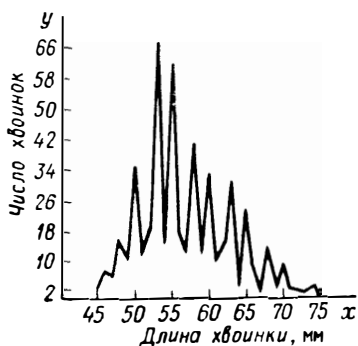


Рис. 18. Многовершинная кривая длины хвоинки у *Pinus sosnowskyi* Nakai (по: [Магомедмирзаев М. М., 1968]).

был Д. А. Сабинин [1940, 1957, 1963], пришедший к выводу о строгой ритмичности роста. Говоря о сезонных ритмах роста, Сабинин указывал, что они слагаются из «элементарных этапов развития», каковыми он считал образование листа с соответствующими ему узлом и междоузлием. При этом он подчеркивал, что «элементарный этап» отнюдь не является «элементарным актом» роста, ибо величина и периодичность последнего зависят от содержания нуклеиновых кислот в клетках растущей ткани. Понижение содержания этих кислот до определенного (порогового) уровня влечет за собой остановку роста на период, достаточный для пополнения убыли, причем период

покоя по времени продолжительнее периодов активного роста.

Сабинин подчеркивал, что рост должен носить прерывистый, периодический характер даже при неизменных и благоприятных для него внешних условиях. Подтверждение тому, что физиологической основой квантированности роста служит его эндогенная ритмичность, можно найти в работе В. Й. Кефели и др. [1964].

Близкие к этому мысли высказывались А. В. Гурским [1948], изучавшим закономерности роста побегов древесных растений. Отмечая ритмичность роста побегов, Гурский связывал ее с физиологией развития почек, порядком дифференциации меристемы точек роста. Им было показано, что число циклов роста удлинённых побегов в зависимости от внешних условий может изменяться, но лишь на целое число этих циклов (квантов).

В связи с квантированностью роста у растений получение

многовершинных кривых размерных признаков в выборке следует признать нормальным явлением. На практике такие кривые возникают сравнительно редко лишь потому, что кванты роста малы и чаще всего просто не улавливаются при той точности измерения признаков, которая обычно вполне устраивает исследователя. Когда же принятая точность измерений оказывается сопоставимой с величиной квантов роста, последние дают о себе знать, проявляясь в виде многовершинной кривой.

Измерить величину квантов роста можно путем построения вариационной кривой без группировки вариант в классы (рис. 19). При этом кривая распадается на множество мелких зубцов,

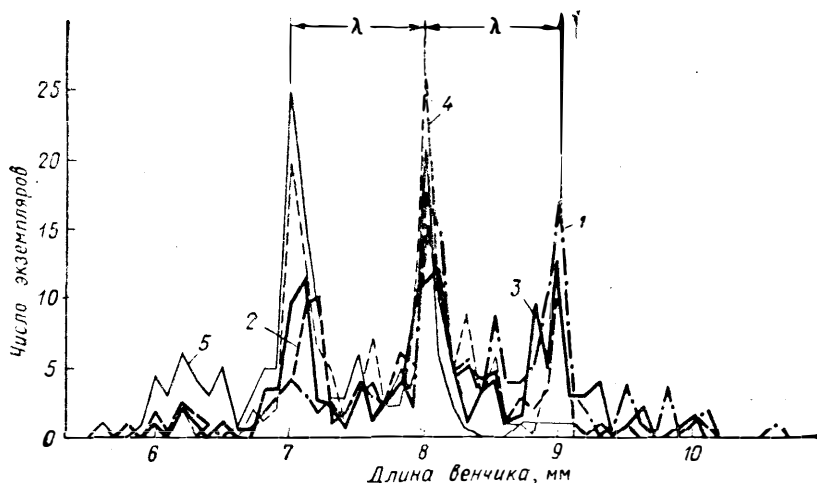


Рис. 19. Кривые длины венчика у видов и экологических форм *Odonites*, построенные без группировки вариант в классы. Обозначения те же, что на рис. 16; λ — величина кванта роста.

на фоне которых в виде пиков выделяются наиболее часто встречающиеся значения признака. Сами пики соответствуют периодам относительного покоя, а расстояние между ними, снятое в масштабе графика, приблизительно характеризует величину кванта роста (λ). Мелкие зубцы такой кривой отражают промежуточные значения признака, зафиксированные в момент активного роста органа и не могущие быть окончательными.

Изучение квантированности роста перспективно во многих аспектах теории морфогенеза, популяционной ботаники и систематики. На важную роль исследования квантов роста в развитии учения Н. И. Вавилова о гомологических рядах изменчивости указывала Е. Н. Синская [1963, с. 102]. По современным представлениям [Магомедмирзаев, 1970а, б, 1972, и др.], каждый линейный признак представляет собою «полимерный блок»,

слагающийся из ряда осуществившихся в ходе роста органа равных отрезков-квантов. Индивидуальные и видовые отличия по такому признаку могут быть обусловлены тремя причинами: неодинаковым размером квантов, неодинаковым их числом и, наконец, тем и другим одновременно. Например, онтогенетическая изменчивость формы листовой пластинки у видов рода *Acer* связана с неравным числом квантов роста жилок и угла их расхождения, который также квантирован.

Такой подход в будущем должен привести к возможности количественной оценки формы листьев. Согласно воззрениям д'Арси Томпсона [Thompson, 1942], форма органа есть результат деформации системы координат, в которую вписана исходная форма. Теперь к этому можно добавить, что такая деформация, осуществляющаяся в процессе роста, ограничена генетически заданными на осях отсчета размерами и количеством квантов, присущих данному виду. Именно это и определяет «физико-механические» возможности изменчивости формы, например, листовой пластинки, ограничивая ее определенными пределами.

Показано также, что квантированность роста *иерархична*: существуют мелкие и более крупные кванты, причем последние выступают как объединение целого числа первых. Крупные кванты, в свою очередь, объединяются вместе, что в конечном счете и определяет наличие сезонных ритмов роста.

Установлено также, что наряду с квантами роста в обычном (онтогенетическом) смысле этого термина, существует и квантированность различий между особями в среднем (т. е. на популяционном уровне). Соответствующая величина получила название *кванта дискриминации*. Кванты роста и кванты дискриминации либо равны, либо связаны кратными соотношениями. Судя по некоторым данным, популяционная изменчивость признаков тем выше, чем меньше величина кванта дискриминации.

6.1.4. «Принцип кратных отношений» А. В. Гурского и числовые соотношения органов растений. А. В. Гурским [1948] было подмечено, что количество листьев на ауксибластах древесных пород обычно в целое число раз превышает количество их на брахибластах. Рост ауксибласта подчинен, таким образом, целому числу квантов. Впоследствии выяснилось, что эта закономерность, которую Гурский назвал «принципом кратных отношений», в более общей форме справедлива для любых жизненных форм растений. Так, явление квантированности в структуре побегов злаков подробно проанализировано Т. И. Серебряковой [1961, 1969, 1971]. Сходные тенденции отмечены Н. В. Шиловой [1962, 1967] у ряда вечнозеленых кустарничков из сем. Ericaceae. Н. П. Соколова [1974] обнаружила квантированность строения побегов у представителей рода *Rubus*. У *R. arcticus* L. и *R. caesius* L. квант состоит из 4 метамеров, а у *R. canadensis* L. — из 8, причем число метамеров побега во-

обще не может быть меньше этих величин, а терминальный цветок на побеге возобновления закладывается лишь после развития кратного величине кванта количества метамеров.

В работе Н. П. Кренке [1940] приводятся данные об изменчивости длины междоузлий у сеянцев *Morus alba* L. Полученные им кривые имеют перегибы в одном случае на 2, 4, 6, 8, 10, 13 и 16-м междоузлиях, а в другом случае — на 10, 14, 18, 20, 24, 26, 28 и 30-м междоузлиях. Отчетливо видна тенденция наибольшего прироста четных междоузлий, т. е. квант в этих примерах составляет два междоузлия, а изменчивость признака описывается рядом $2+2n$. Такая закономерность обращает на себя внимание еще и тем, что для *M. alba* характерен тетрамерный план строения цветка и гинецей из двух плодолистиков.

В мире растений широко распространено интересное явление: изменчивость числа (и формы) разных органов или их частей нередко связана определенными кратными соотношениями. Приведем некоторые примеры, количество которых при желании можно увеличить во много раз.

Пентамерному или тетрамерному околоцветнику представителей сем. Fabaceae, Ericaceae, Vacciniaceae, Caryophyllaceae соответствует строго фиксированное вдвое большее число тычинок. У многих однодольных шести листочкам околоцветника, расположенным в два круга, соответствует такое же расположение шести тычинок, а в случае редукции андроеца (Iridaceae, Sauraceae, отчасти Juncaceae и Poaceae) число тычинок, как правило, сокращается до трех, т. е. теряется целый квант. С планом строения цветка кратными соотношениями могут быть связаны число развиваемых растением листьев (*Paris quadrifolia* L., *Maianthemum bifolium* (L.) F. W. Schmidt, *Trientalis europaea* L., виды рода *Galium* и др.) или форма листьев (*Butomus umbellatus* L., *Sagittaria sagittifolia* L., роды *Liriodendron*, *Berberis* и т. д.). У видов рода *Carex* с тремя тычинками мужского цветка и с нередко трехчленным гинецеем гармонирует трехгранный стебель.

Переход к зигоморфии в ходе эволюции ряда семейств сопровождался возникновением серий числовых соотношений органов и их частей, в основе варьирования которых оказалось квантовое число 2. Так обстоит дело у некоторых представителей сем. Ranunculaceae (роды *Delphinium*, *Aconitum* — два превращенных в нектарники лепестка венчика), Fumariaceae (становление андроеца из двух тычинок), Caprifoliaceae (роды *Lonicera*, *Linnaea* — двугубые венчики, 4 тычинки, расположение цветков по два, у последнего растения — два прицветника). Еще более отчетливо описываемая тенденция проявляется у представителей сем. Scrophulariaceae и особенно Lamiaceae — двугубые венчики, 4 тычинки, гинецей из двух плодолистиков, четверной плод, четырехгранный стебель, двурядное супротивное листорасположение.

Необходимо отметить, что рассматриваемая закономерность нередко маскируется изменчивостью признака и проявляется лишь при статистической обработке достаточно крупных выборок. Примером могут служить вариационные ряды числа лучей первого порядка в соцветиях некоторых видов сем. Ариасеae (табл. 40), образующие при большом числе наблюдений моды

Таблица 40. Число лучей первого порядка в соцветиях некоторых видов сем. Ариасеae

Вид	Место и год сбора	n	3	4	5	6
<i>Aegopodium podagraria</i> L.	Ленингр. обл.: Ст. Петергоф, 1961	519	—	—	—	—
<i>Anthriscus sylvestris</i> (L.) Hoffm.	Там же, 1965	1200	1	7	17	34
<i>Pimpinella saxifraga</i> L.	Сиверская, 1965	1600	—	—	—	—
<i>Chaerophyllum aromaticum</i> L.	Новгор. обл., г. Сол- цы, 1965	493	—	—	—	—

Продолжение табл. 40

7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
—	—	—	2	2	2	5	7	20	31	45	71	80	106	48	38	21	20
70	141	165	253	187	145	85	51	25	10	6	3	—	—	—	—	—	—
—	—	1	3	7	15	62	89	136	145	155	168	173	184	136	99	77	49
—	—	—	—	8	15	24	60	79	54	45	39	35	37	33	23	14	16

Продолжение табл. 40

25	26	27	28	29	30	31	32	33	M	σ	τ	<i>t</i>
12	5	3	1	—	—	—	—	—	19,37	2,78	14,4	0,20
—	—	—	—	—	—	—	—	—	10,22	2,31	22,6	0,21
37	24	16	12	5	4	1	1	1	18,75	3,56	19,0	0,12
6	2	2	1	—	—	—	—	—	17,30	3,40	19,7	0,16

на 10, 15 и 20 лучах с квантовым числом 5, соответствующим пентамерному плану строения цветка у зонтичных. Сильная растянутость вариационных рядов сопровождается, однако, не столь уж высокими значениями коэффициентов вариации (v), и такая общая характеристика изменчивости признака оказывается недостаточной. Следует обратить внимание на невысокую степень стабильности квантового числа 5: соседние с модаль-

ным значения признака также отличаются частой встречаемостью. Поэтому для более полной характеристики изменчивости полезно использовать еще один показатель — коэффициент концентрации ряда (l), представляющий собой процент модальных значений признака в ряду (при $l=100\%$ изменчивость признака равна нулю). Низкие значения этого коэффициента в рассматриваемом примере свидетельствуют о слабой генетической закреплённости квантового числа 5 в изменчивости числа лучей первого порядка в соцветиях зонтичных.

Анализируя строение цветка и обнаружив большое количество одноименных его частей (листочков околоцветника, тычинок, пестиков), ботаники нередко говорят и пишут, что их «много». Между тем не исключено, что в среднем число частей цветка кратно плану его строения. В этом отношении большой интерес представляет изучение изменчивости числа частей цветка у представителей Polycarpicae — группы, стоящей на низшей ступени эволюции покрытосеменных: выяснение соответствующих тенденций должно способствовать установлению филогенетических связей цветковых растений.

В качестве примера рассмотрим данные о проявлении квантированности в строении цветка у некоторых представителей сем. Ranunculaceae (известно, что в пределах этого семейства можно видеть широкий спектр признаков, получивших дальнейшее закрепление в ходе эволюции разных стелов как двудольных, так и однодольных покрытосеменных).

Anemone ranunculoides L., *Caltha palustris* L. В строении цветка этих видов легко найти черты примитивной организации (спиральное расположение, большое и неопределенное число тычинок и пестиков), однако эти черты свойственны подавляющему большинству лютиковых. Наличие простого околоцветника обычно считают проявлением близости к однодольным, но примитивные однодольные (например, Alismatales) нередко имеют двойной околоцветник. К тому же непреходимой грани между простым и двойным околоцветником не существует. Поэтому гораздо более существенным выглядит тот факт, что у обоих видов в основе варьирования числа листочков околоцветника лежит квантовое число 5 (табл. 41 и 42), сопровождающееся крайне высокими значениями коэффициента концентрации ряда (l), почти достигающего единицы в ряде популяций. Это позволяет отнести оба вида к той группе лютиковых, которая характеризуется преобладанием в строении цветка признаков «двудольности».

Hepatica nobilis Mill., *Anemone nemorosa* L., *Pulsatilla latifolia* Rupr. Изменчивость числа листочков околоцветника у этих видов (табл. 43, 44 и 45) свидетельствует о принадлежности их к группе лютиковых, «тяготеющих» к однодольным. Модальные ординаты во всех случаях, кроме числа листочков околоцветника у *Hepatica nobilis*, лежат на варианте 6 (кван-

Таблица 41. Число листочков околоцветника
у *Anemone ranunculoides* L.

Место сбора	n	4	5	6	7	8	9	10	11	M	σ	ν	l
Ленингр. обл.:													
ст. Мсжайская	203	—	118	73	11	1	—	—	—	5,18	0,62	11,3	0,58
Пушкин	98	1	94	3	—	—	—	—	—	5,02	0,20	4,0	0,96
Ст. Петергоф	200	—	181	15	3	1	—	—	—	5,12	0,41	8,0	0,91
ст. Страдное	200	—	170	21	7	2	—	—	—	5,21	0,54	10,4	0,85
ст. Толмачево	200	—	198	2	—	—	—	—	—	5,50	0,05	0,9	0,99
Арханг. обл.:													
пос. Комаринца	81	—	71	9	1	—	—	—	—	5,13	0,39	7,6	0,88
окр. Архангельска	130	—	129	—	1	—	—	—	—	5,01	0,18	3,6	0,99
Тат. АССР:													
пристань Моркваши	112	—	94	14	3	1	—	—	—	5,21	0,52	10,0	0,84
там же	205	1	168	21	12	3	—	—	—	5,26	0,64	12,2	0,82
ст. 774 км	210	—	168	35	3	2	1	1	—	5,27	0,66	12,5	0,80
Раифское лесничество	276	—	263	10	3	—	—	—	—	5,06	0,23	4,5	0,95
пос. Болгары	218	—	201	11	5	—	—	—	—	5,11	0,41	8,0	0,92
Окр. г. Кирсва	563	1	538	20	2	—	1	—	1	5,09	0,67	13,2	0,96

Примечание. Данные по Ст. Петергофу, ст. Страдное и ст. Толмачево заимствованы из работы В. В. Бочанцевой [1972], по Архангельской обл., Тат. АССР и г. Кирсва — из работы В. С. Чунова [1972].

Таблица 42. Число листочков околоцветника у *Caltha palustris* L.

Место сбора	n	3	4	5	6	7	8	9	10	M	σ	v	l
Голландия	416	—	—	300	87	25	4	—	—	5,36	0,64	11,9	0,72
Тюрингия	281	—	—	223	45	6	4	3	—	5,29	0,69	13,0	0,79
Петроград	582	—	8	493	69	9	—	3	—	5,16	0,51	9,9	0,85
Швеция (сумма) . . .	1148	—	21	1082	35	9	1	—	—	5,03	0,30	6,0	0,94
Новгор. обл. (сумма).	4975	—	33	4124	641	143	28	5	1	5,20	0,53	10,2	0,83
Урал (сумма)	8080	4	99	6989	865	87	17	9	—	5,13	0,42	8,2	0,87

Примечание. Данные для Голландии, Тюрингии, Петрограда (Охтинское лесничество) и Новгородской обл. (Княжий двор) приведены по В. И. Сукачеву [1913]; для Урала — по П. В. Терентьеву (1935).

Таблица 43. Число листочков околоцветника (А) и покрывала (Б) у *Hepatica nobilis* Mill. в популяциях из Ленинградской обл. [по: Бочанцева, 1972]

Место сбора	n	3	4	5	6	7	8	9	10	M	σ	v	l
А													
Ст. Можайская . . .	250	—	—	4	111	106	26	2	1	6,66	0,75	11,3	0,44
Там же	200	—	—	—	37	82	63	13	5	7,34	0,93	12,7	0,41
Ст. Толмачево	140	—	—	—	67	57	14	2	—	6,65	0,71	10,7	0,49
Ст. Отрадное	200	—	—	—	57	69	53	19	2	7,20	0,99	13,8	0,34
Там же	200	—	—	—	73	75	37	13	2	6,98	0,95	13,6	0,38
Б													
Ст. Можайская . . .	200	144	42	11	3	—	—	—	—	3,36	0,66	19,5	0,72
Ст. Толмачево	140	112	26	2	—	—	—	—	—	3,21	0,45	14,4	0,80
Ст. Отрадное	200	166	33	1	—	—	—	—	—	3,18	0,39	12,4	0,83
Там же	200	182	16	2	—	—	—	—	—	3,10	0,33	10,7	0,91

товое число 3) при высоких значениях коэффициента концентрации ряда (l). У последнего вида в некоторых популяциях максимум кривой приходится не на 6, а на 7 листочков околоцветника, а коэффициент концентрации ряда примерно вдвое ниже, чем у других видов. Зато по числу листочков покрывала, фактически образующих у этого вида чашечку, все популяции дают односкатные кривые с резким максимумом ($l > 0,7$) на варианте 3. Заметим также, что квантовое число 3 гармонирует у рассматриваемых видов с трехлопастной формой листа (или покрывала).

Таблица 44. Число листочков околоцветника у *Анемоне nemorosa* L.

Место сбора	n	4	5	6	7	8	9	10	11	12	M	σ	σ^2	t
Англия, Бакхэм (сумма)	5000	4	94	2635	1758	403	86	13	5	2	6,57	0,78	12,0	0,53
Зап. Украина, Львов	3126	—	19	2282	694	122	6	3	—	—	6,30	0,57	9,0	0,73
Петербургская губ. (сумма)	3000	—	21	1968	868	123	18	2	—	—	6,38	0,61	9,6	0,66
Детское Село (сумма)	8738	—	20	5627	2069	743	199	58	21	1	6,51	0,83	12,8	0,64
Ленингр. обл.:														
ст. Тайцы — ст. Пудость	200	—	—	158	32	10	—	—	—	—	6,26	0,54	8,6	0,79
ст. Лемболово	237	—	4	131	79	16	5	2	—	—	6,55	0,80	12,2	0,55
Ст. Петергоф	200	—	1	136	46	16	1	—	—	—	6,40	0,66	10,4	0,68
там же	200	—	—	175	24	1	—	—	—	—	6,13	0,35	5,7	0,88
ст. Можайская	200	—	—	142	39	17	2	—	—	—	6,40	0,68	10,7	0,71
ст. Толмачево	200	—	1	164	33	2	—	—	—	—	6,33	0,42	6,7	0,82
ст. Сосново — ст. Петяярви	200	—	8	123	51	13	—	1	1	3	6,50	1,50	23,0	0,62
ст. Отрадное	200	—	1	150	44	4	—	1	—	—	6,47	0,56	8,6	0,75
112-й км шоссе на Псков	108	—	—	72	33	3	—	—	—	—	6,36	0,54	8,5	0,67
ст. Васкелово	120	—	1	66	31	7	13	2	—	—	6,77	1,00	14,8	0,55
там же	129	—	—	62	34	11	18	2	2	—	6,99	1,20	17,2	0,48
Новгородская обл.:														
г. Холм	147	—	3	133	10	1	—	—	—	—	6,06	0,33	5,4	0,90
там же	140	—	1	102	22	6	6	3	—	—	6,45	0,99	15,3	0,73
Псковская обл.:														
г. Невель	128	—	3	112	12	1	—	—	—	—	6,09	0,38	6,2	0,88
пос. Лобок	175	—	1	126	43	4	1	—	—	—	6,30	0,55	8,7	0,72
Львовская обл.:														
г. Борислав	199	—	—	117	76	6	—	—	—	—	6,44	0,55	8,5	0,59
там же	152	—	—	95	49	8	—	—	—	—	6,43	0,59	9,2	0,63

Примечание. Таблица составлена на основании опубликованных данных Р. Э. Регеля, Е. П. Баратынской-Сорокиной, В. В. Бочанцевой, В. С. Чулова, а также подсчетов автора по собранным для него А. И. Толмачевым и П. М. Добряковым материалам из Ленинградской обл.

Trollius europaeus L., *Ficaria verna* Huds. С характеристикой изменчивости строения околоцветника у этих видов дело обстоит сложнее. У *Trollius europaeus* чашелистики располагаются кругами по 3, причем внутренний круг может быть неполным, а само число кругов — различным. Встречаются переходные между чашелистиками и нектарниками образования, стирающие грань между простым и двойным околоцветником. Эти признаки сближают данный вид с группой лютиковых, тяготеющей к однодольным. С другой стороны, множественный характер околоцветника и сильная изменчивость его частей свидетельствуют в пользу примитивности вида. В табл. 46 и 47 приведены данные о варьировании числа чашелистиков и нектарников в географически удаленных друг от друга популяциях *T. europaeus*. Как видно, оба признака варьируют сильно и имеют колеблющиеся моды (10—13 для числа чашелистиков и 10—11 для числа нектарников). Отсутствие строго фиксированной моды и низкий коэффициент концентрации ряда (обычно $l < 0,3$) мы впра-

Таблица 45. Число листочков околоцветника у *Pulsatilla latifolia* Rupr. в окрестностях г. Казани [по: Гордягин, 1920]

Год сбора	n	2	3	4	5	6	7	8	9	10	11	12	M	σ	σ	l
1906	1000	—	—	7	55	920	17	1	—	—	—	—	5,95	0,32	5,4	0,92
1907	1000	1	1	14	50	844	66	19	3	1	0,7	0,3	6,04	0,59	9,8	0,84
1908	1000	—	—	8	59	916	22	2	2	—	—	—	5,97	0,36	6,0	0,92

Примечание. Дробные частоты за 1907 г. обусловлены тем, что все ряды для сравнимости даны в пересчете на 1000 экз.

рассматривать как свидетельства примитивного характера варьирования признаков. Интересно, что сильная изменчивость свойственна здесь не только самим признакам, но и величине корреляционной связи между ними (табл. 47), которая (при общем довольно низком фоне корреляции) колеблется в популяциях от существенной до недостоверной. Это обстоятельство также следует рассматривать в качестве примитивной черты организации околоцветника *T. europaeus*.

С другой стороны, тенденция к соблюдению квантового числа 5 (или 2?), сказывающаяся в частом положении моды обоих признаков на варианте 10 (табл. 46 и 47), служит указанием на известную степень прогрессивности и «двудольности». Остается заключить, что по строению околоцветника *T. europaeus* представляет собой достаточно противоречивый объект и, объединяя в себе признаки примитивности, «однодольности» и «двудольности», должен быть признан промежуточным между всеми этими группами.

Таблица 46. Число чашелистиков у

Место сбора	n	6	7	8	9	10	11	12	13
Средняя Германия (Тюрингия)	340	1	2	14	25	85	78	76	48
Ленингр. обл.:									
Ст. Петергоф	220	—	—	1	7	36	47	60	48
там же	220	—	—	4	11	72	53	37	28
107 км шоссе на Лодейное Поле	178	—	2	11	27	33	38	44	17
89 км "	270	—	—	9	38	73	69	48	28
ст. Тайцы — ст. Пудость	100	—	—	—	2	26	29	23	17
Окр. Сыктывкара:									
с. Вильгорт	96	—	—	—	—	17	23	30	17
32 км шоссе на Княжпогост	105	—	—	1	5	30	29	21	16
Заполярье:									
г. Воркута	88	—	—	—	1	8	13	22	17
Югорский п-ов, р. Люби-Яха	75	—	—	—	1	2	6	9	15

Примечание. Ряд для Тюрингии приводится по: [Ludwig, 1898]. Остальные данные П. М. Добрякова, С. А. Токаревских и автора в 1962—1965 гг.

Таблица 47. Число нектарников у *Trollius europaeus* L. и кор

Место сбора	n	3	4	5	6	7	8	9	10	11
Ленингр. обл.:										
Ст. Петергоф	220	—	—	—	5	—	28	49	53	38
там же	220	1	4	5	9	12	36	41	63	32
107 км шоссе на Лодейное Поле	178	—	2	7	13	19	28	34	36	20
89 км " " "	270	—	2	6	13	26	48	53	53	42
ст. Тайцы — ст. Пудость	100	—	—	1	1	6	14	17	23	19
Окр. Сыктывкара:										
с. Вильгорт	96	—	—	—	—	—	8	10	17	27
32 км шоссе на Княжпогост	105	—	—	—	—	2	7	13	22	26
Заполярье:										
г. Воркута	88	—	—	—	—	—	2	3	6	16
Югорский п-ов, р. Люби-Яха	75	—	—	1	5	4	8	9	8	12

Trollius europaeus L.

14	15	16	17	18	19	20	21	M	σ	ν	l
6	3	—	2	—	—	—	—	11,07	1,57	14,2	0,25
9	7	4	1	—	—	—	—	11,84	1,51	12,7	0,27
8	5	1	1	—	—	—	—	11,19	1,52	13,6	0,33
3	1	1	1	—	—	—	—	10,85	1,65	15,2	0,25
4	4	1	—	—	—	—	—	10,79	1,37	12,7	0,27
3	—	—	—	—	—	—	—	11,36	1,18	10,4	0,29
7	2	—	—	—	—	—	—	11,79	1,25	10,6	0,31
3	—	—	—	—	—	—	—	11,18	1,27	11,4	0,29
12	9	1	—	4	1	—	—	12,80	2,00	15,6	0,25
12	8	7	4	4	3	3	1	14,35	2,64	18,4	0,20

здесь, а также в табл. 47, получены в результате обработки сборов А. И. Толмачева

реляционные отношения η Числа чашелистиков и нектарников

12	13	14	15	16	17	18	19	20	21	22	M	σ	ν	l	$\eta \pm m_\eta$
21	5	2	1	—	—	3	—	—	—	—	9,83	1,89	19,2	0,24	$0,172 \pm 0,065$
11	3	1	1	—	1	—	—	—	—	—	9,20	2,04	22,2	0,29	$0,418 \pm 0,055$
14	2	—	1	1	—	1	—	—	—	—	9,00	2,18	24,2	0,20	$0,135 \pm 0,073$
22	3	1	1	—	—	—	—	—	—	—	9,20	1,92	20,9	0,20	$0,281 \pm 0,055$
14	3	2	—	—	—	—	—	—	—	—	9,78	1,61	16,8	0,23	$0,303 \pm 0,091$
18	9	5	1	—	1	—	—	—	—	—	11,00	1,71	15,6	0,28	$0,185 \pm 0,099$
15	6	4	4	2	1	1	1	1	—	—	11,17	2,37	21,2	0,25	$0,248 \pm 0,092$
5	9	11	12	5	5	2	3	4	3	2	14,12	3,40	24,1	0,18	$0,344 \pm 0,094$
9	4	9	3	1	1	0	1	—	—	—	10,68	2,93	27,4	0,16	$0,420 \pm 0,095$

Тот же вывод, но на основании несколько иных фактов, приходится сделать и в отношении *Ficaria verna* (табл. 48). По числу лепестков венчика этот вид тяготеет к группе «двудольных» лютиковых, ибо модальная ордината неизменно соответствует варианту 8 (квантовое число 2). Однако цветки чистяка обычно снабжены тремя (очень редко четырьмя) чашелистиками. Уэлдон [Weldon, 1901] показал, что модальные ординаты числа пестиков и тычинок соответствуют вариантам 15 и 27 (квантовое число 3), причем в ходе цветения модальное значение числа тычинок у поздноцветущих экземпляров снижается до 18, т. е. теряются три кванта. Поэтому, считая положение *F. verna* промежуточным между группами лютиковых, «тяготеющими»

Таблица 48. Число лепестков венчика у *Ficaria verna* Huds. в популяциях из Ленинградской обл.

Место сбора	n	7	8	9	10	11	M	σ	σ^2	t
Пушкин	310	10	142	117	32	9	8,64	0,82	9,5	0,46
Там же	300	12	181	79	22	6	8,43	0,77	9,1	0,60
Там же	189	4	88	67	29	1	8,66	0,78	9,0	0,47
Ботанический сад ЛГУ . . .	300	15	128	97	48	12	8,71	0,93	10,7	0,43

к двудольным и однодольным, надо отметить, что это положение обусловлено не «блуждающими» модами (как у *Trollius europaeus*), а разными квантовыми числами (2 и 3), определяющими количество разных частей цветка.

Общеизвестно, что в процессе эволюции покрытосеменных число одноименных частей цветка подвергалось сокращению и стабилизации. В связи с этим представляется перспективным изучение числовых соотношений органов и их частей у растений в экологическом и географическом аспектах.

Своеобразный «взрыв изменчивости» числа листочков околоцветника наблюдался у *Pulsatilla latifolia* под влиянием необычайной засухи 1907 г. (см. табл. 45). Сезонный сдвиг моды числа краевых цветков у сложноцветных отмечался рядом авторов. Количество листочков околоцветника *Anemone nemorosa* зависит от фенотазы, в которой находится популяция [Losch, 1916].

Как видно из таблиц 46 и 47, число чашелистиков и нектарников у *Trollius europaeus* обнаруживает географическую изменчивость, увеличиваясь с юго-запада на северо-восток: северные популяции характеризуются достоверным сдвигом средних арифметических значений признаков в большую сторону. Напротив, число листочков околоцветника у *Anemone nemorosa*

и *A. ranunculoides*, по данным В. С. Чупова [1972], географической изменчивости не проявляет.

Интересные данные получены В. В. Бочанцевой [1972], изучившей изменчивость околоцветника *Anemone nemorosa*, *A. ranunculoides* и *Hepatica nobilis* в популяциях из Ленинградской обл. Оказалось, что наиболее сильная изменчивость этого признака свойственна самым древним популяциям, пережившим суровые условия последних стадий валдайского оледенения. Растения молодых популяций, занимающих территории, которые стали заселяться растительностью только в послеледниковое время, обладают, наоборот, более стабильным числом листочков околоцветника.

Таким образом, явление квантированности роста и органогенеза у растений реально имеет место в природе и находит свое выражение в трех основных формах: 1) в виде многовершинных кривых числа органов или их частей с максимумами, подчиняющимися ряду Фибоначчи, ряду $2+2n$ и, возможно, некоторым другим рядам; 2) в виде многовершинных кривых размеров растений, их органов или частей с расстояниями между пиками, соответствующими определенному интервалу роста, и 3) в виде соблюдения «принципа кратных отношений».

В связи с этим на онтогенетическом уровне можно выделить две категории квантов: кванты числа (органогенеза) и кванты роста (размера) органов растений. Кванты числа соответствуют первой и третьей формам проявления квантированности, а кванты роста — второй ее форме. На популяционном уровне существуют кванты дискриминации, кратные квантам онтогенетического уровня, но характеризующие дискретность различий признаков у особей.

6.2. МАТЕМАТИЧЕСКИЙ АНАЛИЗ РОСТА

Явление квантированности роста и органогенеза, рассмотренное в предыдущем разделе, представляет собой интересную, но частную особенность ростовых процессов у растений. Нашей дальнейшей задачей будет математическое описание роста в целом — от его начала до окончания. Оставляя в стороне многообразие аспектов роста, сосредоточим внимание на феноменологической стороне явления, выражающейся в необратимом увеличении размера тела или органов растения в результате самовоспроизводства живого вещества в ходе онтогенеза. Как справедливо подчеркивает Э. Синнот [1963], изучение «геометрического» результата роста составляет одну из важнейших задач динамической морфологии (учения о морфогенезе) растений.

Рост растений и животных в подавляющем большинстве случаев подчиняется ходу S-образных (сигмоидных) кривых: вна-

чале рост идет медленно, потом ускоряется, достигает максимальной скорости, а затем начинает замедляться и, наконец, останавливается. Аналогичные кривые получаются при изучении роста популяций организмов в замкнутой среде или при освоении ими новых экотопов. Во всех этих, казалось бы достаточно различных, ситуациях мы сталкиваемся с проявлением наиболее общих закономерностей ростовых процессов, позволяющих использовать одинаковый математический аппарат для их описания и изучения.

Тем не менее известно несколько попыток создания теорий роста животных и растений, предусматривающих использование разных математических моделей ростового процесса. Обзор этих теорий можно найти в ряде работ [Thompson, 1942; Medawar, 1945; Сабинин, 1963; Стрелер, 1964; Зотин, 1974, и др.], поэтому мы остановимся лишь на некоторых положениях и приемах.

Как в зоологии, так и в ботанике неоднократно предлагалось описывать рост параболическими или экспоненциальными функциями. Одно из предложенных уравнений [Blackman, 1919] имеет, например, вид формулы сложных процентов: $P_1 = P_0 \cdot e^{rt}$, где P_1 — размер органа в момент времени t ; P_0 — размер органа в начале роста; e — основание натуральных логарифмов (2,718); r — скорость роста, выраженная в процентах от размера за каждую предыдущую единицу времени.

Аналогичная формула для описания роста клеток корня предложена недавно В. Б. Ивановым [1974]: $l = l_0 \cdot e^{kt}$, где l — средняя длина клетки в момент времени t ; l_0 — средняя длина клетки в начале изучаемого участка корня; k — константа, справедливая на том участке меристемы, где клетки метаксилемы корня не делятся.

Как уже говорилось, эмпирические кривые роста имеют S-образную форму. Это означает, что параболические и экспоненциальные функции заведомо не могут описать весь ростовой процесс в целом: они удовлетворительно интерполируют лишь отдельные участки кривой.

В зоологии пользуется признанием теория роста Пюттера — Берталанфи, в основу которой положены два постулата: 1) скорость роста пропорциональна разности между ассимиляцией и диссимиляцией и 2) скорость синтеза пропорциональна поверхности тела, а скорость распада — его объему [Pütter, 1920; Bertalanffy, 1934, 1938, и др.].

Теория Пюттера — Берталанфи исходит из предположений, которые не могут быть приняты в ботанике. В силу высокой степени расчлененности тела соотношение между его поверхностью и объемом у растений совершенно иное, чем у животных. Локализация меристем и рост посредством растяжения, которого нет у животных, не учитывается этой теорией. Нако-

Таблица 49. Выравнивание эмпирического ряда регрессии роста листа *Nicotiana rustica* L. в длину (y) по дням наблюдений (x) уравнением логистической функции (y_T)

Эмпирические данные ($A=263; n=12$)							Расчет теоретических значений					Производные			
x	x^2	y	$\frac{A}{y}$	$\frac{A}{y}-1$	$\lg\left(\frac{A}{y}-1\right)=z$	xz	bx	$a+bx=\lg\left(\frac{A}{y_T}-1\right)$	$\frac{A}{y_T}-1$	$\frac{A}{y_T}$	y_T	y'	y''		
0	—	0	—	—	—	—	0	1,6523	44,90	45,90	5,7	1,7	0,5		
6	36	32	8,2188	7,2188	0,8584	5,1504	-0,8106	0,8417	6,945	7,945	33,1	9,0	2,1		
8	64	50	5,2600	4,2600	0,6294	5,0352	-1,0808	0,5715	3,728	4,728	55,6	13,6	2,4		
9	81	70	3,7571	2,7571	0,4404	3,9636	-1,2159	0,4364	2,732	3,732	70,5	16,0	2,3		
10	100	85	3,0941	2,0941	0,3209	3,2090	-1,3510	0,3013	2,001	3,001	87,6	18,2	1,9		
11	121	105	2,5048	1,5048	0,1775	1,9525	-1,4861	0,1662	1,467	2,467	106,7	19,7	1,2		
12	144	132	1,9924	0,9924	1,9967	-0,0033	-0,0396	-1,6212	0,0311	1,074	2,074	126,8	20,4	0,2	
15	225	185	1,4216	0,4216	1,6249	-0,3751	-5,6265	-2,0265	-0,3742	1,6258	0,4225	1,4225	184,9	17,1	-2,2
17	289	216	1,2176	0,2176	1,3377	-0,6623	-11,2591	-2,2967	-0,6444	1,3556	0,2268	1,2268	214,4	12,3	-2,4
19	361	237	1,1097	0,1097	1,0403	-0,9597	-18,2343	-2,5669	-0,9146	1,0854	0,1217	1,1217	234,5	7,9	-1,9
20	400	242	1,0868	0,0868	2,9385	-1,0615	-21,2300	-2,7020	-1,0497	2,9503	0,0892	1,0892	241,5	6,1	-1,6
22	484	255	1,0314	0,0314	2,4969	-1,5031	-33,0682	-2,9722	-1,3199	2,6801	0,0479	1,0479	251,0	3,6	-1,0
27	729	259	1,0154	0,0154	2,1875	-1,8125	-48,9375	-3,6477	-1,9954	2,0046	0,0101	1,0101	260,4	0,8	-0,2
30	—	263	—	—	—	—	-4,0530	-2,4007	3,5993	0,0040	1,0040	262,0	0,3	-0,1	
176	3034	—	—	—	-3,9509	-119,0845	—	—	—	—	—	—	—	—	

нец, формула Берталанфи $x = a(1 - be^{-kt})$, где x — длина тела; a , b , k — параметры; t — число дней роста, описывает рост ничуть не лучше тех формул, о которых пойдет речь в дальнейшем и которые без каких-либо гипотетических допущений просто соответствуют феноменологической сути ростового процесса, подчиняющегося ходу S-образных кривых.

6.2.1. Описание роста логистической функцией и функцией Гомпертца. Логистическая функция, предложенная в 1838 г. бельгийским математиком П. Ферхюльстом, в первые десятилетия нашего века широко использовалась статистиками для

Таблица 50. Выравнивание эмпирического ряда регрессии роста по дням наблюдений (x) урав

Эмпирические данные ($A=242,5$; $n=7$)							
x	x^2	y	$\frac{A}{y}$	$\lg \frac{A}{y}$	$\lg \lg \frac{A}{y} = z$	xz	
13	---	0	--	—	—	—	
18	324	4,0	60,6250	1,7823	0,2511	4,5198	
22	484	8,0	30,3125	1,4815	0,1707	3,7554	
26	676	20,0	12,1250	1,0837	0,0350	0,9100	
39	1521	128,0	1,8945	0,2775	1,4433	-0,5567	-21,7113
53	2809	183,5	1,3215	0,1210	1,0828	-0,9172	-48,6116
67	4489	227,5	1,0659	0,0278	2,4440	-1,5560	-104,2520
86	7396	238,5	1,0168	0,0073	3,8533	-2,1367	-183,7562
99	—	242,5	—	—	—	—	
311	17699	—	—	—	-4,7098	-349,1459	

прогнозирования роста народонаселения. Впоследствии оказалось, что она хорошо описывает ростовые процессы у растений. Существует несколько способов записи соответствующей формулы, из которых наиболее удобен следующий: $y = \frac{A}{1 + 10^{a+bx}} + c$, где A — расстояние между нижней и верхней асимптотами (для наших целей — окончательный размер растения или органа); x — время, прошедшее с начала роста; a и b — константы, определяющие наклон, изгиб и точку перегиба кривой; c — исходный размер (при описании роста от его начала $c=0$).

График логистической функции представляет собой симметричную S-образную кривую, поэтому данной функцией следует пользоваться тогда, когда эмпирические точки, характеризующие рост органа во времени, располагаются на графике симмет-

ричным образом. Если же эмпирические данные указывают на асимметрию кривой за счет растянутости верхней ветви (затухание роста происходит медленно), то вместо логистической функции следует использовать функцию, предложенную Б. Гомпертцом в 1825 г. и после некоторых преобразований имеющую вид

$$y = \frac{A}{10^{10^{a+bx}}},$$

где обозначения те же, что и для логистической функции.

листа *Bryophyllum daigremontianum* Berger в длину (y) нением функции Гомпертца (y_T)

Расчет теоретических значений					Производные	
b.x	a+b.x → lg lg $\frac{A}{y_T}$	lg $\frac{A}{y_T}$	$\frac{A}{y_T}$	y _T	y'	y''
-0.4680	0.4583	2.875	749.9	0.3	0.2	0.1
-0.6480	0.2786	1.900	79.43	3.1	1.1	0.3
-0.7920	0.1346	1.363	23.07	10.5	2.7	0.5
-0.9360	-0.0094 1.9906	0.9786	9.519	25.5	4.8	0.5
-1.4040	-0.4774 1.5226	0.3332	2.154	112.6	7.2	-0.1
-1.9080	-0.9814 1.0183	0.1043	1.272	190.6	3.8	-0.2
-2.4120	-1.4854 2.5146	0.0327	1.078	225.0	1.4	-0.1
-3.0960	-2.1694 3.8306	0.0068	1.016	238.7	0.3	-0.02
-3.5640	-2.6374 3.3626	0.0023	1.006	241.1	0.1	-0.01
-	-	-	-	-	-	-

Задача построения по эмпирическим данным теоретической логистической или гомпертцовой кривой представляет собой частный случай регрессионного анализа, изложенного в главе 5. Рассмотрим конкретные примеры.

Изучался рост 8-го листа *Nicotiana rustica* L. и 7-го листа *Bryophyllum daigremontianum* Berger, выращиваемых на гидропонике в константных условиях среды. Эмпирическая линия регрессии для первого вида оказалась почти симметричной, что позволило использовать логистическую функцию (табл. 49). У второго вида верхняя ветвь эмпирической кривой оказалась заметной растянута, и для моделирования роста в этом случае была применена функция Гомпертца (табл. 50). Для удобства сравнения проследим порядок работы по вычислению теоретических кривых параллельно на обоих объектах.

Описание роста листа
Nicotiana rustica
(логистическая функция).

Описание роста листа
Bryophyllum daigremontianum
(функция Гомпертца)

1. Исходная интерполяционная формула:

$$y = \frac{A}{1 + 10^{a+bx}}; \quad y = \frac{A}{10^{10^{a+bx}}}.$$

2. Путем некоторых преобразований и логарифмирования приводим исходные формулы к более удобному для дальнейшей работы виду:

$$\lg\left(\frac{A}{y} - 1\right) = a + bx; \quad \lg \lg \frac{A}{y} = a + bx.$$

3. Для того чтобы получить возможность использовать метод наименьших квадратов (раздел 5.3.1), приравниваем левые части формул к некоторой величине z и записываем функции в линейной форме:

$$\lg\left(\frac{A}{y} - 1\right) = z, \quad z = a + bx; \quad \lg \lg \frac{A}{y} = z, \quad z = a + bx.$$

4. Пользуясь алгоритмом, помещенным в заголовках столбцов табл. 49 и 50, производим в этих таблицах необходимые расчеты для получения сумм Σx , Σx^2 , Σz и Σxz (нижняя строка таблиц). Система нормальных уравнений для обеих функций одинакова:

$$\begin{cases} na + b \Sigma x = \Sigma z, \\ a \Sigma x + b \Sigma x^2 = \Sigma xz, \end{cases}$$

где n — число обрабатываемых наблюдений, равное числу целиком заполненных строк в левой части табл. 49 и 50. Подставляя в эту систему соответствующие суммы, имеем

$$\begin{cases} 12a + 176b = -3,9509, & \begin{cases} 7a + 311b = -4,7098, \\ 311a + 17699b = -349,1459. \end{cases} \\ 176a + 3034b = -119,0845; \end{cases}$$

5. Решая системы, находим значения неизвестных:

$$a = 1,6523, \quad b = -0,1351; \quad a = 0,9266, \quad b = -0,0360.$$

6. Подставляя найденные значения параметров a и b , а также величину A в исходные интерполяционные формулы, записываем уравнения искомых теоретических кривых:

$$y = \frac{263}{1 + 10^{1,6523 - 0,1351x}}; \quad y = \frac{242,5}{10^{10^{0,9266 - 0,0360x}}}.$$

7. Путем обратного расчисления, алгоритм которого дан в заголовках средней части табл. 49 и 50, получаем ординаты точек теоретических кривых (y_T) и строим графики кривых рос-

та (рис. 20), на которых видно хорошее совпадение эмпирических данных с моделирующими рост листа функциями. Проверку этого можно выполнить с помощью критерия χ^2 (раздел 1.5.2), что мы предлагаем проделать читателю самостоятельно (при этом табличное значение критерия следует брать для числа степеней свободы $\nu = k - 3$, где k — число рассчитанных точек (y_T) теоретических кривых).

6.2.2. Расчет основных параметров роста. Анализ логистической функции и функции Гомпертца позволяет углубить исследование процесса роста путем определения ряда важных его характеристик (параметров). Покажем это,

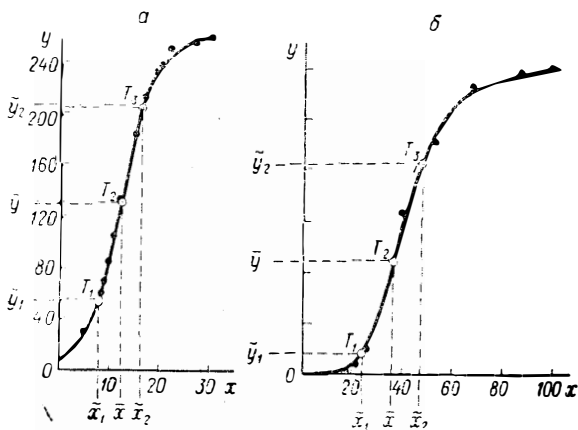


Рис. 20. Кривые роста.

a — логистическая кривая роста 8-го листа *Nicotiana rustica* L., *б* — гомпертцова кривая роста 7-го листа *Bryophyllum daigremontianum* Berger. Точками показаны эмпирические значения, остальные пояснения в тексте.

продолжив начатый в предыдущем разделе анализ роста листа *Nicotiana rustica* и *Bryophyllum daigremontianum*, причем, как и ранее, будем вести работу с двумя функциями параллельно. Излагая технику расчетов, мы не станем касаться математического обоснования выведенных формул [Васильев и др., 1973, 1975], а приведем их в готовом виде, одновременно производя соответствующие вычисления.

Нахождение параметров роста листа *Nicotiana rustica* (логистическая функция)

Нахождение параметров роста листа *Bryophyllum daigremontianum* (функция Гомпертца)

1. Кривые роста, изображенные на рис. 20, имеют перегиб в точке T_2 , характеризующей момент перехода возрастающей

скорости роста в убывающую. Определим координаты (\bar{x} ; \bar{y}) точки T_2 :

$$\begin{aligned}\bar{x} &= \frac{a}{|b|} = \frac{1,6523}{0,1351} = 12,2 \text{ сут}, & \bar{x} &= \frac{\lg \lg e - a}{b} = \frac{0,3622 + a}{|b|} = \\ \bar{y} &= 0,5A = 0,5 \cdot 263 = 131,5 \text{ мм}; & &= \frac{0,3622 + 0,9266}{0,0360} = 35,8 \text{ сут}, \\ & & \bar{y} &= \frac{A}{e} = 0,3679A = 0,3679 \times \\ & & & \times 242,5 = 89,2 \text{ мм}.\end{aligned}$$

Таким образом, максимальная скорость роста, величина которой нам пока еще неизвестна, достигается в тот момент, когда размер растущего органа составляет половину своего окончательного размера при подчинении роста логистической функции или 0,3679 окончательного размера при описании роста функцией Гомпертца.

2. Определим величину скорости роста (суточного прироста) в любой точке кривых. Для этого необходимо вычислить первые производные (y') от исходных функций:

$$y' = \frac{-A \cdot b \cdot 10^{a+bx}}{(1 + 10^{a+bx})^2} \cdot \ln 10; \quad y' = \frac{b}{(\lg e)^2} \cdot \lg \frac{y}{A} \cdot y.$$

Работать с этими формулами неудобно, поэтому путем некоторых преобразований приводим их к более простому рабочему виду:

$$y' = 2,3026 \cdot |b| \cdot y_{\tau} \left(1 - \frac{y_{\tau}}{A}\right); \quad y' = 5,3022 \cdot b \cdot y_{\tau} \cdot \lg \frac{y_{\tau}}{A}.$$

Подставляя в эти формулы известные нам значения b , y_{τ} и A , рассчитываем скорость роста в точках теоретических кривых по дням наблюдений и записываем результаты в колонке (y') правой части таблиц 49 и 50. Например, для шестой сверху строки каждой из этих таблиц имеем

$$\begin{aligned}y' &= 2,3026 \cdot 0,1351 \cdot 103,7 \times & y' &= 5,3022 \cdot (-0,0360) \cdot 190,6 \times \\ &\times \left(1 - \frac{106,7}{263}\right) = 19,7 \text{ мм/сут}; & &\times (-0,1037) = 3,8 \text{ мм/сут}.\end{aligned}$$

По найденным значениям первой производной (y') строим графики скорости роста листьев (рис. 21), с которых при необходимости можно снять приближенное значение скорости роста за любой день.

3. Подставляя значения $y=0,5A$ (для логистической функции) и $y=0,3679A$ (для функции Гомпертца) в формулы первых производных, после некоторых расчетов можно получить следующие простые выражения для вычисления максимальной

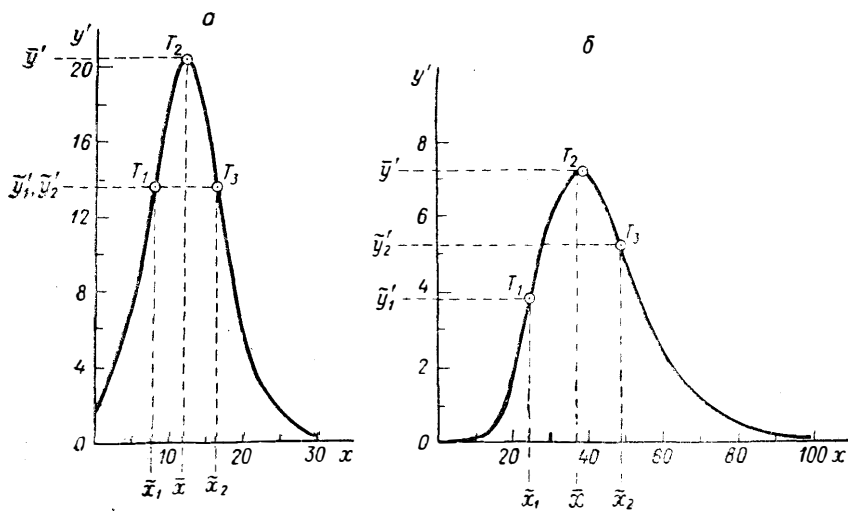


Рис. 21. Кривые скорости роста.
 а — 8-го листа *Nicotiana rustica* L.; б — 7-го листа *Bryophyllum daigremontianum* Berger. Пояснения в тексте.

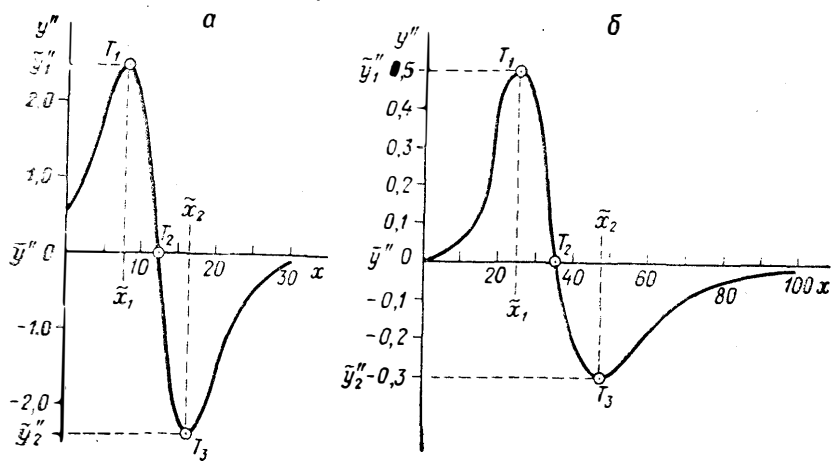


Рис. 22. Кривые ускорения роста.
 а — 8-го листа *Nicotiana rustica* L.; б — 7-го листа *Bryophyllum daigremontianum* Berger. Пояснения в тексте.

скорости роста (\bar{y}'), которая достигается в точке перегиба (T_2) обеих кривых:

$$\begin{aligned} \bar{y}' &= 0,5756 \cdot A \cdot |b| = 0,5756 \times & \bar{y}' &= 0,8479 \cdot A \cdot |b| = 0,8479 \times \\ &\times 263 \cdot 0,1351 = 20,5 \text{ мм/сут}; & &\times 242,5 \cdot 0,0360 = 7,4 \text{ мм/сут} \end{aligned}$$

4. Определяем величину ускорения скорости роста в любой точке кривых, для чего надо вычислить вторые производные (y'') от исходных функций:

$$y'' = 2,3026 \cdot |b| \cdot y' \cdot \left(1 - \frac{2y_T}{A}\right); \quad y'' = 5,3022b \cdot y' \cdot \lg \frac{y_T \cdot e}{A}.$$

Подставляя в эти формулы известные нам значения b , A , y' и y_T , рассчитываем ускорение скорости роста по дням наблюдений и записываем результаты в колонке (y'') правой части табл. 49 и 50. Например, для шестой сверху строки каждой из таблиц имеем

$$\begin{aligned} y'' &= 2,3026 \cdot 0,1351 \cdot 19,7 \times & y'' &= 5,3022 \cdot (-0,0360) \cdot 3,8 \times \\ &\times \left(1 - \frac{2 \cdot 106,7}{263}\right) = 1,2 \text{ мм/сут}^2; & &\times 0,3298 = 0,2 \text{ мм/сут}^2. \end{aligned}$$

По найденным значениям второй производной (y'') строим графики ускорения скорости роста (рис. 22), которые могут служить для приближенного определения изменения скорости роста листьев за любой день.

Обратим внимание на положение точек T_1 , T_2 и T_3 на графиках (рис. 20—22). Точка T_2 на всех графиках имеет известную нам абсциссу \bar{x} . Ее ординаты на кривых роста (\bar{y} на рис. 20) и кривых скорости роста (\bar{y}' на рис. 21), характеризующие момент наибольшей скорости роста, также были определены. Из положения точки T_2 на кривых ускорения роста (\bar{y}'' на рис. 22) следует, что в этот момент ускорение роста является нулевым. Таким образом, координаты точки T_2 на всех графиках нам известны, понятен и ее биологический смысл.

Что касается координат точек T_1 и T_3 , то они пока еще неизвестны. Между тем из рис. 20—22 видно, что они характеризуют некоторые переломные моменты ростового процесса. Например, на рис. 22 эти точки соответствуют максимальному и минимальному ускорению роста. Следовательно, важно определить их координаты, что сопряжено с вычислением третьих производных от исходных функций. Не вдаваясь в математические выкладки, приведем далее простые рабочие формулы и результаты вычислений.

5. Определяем ординаты (\bar{y}_1 и \bar{y}_2) точек T_1 и T_3 на кривых роста (рис. 20):

$$\begin{aligned}\bar{y}_1 &= 0,2113A = 0,2113 \cdot 263 = 55,6 \text{ мм}, & \bar{y}_1 &= 0,0730A = 0,0730 \times \\ & & & \times 242,5 = 17,7 \text{ мм}, \\ \bar{y}_2 &= 0,7887A = 0,7887 \cdot 263 = 207,4 \text{ мм}; & \bar{y}_2 &= 0,6825A = 0,6825 \times \\ & & & \times 242,5 = 165,5 \text{ мм}.\end{aligned}$$

Как видно из формул, эти ординаты в случае логистической кривой соответствуют приблизительно 21 и 79%, а в случае функции Гомпертца — около 7 и 68% окончательного размера органа.

6. Находим ординаты (\bar{y}'_1 и \bar{y}'_2) точек T_1 и T_3 на кривых скорости роста (рис. 21), т. е. определяем скорость роста в этих точках:

$$\begin{aligned}\bar{y}'_1 = \bar{y}'_2 &= 0,3838 \cdot A \cdot b = 0,3838 \times & \bar{y}'_1 &= 0,4405 \cdot A \cdot b = 0,4405 \times \\ & \times 263 \cdot 0,1351 = 13,6 \text{ мм/сут}; & & \times 242,5 \cdot 0,0360 = 3,8 \text{ мм/сут}, \\ & & \bar{y}'_2 &= 0,6014 \cdot A \cdot b = 0,6014 \times \\ & & & \times 242,5 \cdot 0,0360 = 5,3 \text{ мм/сут}.\end{aligned}$$

Функция Гомпертца асимметрична, поэтому $\bar{y}'_1 \neq \bar{y}'_2$

7. Рассчитываем ординаты (\bar{y}''_1 и \bar{y}''_2) точек T_1 и T_3 на кривых ускорения скорости роста (рис. 22), т. е. определяем максимальное и минимальное ускорение роста:

$$\begin{aligned}\bar{y}''_1 &= 0,5102 \cdot Ab^2 = 0,5102 \cdot 263 \times & \bar{y}''_1 &= 1,6385 \cdot Ab^2 = 1,6385 \times \\ & \times 0,1351^2 = 2,5 \text{ мм/сут}^2, & & \times 242,5 \cdot 0,0360^2 = 0,5 \text{ мм/сут}, \\ \bar{y}''_2 &= -0,5102 \cdot Ab^2 = -0,5102 \times & \bar{y}''_2 &= -0,8540 \cdot Ab^2 = -0,8540 \times \\ & \times 263 \cdot 0,1351^2 = -2,5 \text{ мм/сут}^2; & & \times 242,5 \cdot 0,0360^2 = -0,3 \text{ мм/сут}^2.\end{aligned}$$

Поскольку в точках T_1 и T_3 ускорение имеет разный знак, то, в сущности, можно говорить о максимальном ускорении и максимальном торможении роста в этих точках. В случае функции Гомпертца указанные величины отличаются не только своим знаком, но и абсолютным значением.

8. Определяем абсциссы ($\bar{x}_{1,2}$) точек T_1 и T_3 , которые на всех трех графиках (рис. 20—22) для каждой из функций одинаковы:

$$\begin{aligned}\bar{x}_{1,2} &= \bar{x} \pm \frac{0,5719}{b}, & \bar{x}_{1,2} &= \bar{x} \pm \frac{0,4180}{b}, \\ \bar{x}_1 &= 12,2 - \frac{0,5719}{0,1351} = 8,0 \text{ сут}, & \bar{x}_1 &= 35,8 - \frac{0,4180}{0,0360} = 24,2 \text{ сут}, \\ \bar{x}_2 &= 12,2 + \frac{0,5719}{0,1351} = 16,4 \text{ сут}; & \bar{x}_2 &= 35,8 + \frac{0,4180}{0,0360} = 47,4 \text{ сут}.\end{aligned}$$

9. Находим продолжительность основной фазы роста (Φ) между точками T_1 и T_3 :

$$\Phi = \bar{x}_2 - \bar{x}_1 = 16,4 - 8,0 = 8,4 \text{ сут}; \quad \Phi = \bar{x}_2 - \bar{x}_1 = 47,4 - 24,2 = 23,2 \text{ сут.}$$

Тот же результат можно получить еще до расчета значений \bar{x}_1 и \bar{x}_2 по следующим формулам:

$$\Phi = \frac{1,1438}{|b|} = \frac{1,1438}{0,1351} = 8,5 \text{ сут}$$

(разница в 0,1 сут по сравнению с предыдущим результатом возникает за счет промежуточных округлений);

$$\Phi = \frac{0,8359}{|b|} = \frac{0,8359}{0,0360} = 23,2 \text{ сут.}$$

Точки T_1 , T_2 и T_3 , характеризующие переломные моменты ростового процесса, получили название *критических* точек роста. В точке T_1 растущий орган имеет первый критический возраст (возраст максимального ускорения роста) и первый критический размер, составляющий приблизительно 21% (в случае логистической кривой) или 7% (в случае функции Гомпертца) от окончательного своего размера. В точке T_2 орган имеет второй критический возраст (возраст нулевого ускорения и максимальной скорости роста) и второй критический размер, равный 50% (в случае логистической кривой) или около 37% (в случае функции Гомпертца) окончательного размера. Наконец, в точке T_3 орган достигает третьего критического возраста (возраста максимального торможения роста) и третьего критического размера, составляющего примерно 79% (в случае логистической кривой) или 68% (в случае функции Гомпертца) окончательного размера.

В соответствии со сказанным весь период роста можно подразделить на три фазы: 1) фазу увеличения скорости и ускорения роста (от начала роста, когда скорость, ускорение и размер органа равны нулю, до достижения критической точки T_1); 2) фазу максимальной скорости роста и уменьшения ускорения (от первой до третьей критической точки, причем на участке T_1 — T_2 скорость возрастает до максимума, а ускорение, уменьшаясь, падает до нуля, тогда как на участке T_2 — T_3 скорость роста начинает замедляться, а ускорение приобретает отрицательный знак и падает до минимума); 3) фазу уменьшения скорости и увеличения ускорения (от третьей критической точки до остановки роста, когда ускорение из области отрицательных значений возвращается к нулю).

В справедливости выделенных фаз ростового процесса (из которых вторая фаза является основной) легко убедиться, сопоставляя приведенные выше их описания с ходом кривых и по-

ложением критических точек на графиках (рис. 20—22). Любое ростовое явление, описываемое логистической функцией или функцией Гомпертца, характеризуется тремя критическими точками и тремя рассмотренными фазами самого процесса. Вполне понятно, что изучение анатомических, морфологических, физиологических и биохимических аспектов роста, выполненное в момент достижения растущим органом (или растением) критических точек, обеспечивает сопоставимость полученных результатов и позволяет глубже раскрыть сущность ростовых процессов. Поэтому вычисление основных параметров роста является необходимым этапом подобного рода исследований и важным звеном в разработке теории роста растений.

6.2.3. Обобщенный класс функций роста. Логистическая функция и функция Гомпертца, хорошо описывающие ростовые процессы, не исчерпывают, однако, теоретически возможных и фактически известных типов роста у растений. Известны примеры, когда рост (например, рост гипокоты кукурузы, листьев вьющейся фасоли, листьев, развивающихся из перезимовавших почек древесных растений) описывается S-образной кривой с растянутой нижней ветвью. Такую кривую можно получить путем поворота кривой Гомпертца вокруг центра симметрии (\bar{x} ; \bar{y}), и в противоположность последней она характеризуется удлиненной фазой увеличения скорости роста и укороченной фазой замедления его. Формально рост растений в ряде случаев может быть описан функцией Берталанфи, а при описании отдельных этапов ростового процесса возможно использование функций экспоненциального типа.

В связи с этим потребовалось найти такой класс функций, который объединил бы известные в биологии типы кривых роста. Полученная формула [Савинов и др., 1977] имеет вид

$$y = A \left(1 - \frac{e^{a+bx}}{k} \right)^k,$$

где k — любое вещественное число ($-\infty < k < +\infty$), а остальные обозначения — те же, что в формулах логистической и гомпертцовой функций (с. 134), причем всегда $b < 0$.¹

Дополнительный по сравнению с логистической и гомпертцовой функциями параметр k , в известном смысле аналогичный одноименному параметру функции Берталанфи, обладает, однако, более важными и общими свойствами: придавая ему раз-

¹ Расчеты, связанные с использованием этой формулы, целесообразно выполнять с помощью ЭВМ, для чего С. Ф. Колодяжным в БИНИИ ЛГУ разработана специальная программа, выбирающая по эмпирическим данным наиболее подходящий тип теоретической кривой и вычисляющая необходимые параметры роста.

личные значения, можно получить все известные типы ростовых кривых и уточнить их классификацию.

Так, положив $k = -1$, мы получаем логистическую функцию

$$y = A \left(1 - \frac{e^{a+bx}}{-1} \right)^{-1} = \frac{A}{1 + e^{a+bx}},$$

отличающуюся от использованной нами ранее формулы (с. 134) лишь тем, что в ее знаменателе вместо основания десятичных стоит основание натуральных логарифмов.

Принимая $k \rightarrow \pm\infty$, после некоторых преобразований [Савинов и др., 1977] можно убедиться, что в этом случае формула обобщенного класса кривых приводит нас к функции Гомпертца, причем последняя оказывается предельным элементом данного класса:

$$A \left(1 - \frac{e^{a+bx}}{k} \right)^k \xrightarrow{k \rightarrow \infty} A e^{-e^{a+bx}}$$

(правая часть выражения есть один из способов записи функции Гомпертца).

Таблица 51. Некоторые свойства обобщенного класса функций роста

Тип кривой	Формула	Область изменения параметров	Число критических точек	Ордината (\bar{y}) критической точки T_1
Обобщенный класс функций роста	$y = A \left(1 - \frac{e^{a+bx}}{k} \right)^k$	$A > 0; a > 0; b < 0;$ $-\infty < k < +\infty$	0-3	$\bar{y} = A \left(1 - \frac{1}{k} \right)^k$
Кривые Берталанфи	То же	$2 < k < +\infty$	3	$\frac{A}{4} < \bar{y} < \frac{A}{e}$
То же	" "	$1 < k \leq 2$	2	$0 < \bar{y} < \frac{A}{4}$
" "	" "	$0 < k < 1$	0	—
Гипологистическая кривая	" "	$-1 < k < 0$	3	$\frac{A}{2} < \bar{y} < A$
Гиперлогистическая кривая	" "	$-\infty < k < -1$	3	$\frac{A}{e} < \bar{y} < \frac{A}{2}$
Логистическая кривая	$y = A(1 + e^{a+bx})^{-1}$	$k = -1$	3	$\bar{y} = \frac{A}{2}$
Экспоненциальная кривая	$y = A(1 - e^{a+bx})$	$k = +1$	0	—
Кривая Гомпертца	$y = A e^{-e^{a+bx}}$	$k \rightarrow \pm\infty$	3	$\bar{y} = \frac{A}{e}$

При $k=1$ формула обобщенного класса кривых дает экспоненциальную функцию

$$y = A(1 - e^{a+bx}),$$

а при других значениях параметра k в интервале $0 < k < +\infty$ возникают кривые Берталанфи, имеющие экспоненциальный (при $0 < k < 1$) или S-образный характер. Кривые, у которых величина параметра k удовлетворяет условию $-1 < k < 0$, предложено называть *гипологистическими*, а кривые, характеризующиеся значениями $-\infty < k < -1$, — *гиперлогистическими*.

Некоторые свойства различных типов кривых, входящих в обобщенный класс функций роста, указаны в табл. 51, а зависимость типа кривой от параметра k изображена на рис. 23.

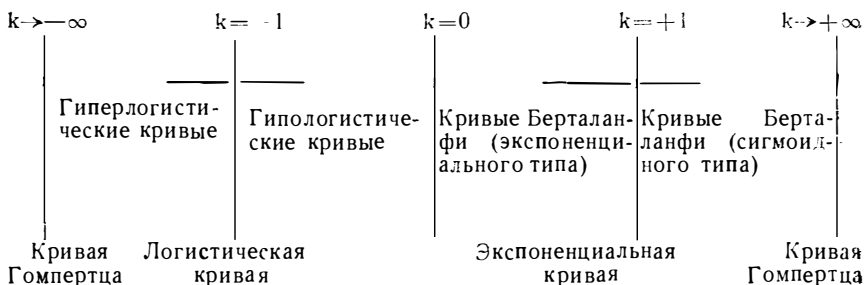


Рис. 23. Зависимость типа кривых роста от параметра k в пределах обобщенного класса функций.

Исследование кривых данного класса показывает, что при значениях $k > 2$ и $k < 0$ соответствующие кривые обладают всеми тремя критическими точками (T_1 , T_2 и T_3), биологический смысл которых был рассмотрен выше. При этом ордината \bar{y} точки T_2 может принимать любое значение в интервале между A и $A:4$. Точка перегиба T_2 существует и для кривых, характеризующихся условием $1 < k < 2$ (при этом ордината \bar{y} этой точки лежит в интервале между 0 и $A:4$), однако в этом случае исчезает точка T_1 . Наконец, при условии $0 \leq k \leq 1$ критические точки T_1 , T_2 и T_3 отсутствуют вовсе, так как в этом интервале мы имеем дело с кривыми экспоненциального типа.

Описываемый класс функций роста отличается некоторым внутренним единством, выражающимся в плавном переходе от одного типа кривых к другому (рис. 23). Формально это означает, что в пределах данного класса не существует принципиальных ограничений характера роста у растений, однако кривые экспоненциального типа, ограниченные условием $0 \leq k \leq 1$, лишены критических точек и неспособны описать процесс роста полностью, соответствуя лишь отдельным его этапам.

На оси диаграммы (рис. 23) логистической и экспоненциальной функциям соответствуют не интервалы, как кривым иного типа, а точки, отвечающие значениям $k = -1$ и $k = +1$. Однако переход кривых из одного типа в другой совершается постепенно, поэтому на практике вполне удовлетворительный результат выравнивания эмпирического ряда с помощью этих функций получается и при некоторых отклонениях от указанных значений параметра k . То же самое относится и к функции Гомпертца ($k \rightarrow \pm \infty$): практически она удовлетворительно выравнивает эмпирические ряды уже при $k = \pm 4$.

6.3. СООТНОСИТЕЛЬНЫЙ РОСТ ОРГАНОВ И ЕГО ОПИСАНИЕ АЛЛОМЕТРИЧЕСКОЙ ФУНКЦИЕЙ

Скорость роста разных органов (или частей) растений и животных различна и, как мы видели, изменяется в ходе онтогенеза. Важнейшей особенностью одновременного роста двух или нескольких органов является то, что изменение скоростей их

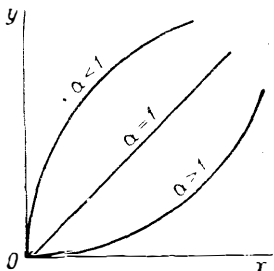


Рис. 24. Схематическое изображение аллометрической функции при пороговых значениях α -константы.

роста происходит синхронно, т. е. скоррелировано таким образом, что отношение скоростей роста остается приблизительно постоянной величиной. В этом и состоит биологическая сущность явления соотносительного (аллометрического) роста, в общей форме известного еще со времен Ж. Кювье, но впервые серьезно изученного Дж. Гексли [Huxley, 1932].

Связь между растущим органом и размером тела (или между двумя растущими органами) является нелинейной и обычно хорошо описывается аллометрической функцией: $y = bx^\alpha$, где y — размер одного органа; x — размер другого органа (или тела растения, животного); b — константа начального роста (при $x=1$ $y=b$); α — константа равновесия (аллометрический показатель), передающая относительную скорость (темп) роста одного органа по сравнению с другим.

Константа α имеет важный биологический смысл. При $\alpha = 1$ оба органа растут с одинаковой скоростью (случай изометрии) и их относительные размеры остаются постоянными. При $\alpha > 1$ (случай положительной аллометрии) из-за преимущества в скорости роста размер одного органа увеличивается относительно размера другого органа. При $\alpha < 1$ (отрицательная аллометрия) в связи с меньшей скоростью роста размер органа уменьшается относительно размера другого органа. Наконец, возможен случай энантиметрии ($\alpha < 0$), при котором абсолютные размеры органа в ходе онтогенеза уменьшаются (в зоологии —

уменьшение длины хвоста головастика по мере превращения его в лягушку). Ход аллометрической функции на графике при трех первых пороговых значениях α -константы схематически изображен на рис. 24.

Методика аллометрии представляет собой еще один частный случай регрессионного анализа. Порядок работы остается обычным. Для использования метода наименьших квадратов аллометрическую функцию предварительно путем логарифмирования приводят к линейной форме: $\lg y = \lg b + \alpha \lg x$. Тогда система нормальных уравнений приобретает вид

$$\begin{cases} n \lg b + \alpha \sum \lg x = \sum \lg y, \\ \lg b \sum \lg x + \alpha \sum (\lg x)^2 = \sum \lg x \cdot \lg y. \end{cases}$$

Обратимся к конкретному примеру, продолжив начатый в разделах 5.3 и 5.4 анализ регрессии длины листа *Melampyrum polonicum* на его ширину (напомним, что использование уравнений прямой линии, парабол 2-й и 3-й степеней приводило нас к удовлетворительной интерполяции эмпирических данных, но не позволяло производить экстраполяцию). Попробуем применить к этому случаю аллометрическую функцию. Расчеты, показанные в левой части табл. 52, приводят к следующей системе нормальных уравнений:

Таблица 52. Выравнивание эмпирической линии регрессии длины листа (y) на его ширину (x) у *Melampyrum polonicum* (Beauv.) Soó уравнением аллометрической функции (y')

x	y	$\lg x$	$\lg y$	$\lg x \cdot \lg y$	$\lg(x)^2$	$\alpha \lg x$	$\lg b + \alpha \lg x = \lg y'$	y'
1	19,5	0	1,2900	0	0	0	1,2681	18,5
4	30,2	0,6021	1,4800	0,8911	0,3625	0,2557	1,5238	33,4
7	41,8	0,8451	1,6212	1,3701	0,7142	0,3588	1,6269	42,4
10	51,5	1,0000	1,7118	1,7118	1,0000	0,4246	1,6927	49,3
13	52,4	1,1139	1,7193	1,9151	1,2408	0,4730	1,7411	55,1
16	64,5	1,2041	1,8096	2,1789	1,4499	0,5113	1,7794	60,2
—	—	4,7652	9,6319	8,0670	4,7674	—	—	—

$$\begin{cases} 6 \lg b + 4,7652 \alpha = 9,6319, \\ 4,7652 \lg b + 4,7674 \alpha = 8,0670. \end{cases}$$

Решая эту систему, получаем: $\alpha = 0,4246$; $\lg b = 1,2681$ и $b = 18,54$. Следовательно, искомая аллометрическая функция

имеет вид: $y = 18,54x^{0,4246}$, или в логарифмической форме: $\lg y = 1,2681 + 0,4246 \lg x$. Используя последнюю запись, производим обратное расчисление (правая часть табл. 52), находим точки теоретической линии регрессии (y') и строим ее график (рис. 25, а). Как видно, аллометрическая функция хорошо описывает эмпирическую регрессию и позволяет экстраполировать опытные данные (в частности, соблюдается условие: $y=0$ при $x=0$).

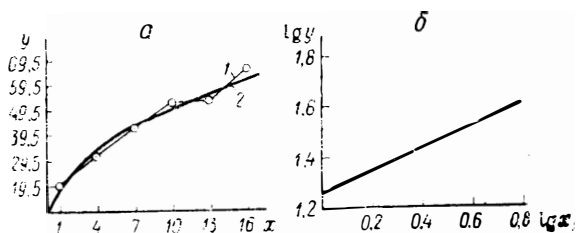


Рис. 25. Аллометрическая функция, отражающая регрессию длины листа *Melampyrum pratense* (Beauv.) Soó на его ширину в натуральных (а) и логарифмическом (б) масштабах.

1 — эмпирическая и 2 — теоретическая линии регрессии.

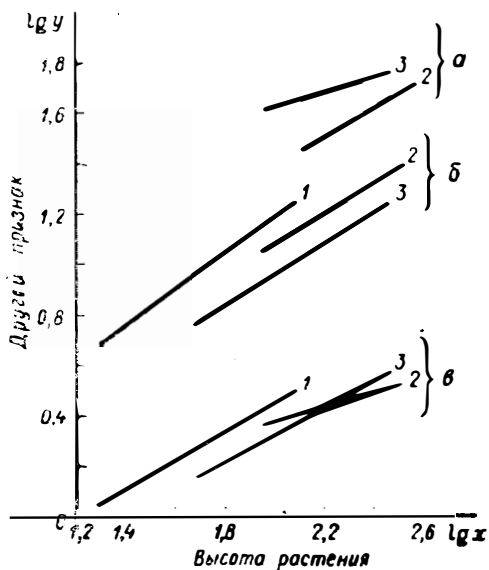


Рис. 26. Примеры сравнения аллометрического роста органов у видов *Odontites*. 1 — *O. aitoralis* Fr.; 2 — *O. verna* (Bell.) Dum.; 3 — *O. serotina* Dum. а — число пар ветвей; б — длина верхнего листа; в — число зубцов с одной стороны листа.

График аллометрической функции, построенный в логарифмическом масштабе (рис. 25, б), имеет вид прямой линии с угловым коэффициентом α . На таком графике нагляднее видна относительная скорость роста: чем больше угол наклона прямой, тем эта скорость выше. Кроме того, именно так удобнее сравнивать характер аллометрического роста одноименных органов у разных видов (рис. 26), что небесполезно для целей систематики.

Из уравнения аллометрической функции $\lg y = \lg b + \alpha \lg x$ видно, что α -константа есть не что иное, как коэффициент регрессии $R_{\lg y / \lg x}$. Таким образом, α -константа передает не только относительный темп роста двух органов (частей), но косвенно отражает и размерную зависимость изменения одного из них при изменении другого.

Поскольку мы имеем дело с двумя признаками (x и y), то существует возможность вычисления двух значений α - и b -констант, характеризующих темп роста признака y относительно признака x (из уравнения $\lg y = \lg b_1 + \alpha_1 \lg x$) и признака x относительно признака y (из уравнения $\lg x = \lg b_2 + \alpha_2 \lg y$), причем $\alpha_1 \neq \alpha_2$ и $b_1 \neq b_2$, но при полной корреляционной связи ($r = \pm 1$) соблюдается соотношение $\alpha_1 = 1 : \alpha_2$.

С учетом сказанного в разделе 5.3.2 можно записать:

$$\alpha_1 = R_{\lg y / \lg x} = r \frac{\sigma_{\lg y}}{\sigma_{\lg x}} \quad \text{и} \quad \alpha_2 = R_{\lg x / \lg y} = r \frac{\sigma_{\lg x}}{\sigma_{\lg y}},$$

где r — коэффициент линейной корреляции; σ — соответствующие средние квадратичные отклонения.

Тогда соотношение между коэффициентом корреляции и α -константами может быть выражено уравнением

$$r = \sqrt{\alpha_1 \cdot \alpha_2}.$$

Исследование аллометрических зависимостей может преследовать разные цели. Различен и биологический смысл получаемых при этом α -констант.

При непосредственном наблюдении за ростом органов или их частей мы имеем дело с аллометрией в ее «чистом» виде, изучаем динамику ростового процесса и получаем α -константы, определяющие относительный темп роста органов. Эти показатели имеют подчеркнуто онтогенетический смысл. Иная ситуация складывается при анализе выборки из множества особей, когда фиксируются размеры органов (или их частей), как достигших окончательной или квантовой промежуточной величины, так и наблюдаемых в период осуществления очередного кванта роста. В этом случае α -константы служат показателями «аллометрии размера». Первый из указанных аспектов изучения аллометрии важен при исследовании проблем роста и органогенеза, а второй — для систематики, включая изучение структуры популяций по величине α -констант.

На ряде ботанических и зоологических объектов показано, что величина α -констант наследственно закреплена и слабо изменяется под воздействием экологических факторов. Напротив, географическая дифференциация вида может сопровождаться существенным сдвигом величины α -констант. Показано также, что таксономически значимые признаки обычно существенно отличаются величиной своих α -констант по отношению к размерам тела растения или животного, тогда как для признаков, не имеющих таксономического значения, справедливым оказывается обратное утверждение. Сказанное указывает на перспективность использования методики аллометрии для целей систематики.

Глава 7

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В СИСТЕМАТИКЕ

7.1. ПРОСТЕЙШИЕ ПРИЕМЫ СРАВНЕНИЯ ТАКСОНОВ

Рассмотренные в главах 1—5 методы используются в систематике для изучения и сравнения изменчивости отдельных признаков, их зависимости от факторов среды, для выявления связи между признаками. Нередко эти методы оказываются достаточными для решения частного систематического вопроса, однако в общем случае составляют лишь начальный этап биометрического исследования.

Дальнейшие действия систематика должны состоять: а) в выборе и расчете обобщенной меры сходства между изучаемыми объектами по комплексу признаков и б) в группировке (классификации) объектов в зависимости от степени сходства.

Иногда примененные даже самых простых способов, описываемых в данном разделе, приносит полезные результаты. Однако в целом современная систематика должна ориентироваться на использование более информативных методов, изложенных в последующих разделах этой главы.

7.1.1. Полигональные графики (или «полиграфы») в качестве приема наглядного сравнения признаков изучаемых объектов были предложены А. Хатчинсоном в 1936 г., но мало использовались систематиками.

Способ их построения крайне прост. От исходной точки по радиусам в произвольно выбранном масштабе, одинаковом для

сравниваемых таксонов, откладываются значения выборочных параметров (M , σ или V) изученных признаков. Полученные точки соединяются ломаной линией, образующей полигональную фигуру (рис. 27). Сравнение таких фигур (особенно удобное в том случае, если они выполнены на кальке и их можно накладывать друг на друга) дает глазомерное представление как о степени совпадения объектов по отдельным признакам, так и о степени их общего сходства по всему комплексу признаков.

7.1.2. Коэффициент дивергенции признаков (КД), предложенный С. Р. Царапкиным в 1960 г., основан на изучении степени расхождения признаков сравниваемых объектов.

Для учитываемых признаков каждого объекта предварительно должны быть рассчитаны средние арифметические (M) и средние квадратические отклонения (σ). Затем один из объектов (соответствующая ему выборка) принимается за стандарт. Удобно, если в качестве стандарта выбирается исходный в каком-либо отношении объект, например предполагаемая предковая форма или типичная популяция некоторого вида, тогда как другие выборки предположительно характеризуют производные формы или (географические, экологические и т. д.) отклонения от стандарта.

Для определения величины дивергенции важна не абсолютная разница между средними арифметическими значениями признаков стандарта и остальных выборок, а величина колебания этих отклонений, выраженная в едином масштабе. Это достигается отнесением отклонений к сигме стандарта, т. е. их нормированием:

$$\delta = \frac{M_i - M_{st}}{\sigma_{st}},$$

где δ — нормированное отклонение; M_i — выборочная средняя

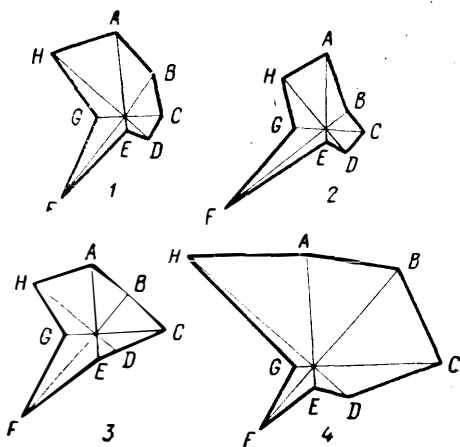


Рис. 27. Полигональные графики, построенные по средним арифметическим значениям признаков спор у видов *Osmunda* (по: Бобров А. Е., 1966).

1 — *O. javanica* Bl.; 2 — *O. cinnamomea* L.; 3 — *O. claytoniana* L.; 4 — *O. regalis* L. Признаки: А — диаметр спор, В — высота бугров, С — ширина бугров; D — длина луча рубца, E — отношение ширины бугра к толщине оболочки, F — отношение длины луча к диаметру споры, G — процент деформации спор, H — толщина оболочки.

арифметическая; M_{st} и σ_{st} — средняя арифметическая и среднее квадратическое отклонение стандарта.

По этой формуле рассчитываются нормированные отклонения каждого признака всех сравниваемых выборок. Нанесение их на график (профиль отклонений) дает наглядное представление о степени расхождения признаков изучаемых объектов. Математическое выражение дивергенции объектов (КД) по комплексу включенных в анализ признаков получают путем расче-

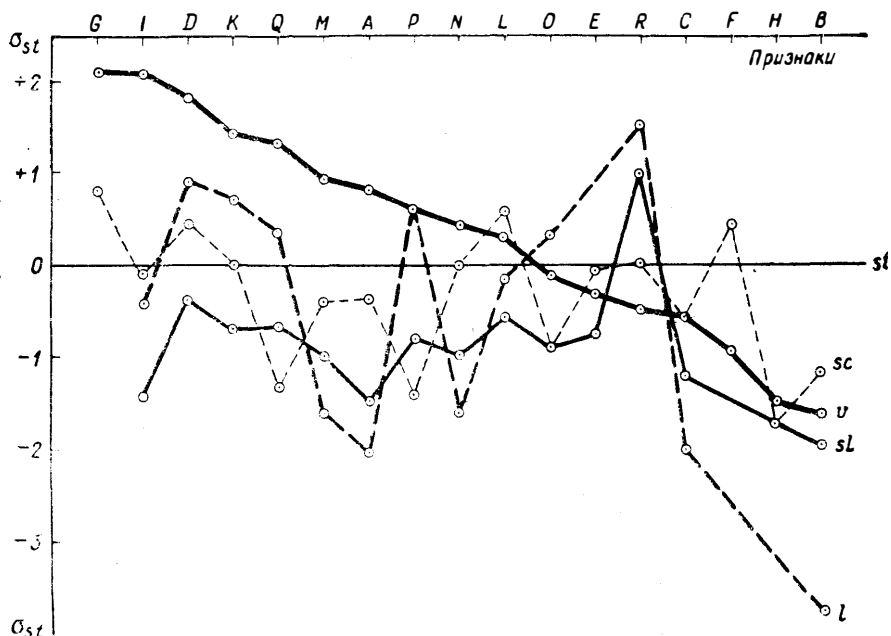


Рис. 28. Профиль отклонений признаков у видов и форм *Odontites* от стандарта.

st — *O. serotina* Dum. (стандарт); sc и sl — сорно-полевая и береговая экологические формы того же вида; v — *O. verna* (Bell.) Dum.; l — *O. litoralis* Fr. Признаки расположены в порядке убывания их отклонений у *O. verna* (Bell.) Dum.

та среднего квадратического отклонения (σ_{δ}) ряда, составленного из нормированных отклонений (δ): $КД = \pm \sigma_{\delta}$.

Для пояснения хода работы обратимся к примеру, который в дальнейшем будет использован для иллюстрации и других биометрических приемов. У трех видов рода *Odontites* — *O. serotina* Dum. (st), *O. verna* (Bell.) Dum. (v) и *O. litoralis* Fr. (l), а также у двух экологических форм первого из них — береговой (sl) и сорно-полевой (sc) — были подвергнуты изучению следующие признаки: А — высота растения, В — число междоузлий, С — число укороченных междоузлий, D — средняя дли-

на междоузлия, E — число пар ветвей, F — угол отклонения верхней ветви, G — уровень ветвления, H — число интеркалярных листьев, I — длина верхнего листа, K — ширина того же листа, L — число зубцов с одной стороны листа, M — длина осевой кисти, N — число пар цветков в ней, O — длина венчика, P — длина чашечки, Q — длина зубца чашечки и R — ширина зубца чашечки (последние 4 признака измерялись на нижней цветке осевой кисти).

В качестве стандарта был принят вид *O.serotina*, параметры которого и ход вычисления отклонения признаков остальных выборов от него показаны в табл. 53. Полученные в этой таблице ряды отклонений (δ) позволяют построить соответствующий график (рис. 28), из которого видно, какие признаки у каких объектов в той или иной степени отклоняются от признаков стандарта.

Чтобы получить коэффициент дивергенции, необходимо для каждого ряда сначала подсчитать сумму квадратов отклонений признаков от стандарта ($\sum \delta^2$, см. табл. 53), а затем вычислить среднее квадратическое отклонение (σ_{δ}), которое и характеризует значение КД:

$$\text{КД} = \sigma_{\delta} = \sqrt{\frac{\sum \delta^2}{n-1}},$$

где n — число признаков.

В нашем примере это дает:
для *O.verna*

$$\text{КД}_v = \sqrt{\frac{24,55}{17-1}} = 1,24;$$

для *O.litoralis*

$$\text{КД}_l = \sqrt{\frac{31,56}{13-1}} = 1,62;$$

для береговой формы *O.serotina*

$$\text{КД}_{sl} = \sqrt{\frac{15,88}{14-1}} = 1,10;$$

для сорно-полевой формы того же вида

$$\text{КД}_{sc} = \sqrt{\frac{10,69}{17-1}} = 0,82.$$

Полученные результаты находят удовлетворительное биологическое истолкование. Вполне понятно, что экологические формы дивергируют от исходного вида заметно слабее, чем родственные ему, но все же самостоятельные виды. Вид *O.litoralis* Fr. сильнее, чем *O.verna* (Bell.) Dum., уклоняется от

Таблица 53. Расчеты для построения профиля отклонений и вычисления коэффициента дивергенции (КД)

Признаки	st		v			l			sl			sc		
	M_{st}	σ_{st}	M_v	$\frac{M_v - M_{st}}{\sigma_{st}} = \delta_v$	σ_v^2	M_l	$\frac{M_l - M_{st}}{\sigma_{st}} = \delta_l$	σ_l^2	M_{sl}	$\frac{M_{sl} - M_{st}}{\sigma_{st}} = \delta_{sl}$	σ_{sl}^2	M_{sc}	$\frac{M_{sc} - M_{st}}{\sigma_{st}} = \delta_{sc}$	σ_{sc}^2
A	166,8	51,3	207,6	+0,80	0,64	62,7	-2,03	4,12	91,1	-1,48	2,19	146,7	-0,39	0,15
B	13,3	2,4	9,4	-1,63	2,66	4,3	-3,75	14,06	8,6	-1,96	3,84	10,5	-1,17	1,37
C	4,8	1,9	3,7	-0,58	0,34	1,0	-2,00	4,00	2,5	-1,21	1,46	3,8	-0,53	0,28
D	8,2	2,7	13,1	+1,81	3,28	10,6	+0,89	0,79	7,2	-0,37	0,14	9,4	+0,44	0,19
E	2,0	2,6	1,2	-0,31	0,10	—	—	—	0,1	-0,73	0,53	1,8	-0,08	0,01
F	48,3	10,6	38,4	-0,93	0,86	—	—	—	—	—	—	53,0	+0,44	0,19
G	45,6	21,0	90,2	+2,12	4,49	—	—	—	—	—	—	62,4	+0,80	0,64
H	4,3	2,2	1,0	-1,50	2,25	—	—	—	—	—	—	0,6	-1,68	2,82
I	12,3	3,1	18,8	+2,10	4,41	11,0	-0,42	0,18	7,9	-1,42	2,02	12,0	-0,10	0,01
K	2,5	0,7	3,5	+1,43	2,04	3,0	+0,71	0,50	2,0	-0,71	0,50	2,5	0	0
L	2,5	1,4	2,9	+0,29	0,08	2,3	-0,14	0,02	1,7	-0,57	0,32	3,3	+0,57	0,32
M	58,8	26,7	83,1	+0,91	0,83	17,8	-1,58	2,50	31,6	-1,02	1,04	47,1	-0,44	0,19
N	9,1	4,2	10,9	+0,43	0,18	2,4	-1,60	2,56	5,0	-0,98	0,96	9,1	0	0
O	8,1	0,9	8,0	-0,11	0,01	8,4	+0,33	0,11	7,3	-0,89	0,79	7,8	+0,89	0,79
P	5,2	0,5	5,5	+0,60	0,36	5,5	+0,60	0,36	4,8	-0,80	0,64	4,5	-1,40	1,96
Q	2,1	0,3	2,5	+1,33	1,77	2,2	+0,33	0,11	1,9	-0,67	0,45	1,7		1,77
R	1,2	0,2	1,1	-0,50	0,25	1,5	+1,50	2,25	1,4	+1,00	1,00	1,2	0	0
Σ	—	—	—	—	24,55	—	—	31,56	—	—	15,88	—	—	10,69

Примечания. Значения признаков см. на с. 150. Обозначения таксонов: st — *O. serotina* Dum. (стандарт), v — *O. verna* (Bell.) Dum., l — *O. litoralis* Fr., sl и sc — береговая и сорно-полевая экологические формы *O. serotina* Dum.

O.serotina Dum., но этого и следовало ожидать, ибо последние два вида настолько близки, что нередко смешиваются ботаниками. Береговая форма *O.serotina* Dum. подвержена гораздо более жесткому отбору, нежели сорно-полевая форма этого вида, и неудивительно, что рассчитанные значения КД указывают на ее более сильную дивергенцию от исходного типа.

Интересная модификация рассматриваемого метода, предложенная в 1973 г. Е. Я. Зяблицкой, предусматривает оценку отклонения значений признака от стандарта не в сигмах последнего, а посредством расчета критерия t (раздел 1.5.1):

$$t = \frac{M_i - M_{st}}{\sqrt{m_i^2 + m_{st}^2}},$$

где M_i и M_{st} — средние арифметические значения признаков выборки и стандарта; m_i и m_{st} — ошибки этих средних.

Тогда формула коэффициента дивергенции приобретает вид

$$\text{КД} = \sqrt{\frac{\sum t^2}{N}},$$

где N — число признаков, по которым ведется сравнение.

7.1.3. Понятие о таксономическом коэффициенте. Известны попытки выведения формул, способствующих практическому определению особей в трудных случаях, когда принятие решения затрудняется трансгрессией (захождением) признаков у близких видов. Одна из первых попыток такого рода, принятая Ф. Хейнке еще в 1898 г., основана на использовании нормированного отклонения. Предложенная им формула может быть записана следующим образом:

$$\sum \left(\frac{x - M_i}{\sigma_i} \right)^2 \rightarrow \min,$$

где x — индивидуальное значение признака определяемой особи; M_i и σ_i — средняя арифметическая и среднее квадратическое отклонение признака того вида, принадлежность особи к которому определяется. Сравнивая результаты вычислений, определяемую особь относят к тому виду, по отношению к которому формула Хейнке дает наименьшее значение.

Рассмотрим простой произвольный пример. Требуется определить принадлежность определяемой особи к одному из двух видов, средние арифметические и средние квадратические отклонения признаков которых заранее известны (табл. 54).

Подставляя нужные значения в формулу Хейнке, имеем

$$\text{I. } \left(\frac{12 - 10}{2} \right)^2 + \left(\frac{47 - 40}{4} \right)^2 + \left(\frac{5 - 4}{0,5} \right)^2 + \left(\frac{175 - 120}{20} \right)^2 = 15,62;$$

$$\text{II. } \left(\frac{12 - 15}{3} \right)^2 + \left(\frac{47 - 50}{4} \right)^2 + \left(\frac{5 - 7}{1} \right)^2 + \left(\frac{175 - 200}{40} \right)^2 = 6,19,$$

и в соответствии с «правилом Хейнке» определяемую особь следует отнести ко второму виду.

Впоследствии выяснилось, однако, что рассматриваемый прием приводит иногда к явно ошибочному результату, заставляя отнести особь, видовой принадлежность которой точно известна, к другому виду. Интерес к формуле Хейнке был потерян. Такая же участь постигла «таксономический коэффициент», предложенный в 1923 г. Г. Г. Виттенбергом в аналогичных целях.

Разгадка указанного явления принадлежит П. В. Терентьеву [1959]. Оказалось, что формула Хейнке дает правильный результат лишь тогда, когда в нее включаются слабо или вовсе не коррелирующие друг с другом признаки. В противном

Таблица 54. Произвольный пример вычисления таксономического коэффициента

Признаки	Средние арифметические (M) и средние квадратические отклонения (σ) признаков видов				Значения признаков (x) у определяемой особи
	I		II		
	M_1	σ_1	M_2	σ_2	
A	10	2	15	3	12
B	40	4	50	4	47
C	4	0,5	7	1	5
D	120	20	200	40	175

случае удельный вес коррелирующих признаков непомерно возрастает, что и приводит к неверному выводу. Лучше всего пользоваться «признаками-индикаторами» корреляционных плеяд (см. раздел 7.3.2), связь между которыми практически равна нулю.

Таким образом, использование формулы Хейнке и других подобного рода «таксономических коэффициентов» возможно лишь на основе некоррелирующих признаков. Необходимость выбора таких признаков путем предварительного изучения корреляционной структуры видов ограничивает практическое значение рассматриваемого приема.

7.2. ТАКСОНОМИЧЕСКИЙ АНАЛИЗ

Понятие о систематике как науке, занимающейся изучением всего многообразия живых существ, является весьма широким. Оно включает в себя общее описание организмов, их классификацию, построение и объяснение системы. Та часть систематики, которая разрабатывает принципы, методы и правила классификации и которая призвана стать теоретической осно-

вой последней, называется таксономией [Тахтаджян, 1966; Бейли, 1970].

Важнейшей задачей таксономии, от решения которой зависит построение классификационных схем, является изучение структуры таксонов и структуры их связей (взаимоотношений) в пределах таксона следующего ранга.

Существуют два аспекта в решении указанной задачи, требующие применения разных математических методов. Первый аспект состоит в изучении и сравнении структуры собственно признаков таксонов. Относящиеся сюда математические методы (которым и посвящен настоящий раздел) мы объединяем под общим названием таксономического анализа. Второй аспект, заключающийся в анализе и сравнении структуры корреляционных связей между признаками таксонов и между самими таксонами, будет рассмотрен в разделе 7.3.

Изучение структуры признаков и структуры связей между ними преследует в конечном счете одну и ту же таксономическую цель, однако между этими двумя аспектами существует разница не только методического, но и принципиального свойства: в одном случае речь идет об изучении «внешней» (наблюдаемой), а в другом — «внутренней» (непосредственно не наблюдаемой) структуры признаков таксонов. Если учесть, что в процессе эволюции корреляционные связи между признаками изменяются медленнее, чем сами признаки, то становится ясным, что результаты анализа, выполненного в двух рассматриваемых нами аспектах, могут частично и не совпадать. Поэтому второй аспект логически дополняет первый, и наиболее эффективным в таксономической практике следует признать последовательное приложение к одному и тому же объекту методов таксономического анализа и методов анализа корреляционной структуры.

7.2.1. О двух направлениях таксономического анализа. Существующие и развивающиеся в настоящее время методы таксономического анализа можно подразделить на две основные группы, различающиеся не только математическим инструментарием, но прежде всего разными биологическими предпосылками.

Одну из таких групп составляют методы «числовой таксономии» (Numerical Taxonomy), основоположниками которой являются энтомолог Р. Сокэл и генетик-микробиолог П. Снит [Sokal, Sneath, 1963; Sneath, Sokal, 1973]. Это направление получило бурное развитие за рубежом, но не нашло особого признания в отечественной биологии. Поскольку в зарубежной литературе за минувшие годы появилось множество модификаций и частных приемов, в значительной мере видоизменивших первоначальную концепцию «числовой таксономии», имеет смысл обратиться к тем исходным установкам, на основе которых она была выработана.

Опорным пунктом «числовой таксономии» явился отвергнутый систематиками постулат старого французского ботаника Мишеля Адансона (1727—1806 гг.) о таксономической равноценности всех признаков организма.

Исходя из этого ошибочного утверждения¹, «числовая таксономия» (в своей ортодоксальной форме) считает идеальной такую классификацию, которая содержит максимум информации, достигаемой не качеством (т. е. не предварительным отбором), а количеством используемых признаков, и предполагает следующую схему действий таксономиста.

Исследователь имеет дело с ОТЕ — «Операбельными Таксономическими Единицами» (Operational Taxonomic Units, OTU), в качестве которых фигурируют таксоны самого низкого в данном анализе ранга (ими могут быть расы и даже особи). Единственное требование, предъявляемое к ОТЕ, — ее постоянство во всей исследуемой группе. Задачей работы является формальная группировка всех ОТЕ в единицы классификации более высокого ранга («феноны»), проявляющиеся на том или ином уровне сходства вне зависимости от каких бы то ни было филогенетических соображений.

Для выполнения этой задачи берется как можно больше (не менее 50—100, а чаще — более) признаков, производятся их измерения или подсчеты, после чего вся информация передается ЭВМ для вычисления коэффициентов сходства между ОТЕ и построения соответствующей дендрограммы, отражающей группировку ОТЕ в «феноны».

При всей логичности построений «числовой таксономии» ей свойственны как биологические, так и чисто математические недостатки. Первые из них, заключающиеся в «неоадансоновском» подходе к делу, уже отмечались нами. Отсюда же вытекает и требование к подключению в анализ возможно большего числа признаков. Странники ортодоксальной концепции «числовой таксономии» отрицают возможность объективного отбора тех признаков, которые следует включать в исследование. Ниже будет показано, что в действительности объективные критерии для «взвешивания» признаков существуют. Отказ от этой процедуры ведет к дебиологизации исходных данных.

Существует и еще один аспект постулата о равноценности всех признаков, который также не выдерживает критики. «Числовая таксономия» считает, что любой набор большого числа признаков равноценно представляет генотип. Эта «гипотеза неспецифичности признаков» допускает возможность по-

¹ Подавляющее большинство современных систематиков убеждено в том, что для качественной характеристики любого таксона необходимо и достаточно иметь сравнительно небольшой набор таксономически значимых (а не всяких!) признаков.

строения классификации на группе признаков любого отдела тела растения или животного, на основе признаков отдельных тканей, онтогенетических стадий и т. д., т. е. не принимает во внимание наличие корреляционных плеяд (раздел 7.3). На практике это открывает путь к построению системы на базе группы тесно коррелированных признаков, что следует признать недопустимым.¹

Что касается математических недостатков «числовой таксономии», то используемые ею многочисленные коэффициенты сходства дают разные результаты, на что обратил внимание один из ее авторов — Р. Сокэл [1968]. Существуют три типа таких коэффициентов: коэффициенты ассоциации, корреляции и расстояния, но Сокэл не отдает предпочтения ни одному из них.

Нет надобности приводить здесь все применяемые числовой таксономией коэффициенты сходства, которых насчитывается более двух десятков, но имеет смысл указать на то, что наиболее часто используемые ею коэффициенты ассоциации типа $S_1 = \frac{m}{n}$, $S_2 = \frac{m-u}{n}$, $S_3 = \frac{m}{n+u}$ и др., где m — число совпадающих признаков; u — число несовпадающих признаков; $n = m + u$ — общее число признаков, являются довольно грубыми невесовыми мерами таксономического сходства. К тому же одни из них учитывают, а другие не учитывают совпадение ОТЕ по отсутствию признаков — явление, на которое систематик не может не обращать внимания.

Оценивая ортодоксальные приемы «числовой таксономии», нельзя не прийти к выводу о том, что классификации, получаемые с ее помощью, являются чисто «фенетическими». Поначалу это не отрицалось и ее авторами, но впоследствии многочисленные сторонники этого направления за рубежом стали придавать получаемым «фенонам» филогенетический смысл.

Надо сказать, что уязвимость адансоновского постулата, положенного в основу «числовой таксономии», а также многочисленные критические замечания, поступившие в ее адрес, способствовали тому, что некоторые сторонники этого направления стали искать пути к предварительному таксономическому «взвешиванию» включаемых в анализ признаков. К настоящему времени уже предложено несколько способов «взвешивания» признаков, и очень может быть, что идя таким путем «числовая таксономия» со временем войдет в русло второго направления таксономического анализа, успешно развивающегося в отечественной систематике.

Это направление, основоположником которого явился Е. С. Смирнов [1960, 1969, и др.], зиждется на признании так-

¹ В действительности для построения классификации необходимо использовать набор возможно более независимых друг от друга признаков [Техтаджян, 1966, с. 42].

сономической неравноценности признаков и с самого начала приняло на вооружение идею «весовых» методов в таксономии.

Проблема «взвешивания» признаков имеет свои трудности. Если мы рассчитываем на получение классификации, отражающей истинное родство таксонов, то к определению веса признаков следовало бы подходить с точки зрения их филогенетической значимости, придавая больший вес тем признакам, которые больше говорят о родстве сопоставляемых объектов. По отношению к таксонам крупного ранга, такая процедура нередко представляется вполне выполнимой, но этого нельзя сказать о таксонах низкого ранга, в частности о видах — основных единицах систематики. К сожалению, мы не располагаем критериями филогенетической значимости подавляющего большинства видовых признаков.

Практически не остается ничего иного, как искать другие критерии, косвенным образом отражающие филогенетическую значимость признаков и позволяющие оценить их таксономический вес. Речь идет о том, что если нельзя или крайне трудно оценить филогенетическую значимость признаков как таковую, то можно и нужно пытаться искать следы исторического родства таксонов в соотношении и структуре их признаков.

При такой постановке вопроса открываются два возможных пути для оценки таксономического веса признаков: на основе изучения корреляций между ними и на основе изучения частоты встречаемости (фреквенции) самих признаков (например, у видов данного рода).

Первый путь, вообще говоря, является более глубоким средством познания, но в то же время он весьма трудоемок, требуя полного корреляционного анализа с выделением корреляционных плеяд и их признаков-индикаторов (см. раздел 7.3), так как именно последним, как наиболее независимым из всех признаков, следует придавать наибольший таксономический вес. Сложность этого пути и явилась причиной того, что в таксономическом анализе преобладают методы, основанные на изучении фреквенции признаков.

Два отечественных весовых метода таксономического анализа — метод Е. С. Смирнова [1960, 1969, 1971, и др.] и метод П. П. Гамбаряна [1964, и др.] — противоположным образом решают вопрос о способе «взвешивания» признаков на основе их фреквенции, используя при этом совершенно различный математический аппарат.

По методу П. П. Гамбаряна наибольший таксономический вес придается наиболее часто встречающемуся признаку. Мотивируется это тем, что, по мнению данного автора, признаки, общие всем видам рода, в таксономическом отношении имеют более высокий «уровень общности», нежели признаки редкие, свойственные только отдельным видам. Фреквенция призна-

ков в этом методе оценивается количеством информации, численно равной энтропии рассматриваемой системы. Как известно, энтропия есть мера неопределенности состояния системы и равна сумме произведений вероятностей возможных состояний системы, взятой с обратным знаком: $H(X) = -p \log_2 p + -q \log_2 q$, где готовые значения $-p \log_2 p$ и $-q \log_2 q$ берутся из специальной таблицы (Приложение X).

Совпадения видов по наличию (тезе) и по отсутствию (антитезе) признаков оцениваются по приведенной выше формуле, а несовпадение признаков не учитывается вовсе (т. е. получает нулевую оценку). Таксономическое отношение (показатель близости двух видов) рассчитывается затем как сумма энтропий совпадающих тез и антитез:

$$T_{xy} = \sum_k H(X),$$

где k — число совпадающих тез и антитез. Величина T_{xy} зависит от числа учитываемых признаков (n) и может варьировать в пределах $n-1 \geq T_{xy} \geq 0$.

В дальнейшем метод подвергся некоторой модификации, позволяющей включать в анализ признаки, имеющиеся только у части таксонов исследуемой группы. Для этого признаки «взвешиваются» произведением частот тезы (p) и антитезы (q) по формуле

$$w = \frac{2p_i q_j}{n},$$

где n — число признаков, а $i \neq j$. Тогда сходство двух видов определяется суммой таких весов:

$$T_{xy} = \sum_k w.$$

По методу Е. С. Смирнова, рассмотрению которого посвящен следующий раздел, наибольший таксономический вес придается, наоборот, самому редкому признаку. Тем самым задача классификации решается не путем сближения таксонов, обладающих совпадающими частыми признаками, а посредством выделения из общей массы тех таксонов, которым свойственны совпадающие редкие признаки.

Строгое математическое обоснование метода Е. С. Смирнова было дано П. П. Тамириным [1971а, б, 1972, 1973], на основе теории корреляций доказавшим наибольшую таксономическую ценность редких признаков. Редкие признаки точнее частых определяют фреквенцию остальных признаков, и поэтому на их основе достигается более четкая классификация. Они наименее коррелированы и по своей таксономической значимости близки к значимости признаков-индикаторов корреляционных плеяд (раздел 7.3). Наконец, с точки зрения тео-

рии информации редкие признаки также заслуживают придания им большего веса, ибо редкое (оригинальное) событие, будучи осуществившимся, дает гораздо больше информации, чем событие частое (банальное), осуществления которого явно можно было ожидать.

С иных позиций — с помощью булевых функций — метод Е. С. Смирнова был математически обоснован А. А. Дюльдиным [1973], а еще ранее включен в сводку алгоритмов автоматической классификации [Дорофеев, 1971].

Заметим также, что результаты применения методов П. П. Гамбаряна и Е. С. Смирнова в общем случае не совпадают, хотя иногда и приводят к близким выводам [Шмидт, 1970]. Более того, существует лишь одна теоретически возможная, но практически почти невероятная ситуация, при которой результаты использования двух методов оказываются идентичными: это может произойти только в том случае, если все учитываемые признаки имеют биеквимодальное распределение, т. е. обладают равными вероятностями тез и антитез ($p=q=0,5$), что ведет к одинаковым для всех признаков значениям $H(X)=1$.

Неоспоримые, на наш взгляд, преимущества метода Е. С. Смирнова диктуют необходимость более подробного его рассмотрения.

7.2.2. Метод таксономического анализа Е. С. Смирнова. Для уяснения сущности этого метода [Смирнов, 1960, 1969, 1971, и др.] необходимо обратиться к его первоначальному варианту, предусматривавшему включение в анализ только качественных признаков таксонов.

Е. С. Смирнов предложил различать *свойства* и *признаки*, считая вторые модальностями первых. Например, окраска венчика — это свойство (A), тогда как конкретное проявление (цвет венчика) — красный (A_1), белый (A_2), голубой (A_3) и т. д. — является признаком.

Допустим, что род содержит 10 видов, из которых 2 вида обладают красным, 7 видов — белым и 1 вид — голубым венчиком. Тогда общее распределение видов рода по признакам данного свойства можно записать в следующей форме: $2A_1 + 7A_2 + 1A_3 = 10$. Пусть далее мы принимаем в расчет еще одно свойство — форму листа (B), причем 4 вида обладают цельным (B_1), а 6 видов — лопастным (B_2) листом: $4B_1 + 6B_2 = 10$.

Распределение признаков свойства A является тримодальным, а распределение признаков свойства B — бимодальным. Частный случай бимодального распределения типа $nA_1 + nA_2 = 2n$ называется биеквимодальным. Могут иметь место и мультимодальные распределения признаков. Одномодальные распределения (все виды рода обладают одним и тем же признаком данного свойства) таксономического значения не имеют

и данным методом не учитываются, что соответствует обычной практике таксономических исследований.

Любое мультимодальное распределение признаков может быть выражено системой равносильных ему бимодальных распределений. В приведенном выше примере 2 вида из 10 обладают красным (A_1) венчиком. Это означает, что у остальных 8 видов рода венчик любого другого, но только не красного цвета, что можно обозначить символом a_1 . Точно так же 7 видов имеют белый (A_2), а остальные — не белый (a_2) венчик и т. д. В итоге исходное тримодальное распределение можно разложить на систему из трех бимодальных распределений:

$$2A_1 + 8a_1 = 10,$$

$$7A_2 + 3a_2 = 10,$$

$$1A_3 + 9a_3 = 10.$$

В этой системе признаки, имеющиеся у вида и обозначенные заглавными буквами, являются положительными, а признаки, отсутствующие у вида и обозначенные строчными буквами, — отрицательными.

Приведенное ранее гипотетическое бимодальное распределение свойства B (форма листа) также можно выразить системой

$$4B_1 + 6b_1 = 10,$$

$$6B_2 + 4b_2 = 10.$$

Таким образом, число получаемых с помощью понятия об отрицательных признаках бимодальных распределений равно числу модальностей признака данного свойства.

Представление об отрицательных признаках имеет важное значение как с математической, так и с биологической точки зрения. Последнее ясно хотя бы потому, что отсутствие признака (отрицательный признак) нередко характеризует таксон столь же ярко, как и его наличие (положительный признак). Представление об отрицательных признаках позволяет математически четко характеризовать вид. Например, если в нашем произвольном примере какой-либо вид обладает положительными признаками A_1 и B_1 , то его характеристику следует записать так: $A_1a_2a_3B_1b_2$ (венчик красный, не белый, не голубой; лист цельный, не лопастной).

Такой прямой способ кодировки признаков возможен потому, что чаще всего положительные и отрицательные признаки не могут быть присущи вместе одному и тому же виду: наличие A_1 есть одновременно отсутствие a_1 и присутствие a_2 и a_3 и т. д.¹

¹ Наличие внутривидового полиморфизма или градиента изменчивости требует особой кодировки признаков, рассматриваемой ниже.

Допустим теперь, что какие-либо два вида (x и y) из сравниваемых нами 10 видов предполагаемого рода имеют следующее распределение признаков и перед нами стоит задача оценки степени их таксономического сходства (угловая скобка символизирует акт сравнения видов):

$$\begin{array}{l} x: A_1 a_2 a_3 B_1 b_2 \\ y: a_1 A_2 a_3 B_1 b_2 \end{array} >$$

Как видно, эти виды совпадают по одним (как положительным, так и отрицательным) признакам и не совпадают по другим, что обычно и наблюдается в действительности.

Признаки, по которым виды совпадают, неравноценны в таксономическом отношении. Вспомним, что в нашем произвольном примере признак A_2 достаточно банален, встречаясь у 7 видов из 10, тогда как признак A_1 редок, а признак A_3 специфичен для одного вида. Следовательно, факту совпадения этих признаков у сравниваемых видов необходимо придать разный таксономический вес. Логично исходить при этом из следующих соображений. Чем реже встречается признак, тем ярче он характеризует обладающий им вид, тем информативнее он в таксономическом отношении. Наоборот, чем чаще встречается признак, тем меньше он выделяет своего обладателя среди других видов (в предельном случае — когда все виды рода обладают этим признаком — он теряет свое значение для разграничения видов и переходит в разряд родовых признаков). Совпадение двух видов по редкому признаку больше говорит об их сходстве, нежели совпадение по признаку банальному, частому. Если же виды совпадают по ряду редких признаков, то это обстоятельство в определенной мере может рассматриваться как проявление родства, а не просто внешнего сходства, которое иногда является результатом конвергенции.

Таксономические веса совпадения признаков (ω), по методу Е. С. Смирнова, определяются как обратные отношения частот встречаемости (фреквенций) положительных и отрицательных признаков в бимодальных распределениях, к которым, как уже говорилось, предварительно сводятся все исходные распределения. Если имеется, например, распределение $2A_1 + 8a_1 = 10$, то таксономический вес совпадения двух видов по признаку A_1 составит $\omega_{A_1, A_1} = 8/2 = 4$, а по признаку a_1 — $\omega_{a_1, a_1} = 2/8 = 1/4$.

Вернувшись к нашему примеру с окраской венчика (A) и формой листа (B), из приведенных ранее систем бимодальных распределений признаков этих свойств (с. 161) получаем следующие таксономические веса совпадения этих признаков у сравниваемых видов:

$$2A_1 + 8a_1 = 10, \quad \omega_{A_1, A_1} = 8/2 = 4, \quad \omega_{a_1, a_1} = 2/8 = 1/4,$$

$$\begin{aligned}
7A_2 + 3a_2 &= 10, & \omega_{A_2A_2} &= 3/7, & \omega_{a_2a_2} &= 7/3, \\
1A_3 + 9a_3 &= 10, & \omega_{A_3A_3} &= 9, & \omega_{a_3a_3} &= 1/9, \\
4B_1 + 6b_1 &= 10, & \omega_{B_1B_1} &= 6/4 = 3/2, & \omega_{b_1b_1} &= 4/6 = 2/3, \\
6B_2 + 4b_2 &= 10, & \omega_{B_2B_2} &= 4/6 = 2/3, & \omega_{b_2b_2} &= 6/4 = 3/2.
\end{aligned}$$

Указанным способом оценивается только совпадение признаков. Однако на величину сходства двух видов должны влиять и несовпадающие у них признаки. По ряду соображений, вытекающих из теории вероятностей, таксономический вес несовпадения любых признаков оценивается отрицательной единицей ($\omega = -1$).¹

Исходя из приведенных выше приемов оценки таксономических весов совпадающих и несовпадающих признаков, Е. С. Смирнов предложил в качестве коэффициента сходства двух видов использовать значение среднего веса, назвав его таксономическим отношением:

$$t_{xy} = \frac{1}{n} \sum_i \omega_i,$$

где n — число признаков; ω_i — таксономические веса совпадающих и несовпадающих признаков двух видов.

В рассматриваемом нами произвольном примере это дает

$$x: A_1 a_2 a_3 B_1 b_2 > t_{xy} = \frac{1}{5} \left(-1 - 1 + \frac{1}{9} + \frac{6}{4} + \frac{6}{4} \right) = +0,222.$$

Положительное значение t_{xy} свидетельствует о сходстве видов, отрицательное — об их различии, а абсолютная величина — о степени сходства или различия. Подобным образом таксономические отношения рассчитываются для всех пар видов и сводятся в матрицу, анализ которой приводит к расчленению рода на группы более сходных между собой видов.

Интересно, что существует и имеет реальный смысл таксономическое отношение вида к самому себе (например, t_{xx} или t_{yy}), причем для разных видов этот показатель оказывается различным. На первый взгляд это обстоятельство вызывает сомнение: казалось бы, любой вид абсолютно сходен с самим собой и соответствующий показатель во всех случаях должен быть равен единице. Но на самом деле в различии таксономических отношений видов к самим себе сказывается глубокий смысл «взвешивания» признаков: один вид может характеризоваться набором редких, а другой — банальных признаков. Таксономическое отношение вида к самому себе показывает степень оригинальности (специфичности) его в системе рода.

¹ Математическое доказательство справедливости такой оценки дано П. В. Тамариным [1971а].

Иными словами, это — показатель «информационной емкости» вида.

В справедливости этого нетрудно убедиться на нашем произвольном примере:

$$x: A_1 a_2 a_3 B_1 b_2 > t_{xx} = \frac{1}{5} \left(\frac{8}{2} + \frac{7}{3} + \frac{1}{9} + \frac{6}{4} + \frac{6}{4} \right) = +1,889;$$

$$y: a_1 A_2 a_3 B_1 b_2 > t_{yy} = \frac{1}{5} \left(\frac{2}{8} + \frac{3}{7} + \frac{1}{9} + \frac{6}{4} + \frac{6}{4} \right) = +0,758,$$

откуда следует, что вид x оригинальнее вида y . Это и понятно, ибо признак A_1 встречается в роде гораздо реже признака A_2 , тогда как по остальным признакам оба вида совпадают.

Мы видели, что вычисление таксономических отношений связано с суммированием обратных величин фреквенций совпадающих положительных и отрицательных признаков и оценкой каждого несовпадения отрицательной единицей. Если обозначить фреквенцию символом β_i , то ее наименьшим значением будет $\beta = 1$ (данным признаком обладает только один вид), а наибольшим — $\beta = s - 1$, где s — число видов (данным признаком обладают все виды рода, кроме одного): $s - 1 \geq \beta_i \geq 1$.

Это позволяет представить формулы таксономических отношений иным образом. Так, если в межвидовых сравнениях число совпадающих признаков равно f , то

$$t_{xy} = \frac{s}{n} \sum_f \left(\frac{1}{\beta_i} \right) - 1,$$

где s — число видов; n — общее число признаков. При сравнении вида с самим собой совпадают все n признаков, поэтому

$$t_{xx} = \frac{s}{n} \sum_n \left(\frac{1}{\beta_i} \right) - 1.$$

Поскольку в конкретном сравнении одинаковые значения фреквенций разных признаков могут встретиться каждое по несколько раз, то практически расчет таксономических отношений удобно вести по формуле

$$t = \frac{s}{n} \left(n_1 \frac{1}{1} + n_2 \frac{1}{2} + \dots + n_{s-1} \frac{1}{s-1} \right) - 1,$$

где s — число видов; n — число признаков; n_1, n_2, \dots, n_{s-1} — множители, показывающие, сколько раз встречается соответствующая фреквенция. Это выражение пригодно для расчета как t_{xy} , так и t_{xx} .

Проверим эту формулу на ранее рассмотренном произвольном примере, беря в качестве фреквенций совпадающих при-

знаков значения соответствующих коэффициентов из бимодальных распределений (с. 161):

$$\begin{matrix} x: A_1 a_2 a_3 B_1 b_2 \\ y: a_1 A_2 a_3 B_1 b_2 \end{matrix} > t_{xy} = \frac{10}{5} \left(\frac{1}{9} + 2 \cdot \frac{1}{4} \right) - 1 = 0,222;$$

$$\begin{matrix} x: A_1 a_2 a_3 B_1 b_2 \\ x: A_1 a_2 a_3 B_1 b_2 \end{matrix} > t_{xx} = \frac{10}{5} \left(\frac{1}{2} + \frac{1}{3} + \frac{1}{9} + 2 \cdot \frac{1}{4} \right) - 1 = 1,889;$$

$$\begin{matrix} y: a_1 A_2 a_3 B_1 b_2 \\ y: a_1 A_2 a_3 B_1 b_2 \end{matrix} > t_{yy} = \frac{10}{5} \left(\frac{1}{8} + \frac{1}{7} + \frac{1}{9} + 2 \cdot \frac{1}{4} \right) - 1 = 0,758.$$

Как видим, результаты тождественны полученным ранее, однако при большом числе учитываемых признаков предпочтительнее использовать последнюю формулу, значительно сокращающую в этом случае объем вычислительной работы. Ускорить расчеты можно также с помощью составленных П. В. Тamarin [1972] специальных весовых таблиц.

Расчитанные значения t_{xy} и t_{xx} являются величинами абсолютными, сопоставимыми только в рамках конкретного таксона (например, рода), ибо предельные значения, которые они могут принимать, зависят от числа соподчиненных ему единиц (например, видов):

$$\frac{s-2}{2} > t_{xy} \geq -1;$$

$$s-1 > t_{xx} \geq \frac{1}{s-1} \quad (\text{при } s \geq 3),$$

где s — число видов.

Если же требуется сопоставить таксономические отношения, полученные для родов, содержащих разное число видов, то значения t_{xy} и t_{xx} следует перевести в относительные значения T_{xy} и T_{xx} по следующим формулам:

$$T_{xy} = \frac{2(t_{xy} + 1)}{s} \quad \text{и} \quad T_{xx} = \frac{t_{xx}(s - 1)}{s(s - 2)}.$$

Мы рассмотрели основные принципы таксономического анализа по методу Е. С. Смирнова в предположении, что качественные признаки однозначно характеризуют таксоны, каждому из которых присуща лишь одна модальность того или иного свойства. На практике ситуация нередко оказывается более сложной. Например, может иметь место полиморфизм, выражающийся в том, что таксон обладает набором модальностей одного свойства (скажем, белым, розовым и красным венчиком). В других случаях признаки могут обнаруживать заметный градиент изменчивости. Наконец, крайне желательно найти возможность включения в таксономический анализ количественных признаков, варьирующих по непрерывному типу.

Перечисленные выше случаи требуют особого кодирования признаков [Смирнов, 1971], что мы и покажем на соответствующих примерах.

Кодирование признаков в случае полиморфизма. Допустим, что в роде 5 видов, один из которых обладает венчиком только белого цвета (A_1), другой — розового (A_2) и красного (A_3), третий вид имеет белые (A_1) и голубые (A_4) венчики, четвертый — белые (A_1), розовые (A_2) и красные (A_3), а пятый — голубые (A_4) и синие (A_5). Тогда исходное полимодальное распределение может быть записано следующим образом: $1A_1 + 1A_{2,3} + 1A_{1,4} + 1A_{1,2,3} + 1A_{4,5} = 5$.

Введя отрицательные признаки (a_1 — венчик не белый, a_2 — венчик не розовый и т. д.), получаем полные характеристики (I) каждого из 5 видов по признакам данного свойства и, просматривая вертикальные ряды символов, легко находим бимодальные распределения (II), необходимые для расчета таксономических весов совпадения признаков (который производится затем обычным способом):

I	II
1. $A_1a_2a_3a_4a_5$	$3A_1 + 2a_1 = 5$
2. $a_1A_2A_3a_4a_5$	$2A_2 + 3a_2 = 5$
3. $A_1a_2a_3A_4a_5$	$2A_3 + 3a_3 = 5$
4. $A_1A_2A_3a_4a_5$	$2A_4 + 3a_4 = 5$
5. $a_1a_2a_3A_4A_5$	$1A_5 + 4a_5 = 5$

Кодирование признаков при наличии градиента изменчивости. Качественные или дискретно варьирующие количественные признаки иногда позволяют расположить виды рода в порядке убывания или нарастания степени выраженности соответствующего признака. Предположим, что у первого из 5 видов рода сложный лист состоит из 3 (A_1), у второго — из 4 (A_2), у третьего — из 5 (A_3), у четвертого — из 6 (A_4) и у пятого — из 7 (A_5) пар листочков, т. е. исходное распределение можно записать в форме выражения $1A_1 + 1A_2 + 1A_3 + 1A_4 + 1A_5 = 5$.

Обычный способ кодирования приводит здесь к пяти бимодальным распределениям типа $1A_1 + 4a_1 = 5$, в результате чего все межвидовые таксономические отношения оказываются одинаковыми ($t_{xy} = -0,25$). Поскольку мы наблюдаем ясно выраженный градиент, такой результат явно неудовлетворителен: например, первый вид (3 пары листочков) по данному признаку должен быть ближе к второму (4 пары листочков), чем, скажем, к пятому виду (7 пар листочков).

Выход из положения заключается в том, что кодирование и переход к бимодальным распределениям следует производить с помощью шкалы «не менее такого-то признака». В нашем

примере эта шкала выглядит следующим образом: A_1 — не менее 3 пар листочков, A_2 — не менее 4 пар, A_3 — не менее 5 пар, A_4 — не менее 6 пар и A_5 — не менее 7 пар. При таком подходе каждый последующий вид оказывается обладателем признаков всех предыдущих видов, причем признак A_1 , свойственный всем 5 видам рода, из анализа исключается:

I	II
1. $a_2a_3a_4a_5$	$4A_2 + 1a_2 = 5$
2. $A_2a_3a_4a_5$	$3A_3 + 2a_3 = 5$
3. $A_2A_3a_4a_5$	$2A_4 + 3a_4 = 5$
4. $A_2A_3A_4a_5$	$1A_5 + 4a_5 = 5$
5. $A_2A_3A_4A_5$	

Как видно, в совместном распределении (I) виды располагаются в порядке нарастания признака, что и находит свое отражение как в бимодальных распределениях (II) и рассчитанных на их основе таксономических весах, так и в значениях t_{xy} : сходство каждого вида с соседними оказывается наибольшим и убывает по отношению к более удаленным. Например, здесь: $t_{12} = +0,35$, но $t_{13} = -0,27$, $t_{14} = -0,69$, а $t_{15} = -1,00$. Следовательно, наблюдаемый градиент учтен особым способом кодирования признаков.

Кодирование признаков при одновременном наличии полиморфизма и градиента. Предположим, что по характеру опушения стебля сравниваются 5 видов. Все особи первого вида снабжены только простыми волосками (R_1). Среди экземпляров второго вида попадаются как опушенные лишь простыми волосками (R_1) растения, так и обладающие небольшой примесью железистых волосков (R_2). У третьего вида также встречаются оба варианта (R_1 и R_2), но добавляется еще и третий: в опушении некоторых растений железистые волоски явно преобладают (R_3). Четвертый вид характеризуется вариантами R_2 и R_3 . Наконец, все особи пятого вида опушены исключительно железистыми волосками (R_4). Для удобства указанные соотношения можно изобразить в виде схемы, из которой видно, что наряду с внутривидовым полиморфизмом (виды 2, 3 и 4) среди видов рода наблюдается градиент перехода от простого опушения стебля к железистому (виды 1—5):

1. R_1 — — —
2. R_1R_2 — —
3. $R_1R_2R_3$ —
4. $-R_2R_3$ —
5. — — — R_4

В такой ситуации необходимо объединить те приемы кодирования признаков, которые выше были рассмотрены в отдельности для случаев наличия полиморфизма и градиента.

Градиент можно отразить, кодируя признаки по следующей схеме: A — присутствие простых волосков, B — преобладание простых волосков, C — наличие только простых волосков, D — присутствие железистых волосков, E — преобладание железистых волосков, F — наличие только железистых волосков.

Вводя, как обычно, отрицательные признаки (например, a — отсутствие простых волосков и т. д.) и учитывая все то, что говорилось ранее о кодировании признаков в случае полиморфизма (например, наличие признака B автоматически означает и наличие A ; наличие C есть одновременно наличие A и B , и т. д.), получаем характеристики всех 5 видов по рассматриваемому признаку (I), на основании которых составим бимодальные распределения (II) для последующего расчета таксономических видов обычным способом:

I	II
1. $ABCdef$	$4A + 1a = 5$
2. $ABCDef$	$4B + 1b = 5$
3. $ABCDEf$	$3C + 2c = 5$
4. $ABcDEf$	$4D + 1d = 5$
5. $abcDEF$	$3E + 2e = 5$
	$1F + 4f = 5$

Кодирование количественных признаков, варьирующих по непрерывному типу. При непрерывном типе изменчивости признака мы имеем дело со сгруппированным в несколько классов вариационным рядом (раздел 1.4.3), который не может быть включен в таксономический анализ непосредственно. Необходимо найти способ «перевода» такого количественного признака в разряд качественных.

Наиболее простое решение вопроса состоит в рассечении вариационного ряда на две части по принципу «много — мало» в зависимости от положения средней арифметической (M), рассчитанной по данному признаку для рода в целом. Тогда для перехода к бимодальному распределению достаточно установить, превышает ли средняя арифметическая каждого вида родовую среднюю («много») или не достигает ее («мало»). Разумеется, при таком подходе некоторая часть информации об изменчивости признака теряется, однако это компенсируется возможностью подключения к таксономическому анализу количественных признаков, варьирующих по непрерывному типу.

Покажем это на реальном примере с тремя видами и двумя экологическими формами рода *Odontites*. Из 17 учетных морфологических признаков (см. раздел 7.1.2, с. 150) 11 варьируют по непрерывному типу, а остальные 6 (*B, C, E, H, L, N*) — по дискретному, но образуют при этом в силу большой изменчивости столь длинные ряды, что кодирование их способом, описанным для случая полиморфизма, становится невозможным (число слагаемых при вычислении таксономических отношений непомерно возрастает). Поэтому принцип «много — мало» был использован для кодирования всех 17 признаков, что привело к следующей схеме:

l: *abcDefghijklmNOPQR*
 sl: *abcdefg h i k l m n o p q R*
 st: *ABCdEFGHi k l MNOPQr*
 sc: *ABCdEFGHi k L m N o p q r*
 v: *ABCDEFGHIKLMNOPQr*

где заглавные буквы означают превышение признака над родовой средней, а строчные — обратное соотношение.

Заметим, что роль отрицательных признаков в данном случае играют значения, не достигающие родовой средней (обозначенные строчными буквами), и поэтому из приведенной схемы мы сразу же получаем бимодальные распределения ($3A + 2a = 5$; $3B + 2b = 5$; $3C + 2c = 5$; $2D + 3d = 5$ и т. д.), на основе которых рассчитываются таксономические веса.

Предоставляя читателю ради упражнения произвести необходимые расчеты, приведем итоговую матрицу таксономических отношений (табл. 55), из которой можно извлечь несколько интересных выводов.

Среди видов наибольшей таксономической оригинальностью выделяется *O.litoralis* ($t_{11} = +1,181$) — морфологически наибо-

Таблица 55. Таксономические отношения видов и форм *Odontites*

Виды и формы	l	sl	st	sc	v	Σ
l	+1,181	+0,593	-0,534	-0,828	-0,412	0
sl	+0,593	+1,230	-0,632	-0,191	-1,000	0
st	-0,534	-0,632	+0,691	+0,152	+0,324	+0,001
sc	-0,828	-0,191	+0,152	+0,838	+0,029	0
v	-0,412	-1,000	+0,324	+0,029	+1,059	0
Σ	0	0	+0,001	0	0	+0,001

лее стабильный и экологически узкоспециализированный вид, произрастающий в условиях засоляемой и затопляемой береговой полосы Балтийского моря. Наименьшая таксономическая оригинальность свойственна *O.serotina* ($t_{st. st} = +0,691$) — виду, отличающемуся значительной изменчивостью морфологических признаков и широкой экологической амплитудой.

В результате естественного отбора заметно повышена таксономическая оригинальность экологических форм *O.serotina* ($t_{st. st} = +1,230$ и $t_{sc. sc} = +0,838$), причем это повышение пропорционально силе давления отбора, которая гораздо выше в условиях береговой полосы. Так, береговая форма *O.serotina*

Таблица 56. Шестибалльная логарифмическая шкала для перевода количественных признаков в качественные [по: Расницын, 1965]

Номер класса	Начало класса, %	Конец класса, %	Балл	Номер класса	Начало класса, %	Конец класса, %	Балл
I	0	3,125	1	IV	12,6	25,0	4
II	3,125	6,25	2	V	25,1	50,0	5
III	6,25	12,50	3	VI	50,1	100,0	6

Примечание. Начало и конец классов даны в процентах натуральных частот от объема выборки.

по анализируемому показателю оказывается ближе к *O.litoralis* ($t_{sl. l} = +0,593$), чем к исходной форме ($t_{sl. st} = -0,632$). Заметим, что конвергенция береговых форм *O.serotina* с типичным обитателем береговой полосы — видом *O.litoralis* явилась причиной путаницы в систематике этих видов. Что касается сорно-полевой формы *O.serotina*, то по причине более слабого давления отбора она, несмотря на известную конвергенцию с *O.verna* — сорняком озимых зерновых культур, остается более близкой к исходному типу ($t_{sc. st} = +0,152$ против $t_{sc. v} = +0,029$). Таким образом, экологические формы *O.serotina* дивергируют от типичной формы этого вида в «противоположные» стороны, на что указывают и знаки соответствующих таксономических отношений в табл. 55.

В целом таксономический анализ видов и форм *Odontites* существенно дополняет и развивает те выводы, которые ранее (с. 151) были сделаны путем расчета коэффициентов дивергенции (КД).

Однако, как уже говорилось, рассмотренный выше способ трансформации непрерывных количественных признаков в ка-

ественные ведет к потере части информации. Потеря информации уменьшится, если ряд разбить на большее число частей, сохранив, например, обычную группировку вариант в классы, выполняемую по известным правилам (раздел 1.4.3). Тогда каждому классу вариационного ряда надо будет придать определенный балл, т. е. перейти к дискретному распределению при-

Таблица 57. Перевод непрерывного количественного признака (ширина листа у видов и форм *Odontites*) в качественный при помощи логарифмической шкалы из табл. 56

Объекты	Классы и шифр признаков						n
	0,0—0,9	1,0—1,9	2,0—2,9	3,0—3,9	4,0—4,9	5,0—5,9	
	A	B	C	D	E	F	
l		18	59	53	15	5	150
sl	1	72	64	13			150
st		34	86	25	4	1	150
sc	4	43	56	34	11	2	150
v		7	42	52	36	13	150
Частоты, % от n							
l		12,0	39,4	35,3	10,0	3,3	100
sl	0,7	47,9	42,7	8,7			100
st		22,7	57,2	16,7	2,7	0,7	100
sc	2,7	28,7	37,3	22,7	7,3	1,3	100
v		4,7	28,0	34,6	24,0	8,7	100
Баллы частот по логарифмической шкале							
l	0	3	5	5	3	2	—
sl	1	5	5	3	0	0	—
st	0	4	6	4	1	1	—
sc	1	5	5	4	3	1	—
v	0	2	5	5	4	3	—

знака, кодирование которого особой трудности уже не составит. Вопрос заключается в том, каким образом переводить частоты классов в баллы?

С. П. Расницын [1965] предложил для этой цели шестибалльную логарифмическую шкалу (табл. 56), пользование которой мы поясним на примере такого непрерывного количественного признака, как ширина листа у видов и форм *Odontites*. Ширину листа будем рассматривать как свойство, имею-

щее столько признаков (A, B, C и т. д.), сколько классов имеют вариационные ряды. В табл. 57 приведены натуральные частоты, которые затем переведены в проценты от объема (n) выборки и, наконец, заменены баллами в соответствии с табл. 56.

Используя обозначения признаков с индексом соответствующего балла из табл. 57, можно составить следующее совместное распределение положительных признаков для сравниваемых видов и форм:

$$\begin{aligned} 1: & A_0 B_3 C_3 D_3 E_3 F_2 \\ sl: & A_1 B_3 C_3 D_3 E_0 F_0 \\ st: & A_0 B_4 C_6 D_4 E_1 F_1 \\ sc: & A_1 B_3 C_3 D_4 E_3 F_1 \\ v: & A_0 B_2 C_3 D_3 E_4 F_3 \end{aligned}$$

Дополняя каждый положительный признак соответствующими отрицательными признаками, приходим к схеме

$$\begin{aligned} 1: & A_0 a_1 b_2 B_3 b_4 b_5 C_3 c_6 d_3 d_4 D_5 e_0 e_1 E_3 e_4 f_0 f_1 F_2 f_3 \\ sl: & a_0 A_1 b_2 b_3 b_4 B_3 C_3 c_6 D_3 d_4 d_5 E_0 e_1 e_3 e_4 F_0 f_1 f_2 f_3 \\ st: & A_0 a_1 b_2 b_3 B_4 b_5 c_3 C_6 d_3 D_4 d_5 e_0 E_1 e_3 e_4 f_0 F_1 f_2 f_3 \\ sc: & a_0 A_1 b_2 b_3 b_4 B_3 C_3 c_6 d_3 D_4 d_5 e_0 e_1 E_3 e_4 f_0 F_1 f_2 f_3 \\ v: & A_0 a_1 B_2 b_3 b_4 b_5 C_3 c_6 d_3 d_4 D_3 e_0 e_1 e_3 E_4 f_0 f_1 f_2 F_3 \end{aligned}$$

Из вертикальных рядов символов этой схемы легко составить бимодальные распределения ($3A_0 + 2a_0 = 5$; $2A_1 + 3a_1 = 5$; $1B_2 + 4b_2 = 5$ и т. д.), на основании которых обычным путем рассчитываются таксономические веса.

Недостаток описанного способа перевода непрерывных количественных признаков в качественные заключается в том, что шестибалльная логарифмическая шкала носит искусственный характер. Кроме того, остается открытым вопрос, достаточно ли обычной группировки вариационного ряда для существенного уменьшения потери информации?

Разбивая вариационный ряд на 3, 4, 5, ..., n частей, мы получаем 2, 3, 4, ..., $n-1$ бимодальных распределений. По мере увеличения числа градаций объем вычислительной работы быстро возрастает, а количество восполняемой таким путем информации, наоборот, с каждым шагом уменьшается. Существует какой-то предел, по достижении которого дальнейшее «дробление» ряда теряет смысл, переставая окупаться сколько-нибудь существенным восполнением количества информации.

П. В. Тамариным [1978] математически доказано, что этот предел наступает при разбиении вариационного ряда на 20 ча-

стей, что дает 19 бимодальных распределений количественных признаков, кодировать которые рекомендуется по следующей схеме:

$$\begin{array}{l}
 Q_0 q_1 q_2 q_3 \dots q_{19} q_{20} \\
 Q_0 Q_1 q_2 q_3 \dots q_{19} q_{20} \\
 Q_0 Q_1 Q_2 q_3 \dots q_{19} q_{20} \\
 Q_0 Q_1 Q_2 Q_3 \dots q_{19} q_{20} \\
 \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\
 Q_0 Q_1 Q_2 Q_3 \dots Q_{19} q_{20},
 \end{array}$$

где большими буквами Q обозначено наличие качественного признака, а малыми буквами q — его отсутствие. При этом качественные признаки Q_0 и q_{20} являются родовыми (общими для всех видов) и из таксономического анализа исключаются.

Таким образом, для того чтобы гарантировать себя от потери информации при переводе непрерывного количественного признака в качественный, необходимо и достаточно разбить вариационный ряд на 20 классов, каждый из которых характеризует новый качественный признак с бимодальным распределением типа альтернативы «да — нет». Объем вычислительной работы при этом настолько велик, что даже при использовании нескольких количественных признаков обращение к ЭВМ становится неизбежным.

Заканчивая изложение метода таксономического анализа Е. С. Смирнова, обратим внимание на некоторые положения.

В любом варианте метода имеющаяся информация распределяется между анализируемыми таксонами в строгой зависимости от частоты признаков. Это приводит к тому, что между рассчитанными значениями t_{xy} и t_{xx} существуют определенные соотношения, которые могут быть использованы для проверки правильности вычислений.

Так, например, алгебраическая сумма t_{xy} и t_{xx} во всех строках и колонках итоговой матрицы (см. табл. 55) должна быть равна нулю (или незначительно отличаться от нуля за счет округлений).

Должны соблюдаться также следующие соотношения: $2\sum t_{xy} + \sum t_{xx} = 0$ и $\sum t_{xx} = s$, где s — число таксонов, включенных в анализ. Применительно к табл. 55 имеем, например: $2\sum t_{xy} = -4,999$, а $\sum t_{xx} = +4,998$ при $s=5$, т. е. вычисления верны. Средняя арифметическая из всех межвидовых таксономических отношений определяется величиной $\bar{t}_{xy} = -\frac{1}{s-1}$, а средняя из внутривидовых таксономических отношений составляет $\bar{t}_{xx} = 1$.

Метод таксономического анализа Е. С. Смирнова применим не только в систематике: его можно использовать для опре-

деления степени сходства между любыми объектами. Имеются многочисленные примеры удачного применения этого метода в экологии, физиологии, сравнительной флористике и даже в такой далекой от биологии отрасли, как наукометрия.

7.3. АНАЛИЗ КОРРЕЛЯЦИОННОЙ СТРУКТУРЫ ТАКСОНОВ

Корреляционная структура таксона представляет собой систему связей между всеми его признаками. Поскольку практически невозможно изучить даже большинство признаков объекта, мы будем употреблять термин «корреляционная структура» применительно к тому набору признаков, который включен в конкретное исследование. Это оправдано и потому, что опытный систематик выбирает для анализа наиболее важные с таксономической точки зрения признаки, структура связей между которыми является своеобразным «корреляционным скелетом» объекта и в главных чертах отражает его полную корреляционную структуру.

В свете того, что сказано о сущности корреляционной связи в разделе 4.1, каждый рассчитанный коэффициент корреляции может рассматриваться как некий новый признак таксона, выявленный посредством метода, обладающего большей разрешающей способностью, нежели чисто морфологический метод. Опыт показывает, что существуют корреляции, характеризующие таксономический ранг объекта (видовые, секционные, родовые и т. д.) и могущие в силу этого служить таксономическими признаками.

Следует обратить внимание и на то, что в ходе корреляционного анализа происходит своеобразное «размножение» признаков. Число возможных парных коэффициентов корреляции (или любых других показателей связи) резко возрастает с увеличением числа учитываемых признаков, определяясь формулой сочетаний $n(n-1) : 2$, где n — число признаков. Так, при $n=10$

Таблица 58. Схема полной корреляционной матрицы

Признаки	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	...	<i>n</i>
<i>a</i>	1	r_{ab}	r_{ac}	r_{ad}	...	r_{an}
<i>b</i>	r_{ab}	1	r_{bc}	r_{bd}	...	r_{bn}
<i>c</i>	r_{ac}	r_{bc}	1	r_{cd}	...	r_{cn}
<i>d</i>	r_{ad}	r_{bd}	r_{cd}	1	...	r_{dn}
...
<i>n</i>	r_{an}	r_{bn}	r_{cn}	r_{dn}	...	1

мы получаем 45 коэффициентов корреляции, но увеличив число признаков вдвое, можно рассчитать уже 190 показателей связи. Это обстоятельство, несомненно, способствует более полному познанию объектов и выяснению степени сходства между ними.

Изучение корреляционной структуры признаков таксона означает переход от исследования внешних особенностей к выявлению скрытых от непосредственного наблюдения внутренних взаимозависимостей.

Приступая к рассмотрению приемов анализа корреляционной структуры, мы будем исходить из того, что исследователем предварительно рассчитаны (см. главу 4) и сведены в корреляционную матрицу (табл. 58) все возможные для данного числа признаков коэффициенты корреляции (или иные показатели связи). Таким образом, анализ корреляционной структуры есть не что иное, как исследование системы связей, отражаемых полной матрицей парных коэффициентов корреляции всех изучаемых признаков.

7.3.1. Коэффициент дивергенции корреляций. Для оценки степени дивергенции таксонов по корреляционной структуре их признаков можно, по аналогии с профилем отклонения признаков от стандарта (см. рис. 28), построить профиль отклонения корреляций, причем нет надобности выражать их в сигмах стандарта, ибо коэффициент корреляции есть величина относительная. Такой профиль, несмотря на свою громоздкость, дает некоторое представление о дивергенции корреляционных структур.

Однако составление коэффициента дивергенции корреляций КДК по аналогии с коэффициентом дивергенции признаков КД (см. раздел 7.1.2) невозможно, ибо ряд отклонений корреляций от стандарта имеет (как и распределение самих коэффициентов корреляций) резко выраженную положительную асимметрию (рис. 29). В таких условиях среднее квадратическое отклонение как мерилlo изменчивости ряда утрачивает свое значение. Поэтому в основу описываемой методики [Шмидт, 1964] были положены следующие соображения.

Значение коэффициента дивергенции корреляций должно зависеть от двух причин: величины отклонения отдельных корреляционных связей от стандарта и числа таких отклонений.

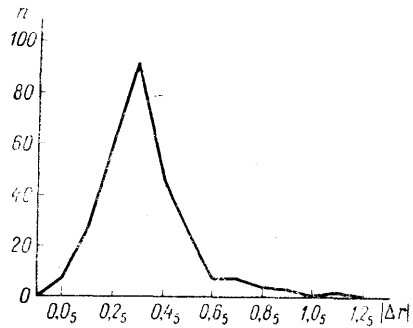


Рис. 29. Кривая распределения 274 достоверных отклонений корреляций от стандарта у видов и форм *Odontites*.

Действительно, одна и та же в среднем степень дивергенции корреляционных структур может быть достигнута как в случае сильных отклонений небольшого числа связей, так и в случае большого числа слабых отклонений.¹ Поэтому необходимо учитывать как величину, так и число отклонений корреляционных связей от стандарта. При этом логично принимать во внимание только статистически достоверные отклонения.

В качестве мерила величины достоверных отклонений корреляционных связей между парами одноименных признаков двух выборок естественно взять среднее отклонение:

$$\frac{\sum |\Delta_r|}{n},$$

где Δ_r — абсолютное значение достоверного отклонения, а n — число таких отклонений из общего количества (N) сопоставляемых корреляций.

Число достоверных отклонений (n) само по себе не дает представления о соотношении дивергирующих и остающихся без изменения связей, поэтому вторым членом искомого коэффициента (КДК) следует взять отношение $n:N$. Нижний предел обоих членов коэффициента равен нулю, но верхние пределы различны: максимальным значением второго члена служит единица (в предельном случае $n=N$), тогда как верхний предел первого члена зависит от типа коэффициента, используемого в конкретном анализе.

В случае применения коэффициентов связи, величина которых колеблется от 0 до 1 (корреляционное отношение η , коэффициент Чупрова T и др.), верхний предел величины $\sum |\Delta_r|:n$ равен единице, и тогда наш коэффициент имеет вид

$$\text{КДК} = \frac{\sum |\Delta_r|}{n} \cdot \frac{n}{N} = \frac{\sum |\Delta_r|}{N}.$$

Если же используются коэффициент линейной корреляции r , а также коэффициенты ранговой корреляции Спирмена ρ или Кендэла τ , варьирующие в пределах от -1 до $+1$, верхний предел величины $\sum |\Delta_r|:n$ повышается до 2 (в предельном случае сумма всех отклонений равна $2n$). В этом случае коэффициент дивергенции корреляции (КДК) следует рассчитывать по формуле

$$\text{КДК} = \frac{\sum |\Delta_r|}{2N}.$$

Нетрудно убедиться, что как в одном, так и в другом случае КДК варьирует в пределах от 0 до 1, равен нулю только при условии $n = \sum |\Delta_r| = 0$ и достигает единицы лишь при $n = N$

¹ Указанные случаи, возможно, характеризуют различные пути дивергенции корреляционных структур в ходе эволюции.

(последнее сопровождается дополнительным условием $\Sigma|\Delta_r| : n=1$ для первой формулы и $\Sigma|\Delta_r| : n=2$ — для второй).

Для приближенной оценки величины КДК может оказаться полезной номограмма (рис. 30), ось ординат которой отградуирована значениями $\Sigma|\Delta_r| : n$, а ось абсцисс — значениями $n : N$. Найдя точку пересечения этих параметров, следует, сообразуясь с ходом кривых номограммы, прочертить мысленную линию до пересечения с правой стороной графика и прочесть на ней соответствующее значение КДК.

Поясним использование КДК на примере с видами и формами *Odontites*. В нашем распоряжении имелись корреляционные матрицы признаков трех видов: *O.litoralis*, *O.verna* и *O.serotina* и двух экологических форм — береговой и сорно-полевой — последнего вида.

Процент достоверных отклонений корреляций в каждой выборке относительно любой другой колебался от 17 до 40%, причем сильных отклонений обнаружено было относительно мало. Пример вычисления суммы достоверных отклонений одной выборки относительно другой дан в табл. 59. В этом примере число коэффициентов корреляции (N) равнялось 136, а рассчитанная по табл. 59 величина $\Sigma|\Delta_r|=14,745$. В результате имеем

$$\text{КДК}_{\text{st.v}} = \frac{14,745}{2 \cdot 136} = 0,054.$$

Подобным же образом были рассчитаны и остальные десять возможных в данном сравнении значений КДК (табл. 60). Из этой таблицы можно извлечь интересные выводы.

Видно, например, что по своей корреляционной структуре экологические формы *O.serotina* дивергировали от этого вида слабее, нежели виды *O.litoralis* и *O.verna*. Заметно также, что они отклонились от исходного типа в противоположные стороны: экологические формы отличаются друг от друга сильнее ($\text{КДК}_{\text{sl. sc}}=0,046$), нежели каждая из них от *O.serotina* ($\text{КДК}_{\text{sl. st}}=0,043$; $\text{КДК}_{\text{sc. st}}=0,042$). Корреляционные структуры обладают гораздо большим консерватизмом по сравнению

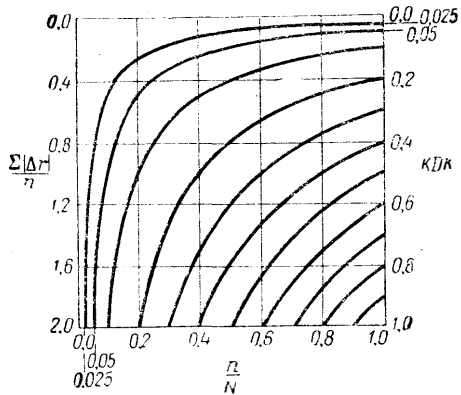


Рис. 30. Номограмма для определения величины КДК. Пояснения в тексте.

Таблица 59. Расчет суммы достоверных отклонений корреляций $|\Delta_r|$ между парами признаков *O. serotina* Dum. (st) и *O. verna* (Bell.) Dum. (v)

Признаки	r_{st}	r_v	$ \Delta_r $	Признаки	r_{st}	r_v	$ \Delta_r $
AB	0,099	0,366	0,267	GM	-0,125	-0,587	0,462
AK	0,416	0,785	0,369	GN	-0,224	-0,518	0,294
AN	0,591	0,698	0,197	HI	-0,570	0,578	1,148
BD	-0,240	0,292	0,532	HK	-0,056	0,472	0,528
BE	0,177	0,703	0,526	HL	-0,498	-0,290	0,298
BG	0,181	0,550	0,369	HM	-0,195	0,716	0,911
BH	0,597	-0,235	0,832	HN	-0,091	0,637	0,728
BI	-0,241	0,085	0,326	IK	0,564	0,777	0,213
BL	-0,128	0,268	0,396	IL	0,528	0,220	0,308
DH	-0,564	0,022	0,586	IM	0,464	0,658	0,194
DK	0,407	0,721	0,314	IN	0,367	0,638	0,241
EG	-0,322	0,210	0,532	KN	0,299	0,583	0,283
EH	0,246	-0,329	0,575	KO	0,368	-0,071	0,439
FH	-0,079	0,385	0,464	MN	0,872	0,917	0,045
FK	0,146	0,485	0,339	MO	0,226	-0,078	0,304
GH	-0,175	-0,622	0,447	MP	0,473	0,074	0,399
GJ	0,024	-0,390	0,414	PQ	0,794	0,668	0,126
GL	0,045	0,374	0,329				
				$\sum \Delta_r = 14,745$			

с морфологическими признаками: *O. litoralis* внешне куда сильнее отличается от *O. serotina*, чем *O. verna*, но значения КДК свидетельствуют об одинаковой дивергенции корреляционных структур этих видов от *O. serotina* ($KDK_{st, 1} = 0,055$; $KDK_{st, v} = 0,054$).

Таблица 60. Коэффициенты дивергенции корреляционных структур (КДК) видов и форм *Odontites*

Виды и формы	l	sl	st	sc	v
l	0	0,069	0,055	0,028	0,052
sl	0,069	0	0,043	0,046	0,041
st	0,055	0,043	0	0,042	0,054
sc	0,028	0,046	0,042	0	0,067
v	0,052	0,041	0,054	0,067	0

В рассмотренном примере мы имели дело с близкими видами и экологическими формами одного из них. Поэтому небольшие значения КДК и слабое различие между ними вполне естественны. При сравнении корреляционных структур резко обособленных видов (а тем более таксонов более высокого ранга) значения КДК будут значительно выше, отражая тем самым более длительный путь дивергенции структуры корреляционных связей этих категорий.

Описанный метод успешно использован Л. П. Рухадзе [1968] при сравнении корреляционной структуры признаков семян ископаемых и современных сортов винограда, а также М. Л. Аристарховой и Л. К. Иванюкович [1976] при сравнении корреляционной структуры количественных признаков культурных видов сорго.

7.3.2. Метод корреляционных плеяд
П. В. Терентьева. Этот метод, предложенный П. В. Терентьевым еще в 1931 г., впоследствии был модернизирован тем же автором [Терентьев, 1959, 1960].

В основу метода положен тот факт, что признаки организма связаны друг с другом корреляциями не хаотично, а образуют некоторые скопления, группы. Признаки, входящие в одну группу, заметно сильнее связаны друг с другом, нежели с признаками других групп. На это явление впервые обратил внимание Е. С. Смирнов [1923, 1924], назвавший такие группы признаков «конгрегациями» и отметивший, что внутри каждой группы имеется «признак-индикатор», который в среднем сильнее всех связан с остальными признаками своей группы. П. В. Терентьев для обозначения того же явления ввел термин «корреляционные плеяды», который и прижился в биологической литературе.

Итак, корреляционная плеяда — это группа признаков, связанных друг с другом сильнее, нежели с признаками других плеяд. Центром плеяды является признак-индикатор, как бы объединяющий вокруг себя другие признаки. Он служит «полномочным представителем» своей плеяды еще и потому, что связь между признаками-индикаторами разных плеяд крайне слаба или отсутствует вовсе. Именно поэтому относительно небольшой набор практически независимых друг от друга признаков-индикаторов ряда корреляционных плеяд наилучшим образом характеризует таксон в качественном отношении.

Обратимся к технике работы по выделению и анализу кор-

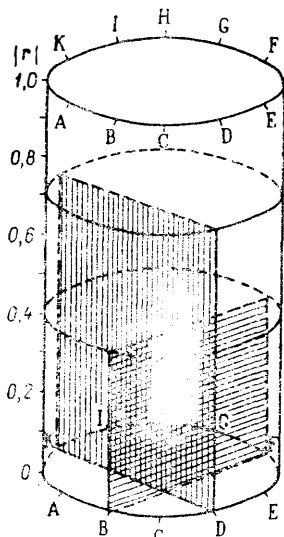


Рис. 31. Корреляционный цилиндр.

Пояснения в тексте.

реляционных плеед, полагая, что мы имеем перед собой полную корреляционную матрицу, и оговорив при этом, что знак связи данным методом не учитывается: нас интересует лишь абсолютная величина коэффициентов корреляции $|r|$, варьирующая от 0 до 1.

Суть метода удобно пояснить путем построения модели «корреляционного цилиндра» (рис. 31). Нижнее основание такого цилиндра градуируется шифрованными обозначениями признаков (A, B, C и т. д.), а боковая грань (вертикальная ось этой трехмерной модели) — нарастающими значениями $|r|$ от 0 до 1. Интервал произволен, но удобнее всего принять его равным 0, 1. Эти цифры будут символизировать уровни корреляции.

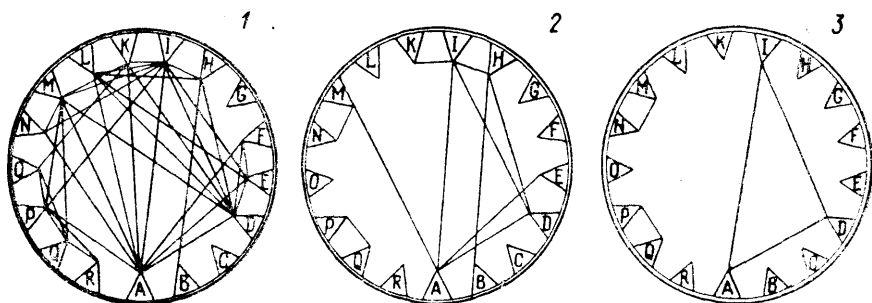


Рис. 32. Корреляционные кольца, отражающие структуру связей между признаками, и корреляционные плееды у *Odontites serotina* Dum.

1 — на уровне связи $|r| \geq 0,4$; 2 — при $|r| \geq 0,6$; 3 — при $|r| \geq 0,7$.

Ясно, что на нулевом уровне (нижнее основание цилиндра) все признаки связаны друг с другом (по условию связь не может быть меньше нуля). Это можно изобразить в виде линий, соединяющих все признаки, — получится «корреляционное кольцо» на уровне $|r|=0$. Для того чтобы показать на рассматриваемой модели связи на более высоких уровнях корреляции, следует возвести «заборчики» соответствующей высоты (рис. 31). След от них на поперечном сечении корреляционного цилиндра даст нам корреляционные кольца на разных уровнях связи. Практически такие сечения получают без построения цилиндра, перенося из корреляционной матрицы на заранее заготовленные кольца все связи, достигающие или превышающие заданный уровень корреляции (скажем, сначала для $|r| \geq 0,1$, затем для $|r| \geq 0,2$ и т. д.). Примеры корреляционных колец даны на рис. 32.

При «подъеме» по корреляционному цилиндру все большее число связей будет выпадать, и в конце концов признаки нач-

нут распадаться на плеяды (могут «отщепляться» и отдельные признаки). Плеяды могут вырисовываться на разных уровнях связи, но наиболее интенсивное их выделение происходит обычно при $|r| \geq 0,6-0,7$. На уровне $|r| \geq 0,9$ большинство плеяд распадается.

Описанный способ выделения корреляционных плеяд прост, но страдает некоторым субъективизмом, ибо выбор того или иного уровня связи для построения корреляционного кольца произволен и выделение плеяд происходит скачкообразно. Бо-

Таблица 61. Корреляционная матрица признаков *Odontites litoralis* Fr.

При- знаки	A	B	C	D	I	K	L	M	N	O	P	Q	R
A	—	62	73	79	85	74	51	86	51	31	42	35	36
B	62	—	32	17	54	47	56	30	24	4	25	16	11
C	73	32	—	6	2	9	13	12	9	8	1	2	5
D	79	17	6	—	71	67	40	63	33	36	43	33	39
I	85	54	2	71	—	78	59	70	44	28	44	34	31
K	74	47	9	67	78	—	45	67	49	18	40	34	36
L	51	56	13	40	59	45	—	43	38	11	24	18	13
M	86	30	12	63	70	67	43	—	71	36	44	44	44
N	51	24	9	33	44	49	38	71	—	7	19	23	19
O	31	4	8	36	28	18	11	36	7	—	38	35	20
P	42	25	1	43	44	40	24	44	19	38	—	71	59
Q	35	16	2	33	34	34	18	44	23	35	71	—	51
R	36	11	5	39		36	13	44	19	20	50	51	—

Примечание. Все значения коэффициентов корреляции для удобства умножены на 100, знак связи не учитывается.

лее дробная градуировка вертикальной оси корреляционного цилиндра смягчает указанный недостаток, но не устраняет его полностью.

Выход из положения был предложен С. Р. Вельдре [1964], показавшим, что предварительное построение дендрита способом «максимального корреляционного пути» по алгоритму Л. К. Выханду [1964] обеспечивает выделение плеяд именно на том уровне связи, на котором они впервые обозначаются при «подъеме» по корреляционному цилиндру.

Покажем выделение корреляционных плеяд этим способом на примере анализа корреляционной матрицы признаков *Odontites litoralis* (табл. 61).

Использование алгоритма «максимального корреляционного пути» продемонстрировано в табл. 62. Ход работы состоит в следующем. В корреляционной матрице (табл. 61) находим наибольшее значение коэффициента корреляции, каковым является $r_{AM}=86$. Переносим в табл. 62 всю строку A , в которой лежит этот коэффициент, снабжая при этом коэффициенты обозначениями тех признаков, связь между которыми они отражают. Отметив наибольший коэффициент, исключаем соответствующий столбец из дальнейшего анализа (делаем прочерки во всех клетках ниже).

Следующая строка табл. 62 должна быть озаглавлена признаком M , так как именно с ним обнаружена наибольшая связь

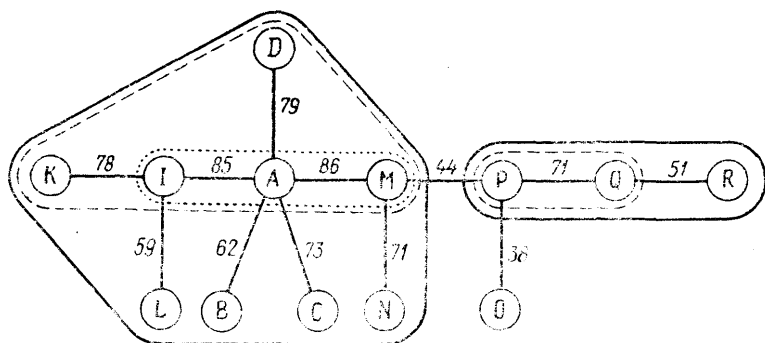


Рис. 33. Дендрит, построенный способом «максимального корреляционного пути» по данным табл. 62, и корреляционные плеяды разного порядка (ср. с данными табл. 64).

Цифры — значения коэффициентов корреляции (ноль целых и запятая опущены).

признака A в предыдущей строке. Сравниваем первую строку табл. 62 со строкой M в корреляционной матрице (табл. 61). Если в последней содержатся меньшие или равные значения коэффициентов, то во второй строке табл. 62 оставляем старые значения, а если более высокие — выписываем новые, например:

B	C	D	...	N	O	P
62	73	79	...	71	36	44
A	A	A	...	M	M	M

При этом не следует забывать, что изменение коэффициентов всякий раз сопровождается изменением обозначения соответствующих признаков.

Заполнив вторую строку табл. 62, находим в ней наибольшее значение коэффициента корреляции $r_{AI}=85$, отмечаем его,

Таблица 62. Алгоритм построения максимального корреляционного пути по данным табл. 61

Прин- таки	A	B	C	D	I	K	L	M	N	O	P	Q	R
A	—	B 62 A	C 73 A	D 79 A	I 85 A	K 74 A	L 51 A	M 83 A	N 51 A	O 31 A	P 42 A	Q 35 A	R 36 A
M	—	B 62 A	C 73 A	D 79 A	I 85 A	K 74 A	L 51 A	—	N 71 M	O 36 M	P 44 M	Q 44 M	R 44 M
I	—	B 62 A	C 73 A	D 79 A	—	K 78 I	L 59 I	—	N 71 M	O 36 M	P 44 M	Q 44 M	R 44 M
D	—	B 62 A	C 73 A	—	—	K 78 I	L 59 I	—	N 71 M	O 36 M	P 44 M	Q 44 M	R 44 M
K	—	B 62 A	C 73 A	—	—	—	L 59 I	—	N 71 M	O 36 M	P 44 M	Q 44 M	R 44 M
C	—	B 62 A	—	—	—	—	L 59 I	—	N 71 M	O 36 M	P 44 M	Q 44 M	R 44 M
N	—	B 62 A	—	—	—	—	L 59 I	—	—	O 36 M	P 44 M	Q 44 M	R 44 M
B	—	—	—	—	—	—	L 59 I	—	—	O 36 M	P 44 M	Q 44 M	R 44 M
L	—	—	—	—	—	—	—	—	—	O 36 M	P 44 M	Q 44 M	R 44 M
P	—	—	—	—	—	—	—	—	—	O 38 P	—	Q 71 P	R 59 P
Q	—	—	—	—	—	—	—	—	—	O 38 P	—	—	R 51 Q
R	—	—	—	—	—	—	—	—	—	O 38 P	—	—	—

ниже в столбце делаем прочерки и, озаглавив третью строку табл. 62 признаком I , сравниваем теперь вторую строку этой таблицы со строкой I в корреляционной матрице (табл. 61).

Описанный шаг алгоритма повторяется до тех пор, пока все признаки не будут исчерпаны. Затем из отмеченных в табл. 62 коэффициентов строим дендрит (рис. 33), в котором все признаки располагаются таким образом, что связи между ними являются наибольшими. Это и есть «максимальный корреляционный путь». Заметим, что такой дендрит может как угодно

Таблица 63. Выделение корреляционных плеяд и отщепление признаков при последовательном повышении уровня связи в дендрите (рис. 33)

Признаки	Уровень связи $ r $	Ход выделения плеяд
<i>OP</i>	0,39	ОПл $\rightarrow O$
<i>MP</i>	0,45	ОПл $\left\{ \begin{array}{l} \rightarrow \text{Пл } PQR \\ \rightarrow \text{Пл } ABCDIKLMN \end{array} \right.$
<i>QR</i>	0,52	Пл $PQR \rightarrow R$
<i>IL</i>	0,60	Пл $ABCDIKLMN \rightarrow L$
<i>AB</i>	0,63	Пл $ABCDIKMN \rightarrow B$
<i>PQ</i>	0,72	Пл $PQ \left\{ \begin{array}{l} \rightarrow \\ \rightarrow \end{array} \right.$
<i>MN</i>		Пл $ACDIKMN \rightarrow N$
<i>AC</i>	0,74	Пл $ACDIKM \rightarrow C$
<i>IK</i>	0,79	Пл $ADIKM \rightarrow K$
<i>AD</i>	0,80	Пл $ADIM \rightarrow D$
<i>AI</i>	0,86	Пл $AIM \rightarrow I$
<i>AM</i>	0,87	Пл $AM \left\{ \begin{array}{l} \rightarrow \\ \rightarrow \end{array} \right.$

Примечание. ОПл — общая плеяда, Пл — плеяда; стрелки показывают отщепление признаков, выделение плеяд и их распад на соответствующих уровнях связи.

ветвиться, но линии связи в нем не должны образовывать замкнутых циклов.

На \uparrow уровне минимальной в дендрите связи все признаки образуют общую корреляционную плеяду, которую в случае надобности легко воссоздать, построив недостающие на этом уровне достоверные связи. Если же «разорвать» в дендрите самую слабую связь, т. е. мысленно повесить минимальный уровень связи на единицу принятой точности расчета коэффициентов корреляции, то начинается распад дендрита на плеяды (или вычленяются отдельные признаки). Величина связи есте-

ственным образом характеризует тот уровень корреляции, на котором в действительности проявляются плеяды. Процесс повышения уровня связи следует продолжать, прослеживая последовательно выделение корреляционных плеяд и их распад на более высоких уровнях связи (табл. 63).

Выделенные корреляционные плеяды полезно проанализировать в нескольких аспектах.

Прежде всего обращают внимание на состав плеяд и на уровень связи, при котором они себя обнаруживают. Плеяды у разных видов, не отличаясь составом признаков, могут выделяться на разных уровнях связи, но может иметь место и обратное соотношение. Затем для каждой плеяды находят признак-индикатор. Он может быть установлен логическим путем (например, высота растения заведомо решающим образом влия-

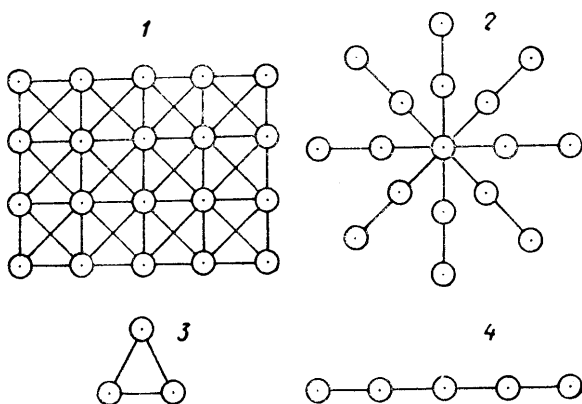


Рис. 34. Структурные типы корреляционных плеяд.
1 — «сеть»; 2 — «звезда»; 3 — «треугольник»; 4 — «цепь».

ет на другие признаки своей плеяды), но лучше определять его статистически как признак, в среднем наиболее сильно связанный со всеми другими признаками данной плеяды (для этого надо подсчитать и сравнить средние арифметические значения внутривидовых коэффициентов корреляции каждого признака плеяды).

Корреляционные плеяды можно классифицировать по их структуре. П. В. Терентьевым [1960] описаны три структурных типа плеяд: «цепь», «звезда» и «сеть» (рис. 34). К этому перечню следует добавить «треугольник», часто встречающийся у растений. «Сети» характерны для низких, а «цепи» — для высоких уровней связи. На практике нередко встречаются различные модификации и комбинации указанных основных структурных типов корреляционных плеяд.

Представляют интерес и некоторые другие показатели, по которым можно сравнивать пляды. Так, мощностью пляды (G) П. В. Терентьев назвал число входящих в нее признаков. С повышением уровня связи мощность пляд падает. Крепость пляды (D) — это среднее арифметическое из всех внутривидных коэффициентов корреляции. Наиболее крепкими обычно оказываются двучленные «цепи», а наименее крепкими — «сети». Крепость корреляционных пляд прямо пропорциональна

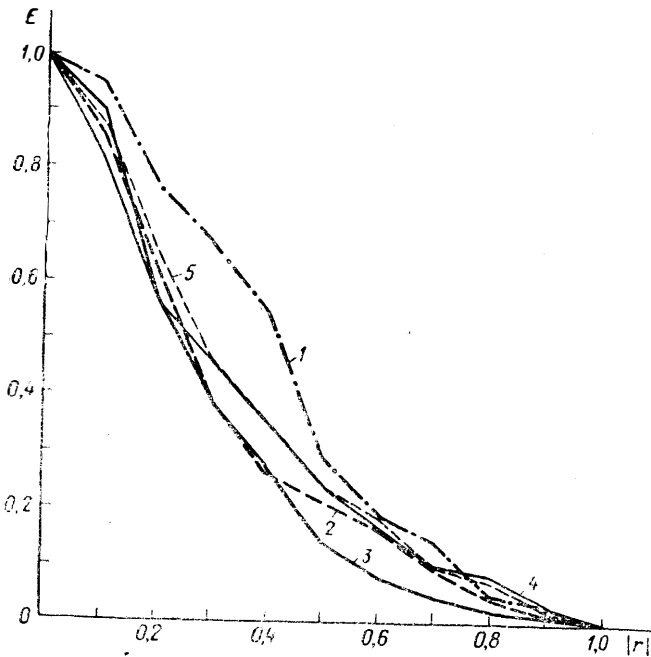


Рис. 35. Изменение коэффициента гомогенности (E) при «подъеме» по корреляционному цилиндру у видов и форм *Odontites*.

1 — *O. litoralis* Fr.; 2 — *O. verna* (Bell.) Dum.; 3 — *O. serotina* Dum.; 4 — береговая и 5 — сорно-полевая экологические формы последнего вида.

уровню связи и обратно пропорциональна их мощности. Коэффициент гомогенности (E) представляет собой отношение числа связей, выявленных на данном уровне корреляции (корреляционном кольце), к общему числу связей между признаками в данной выборке. Определяя значение E для каждого уровня, можно проследить изменение этого показателя при «подъеме» по корреляционному цилиндру. Получающийся при этом график отражает скорость выделения пляд, в известной мере характеризую степень интегрированности объекта. На рис. 35

приведен такой график для видов и экологических форм *Odontites*, из которого следует, что *O.litoralis* является наиболее, а *O.serotina* — наименее интегрированным объектом. Это соответствует морфологической стабильности и узкой экологической амплитуде первого вида и противоположным свойствам второго вида. Интересно также и то, что в ходе естественного отбора экологических форм *O.serotina* степень их интегрированности по сравнению с исходной формой повышается.

Из сравнения корреляционных плеяд признаков видов и экологических форм *Odontites* (табл. 64) следует, что часть признаков плеяд является общей для всех трех видов и должна быть отнесена в разряд родовых (или секционных), а другая часть характеризует специфику отдельных видов. Так, все виды обладают двумя общими плеядами (плеядой чашечки-2 и плеядой роста), которые выделяются на примерно одинаковом уровне связи, имеют почти одинаковую крепость и одни и те же признаки-индикаторы (соответственно длину чашечки и высоту растения). Но для *O.verna* специфично наличие плеяды стебля, а для *O.litoralis* — наличие плеяды чашечки-1 и отсутствие плеяды соцветия. Состав плеяды роста, включающей в себя обязательные признаки *A*, *D*, *I*, в остальном различен у разных видов. Что касается экологических форм *O.serotina*, то анализ корреляционных плеяд, с одной стороны, указывает на принадлежность этих форм к исходному виду, а с другой стороны, подчеркивает их дивергенцию в сторону других видов. Например, у береговой формы этого вида появляется плеяда чашечки-1, свойственная *O.litoralis*, и исчезает плеяда соцветия, сохраняющаяся у сорно-полевой формы, конвергирующей с *O.verna*.

В целом мы должны констатировать некоторое несоответствие между ярко выраженной конвергенцией внешнеморфологических признаков экологических форм *O.serotina* с двумя другими видами и относительно слабой конвергенцией корреляционной структуры тех же объектов. Эволюция корреляционных структур идет медленнее изменения внешних признаков и только путем одновременной перестройки лишь незначительного числа связей при сохранении стабильности «корреляционного скелета» организма, что устраняет опасность его внутренней дезинтеграции. К аналогичному выводу пришли К. М. Завадский и Э. И. Колчинский [1977].

Метод корреляционных плеяд П. В. Терентьева нашел широкое применение в самых различных отраслях биологии и медицины. Он дал прекрасные результаты во множестве работ по морфологии, систематике, экологии и географии растений и животных, в геоботанике, сравнительной флористике, селекции растений и при прогнозировании динамики численности популяций. Важным является указание А. Л. Тахтаджяна [1966] на то, что данный метод существенно способствует решению

трудных филогенетических проблем, связанных с явлениями параллелизма в эволюции цветковых растений.

В заключение этого раздела необходимо отметить два обстоятельства. Во-первых, опыт использования метода корреляционных плеяд показал, что в нем в качестве мерил связи могут применяться не только коэффициенты линейной корреляции (r), но и любые другие показатели связи или коэффициенты сходства. Во-вторых, если сам П. В. Терентьев [1960] считал свой метод принадлежащим к так называемой R -технике корреляционного анализа, изучающей корреляцию признаков особей в выборке, то теперь уже ясно, что этот метод с равным

Таблица 64. Сравнительная характеристика корреляционных

Плеяды	I				SI			
	Состав	Уровень	Мощность	Крепость	Состав	Уровень	Мощность	Крепость
Цветка					<i>OPQR</i>	0,4	4	0,46
Чашечки-1	<i>PQR</i>	0,5	3	0,57	<i>PQR</i>	0,5	3	0,60
Чашечки-2	<i>PQ</i>	0,6	2	0,71	<i>PQ</i>	0,7	2	0,89
Стебля								
Соцветия								
Роста	<i>ADIKM</i>	0,8	5	0,71	<i>AIMN</i>	0,8	4	0,86

правом можно относить и к Q -технике, занимающейся коррелированием особей или таксонов по ряду признаков, т. е. их группировкой. Все это сильно расширяет сферу применения метода корреляционных плеяд, который может использоваться во всех тех ситуациях, когда требуется исследовать внутреннюю систему связей любого биологического объекта.

7.3.3. Основные принципы факторного анализа. Факторный анализ [Лоули, Максвелл, 1967; Харман, 1972; Окунь, 1974], как и метод корреляционных плеяд, имеет задачей исследование системы корреляций, запечатленных в корреляционной матрице, но решает эту задачу иным образом, исходя из других методических соображений.

В методе корреляционных плеяд признаки группируются непосредственно по уровню связи друг с другом, а «центром тяжести» плеяды служит, как уже говорилось, реальный признак-индикатор. В отличие от этого в основу факторного анализа положено предположение, что согласованная изменчивость

группы признаков обусловлена их общей зависимостью от некоей скрытой (непосредственно не наблюдаемой) переменной, именуемой главной компонентой или фактором, вокруг которого и группируются признаки.

В роли фактора может выступать как один из изучаемых признаков, так и признак, не включенный в анализ. Им может быть даже показатель, не поддающийся прямому измерению, о существовании которого исследователю заранее ничего не известно. Биологическое истолкование сущности фактора выполняется на заключительном этапе факторного анализа и представляет собой иногда довольно трудную задачу. Подоб-

плеяд видов и экологических форм *Odontites*

st				sc				v			
Состав	Уровень	Мощность	Крепость	Состав	Уровень	Мощность	Крепость	Состав	Уровень	Мощность	Крепость
								<i>OPQR</i>	0,3	4	0,34
<i>PQ</i>	0,6	2	0,79	<i>PQ</i>	0,6	2	0,85	<i>PQ</i>	0,6	2	0,67
								<i>BE</i>	0,7	2	0,70
<i>MN</i>	0,7	2	0,87	<i>MN</i>	0,9	2	0,91	<i>MN</i>	0,8	2	0,92
<i>ADI</i>	0,7	3	0,73	<i>ADEIKM</i>	0,8	7	0,69	<i>ADIK</i>	0,8	4	0,74

но признакам-индикаторам корреляционных плеяд факторы друг с другом не коррелируют. В итоге факторного анализа исследуемая система корреляций подразделяется на части (факторные плеяды), каждая из которых обусловлена воздействием своего фактора.

Разработано несколько вариантов факторного анализа, предусматривающих использование коэффициентов линейной корреляции (r), ибо применение в этом методе нелинейных мер связи, например корреляционного отношения (η), наталкивается на значительные трудности. Наиболее употребительным вариантом факторного анализа является метод главных компонент, описываемый ниже. Используемый при этом математический аппарат довольно сложен, вычисления весьма трудоемки и все расчеты поэтому ведутся на ЭВМ, для чего существуют специально составленные программы.

Каждый фактор представляет собой линейную комбинацию изучаемых признаков: $F_j = \lambda_j (R_{j1}x_1 + R_{j2}x_2 + \dots + R_{ji}x_i + \dots +$

$+R_{jk}x_k$), где F_j — фактор ($j=1, 2, \dots, m$); x_i — признак ($i=1, 2, \dots, k$); R_{ji} — коэффициент корреляции j -го фактора с i -м признаком; λ_j — некоторый коэффициент. Приведенную выше формулу можно записать в сжатом виде:

$$F_j = \lambda_j \sum_{i=1}^k R_{ji} x_i.$$

С геометрической точки зрения величина R_{ji} представляет собой проекцию вектора i -го признака на ось j -го фактора. Эта величина получила название «факторной нагрузки». Для нахождения факторов и для их интерпретации используется еще одна величина — факторная дисперсия R_{ji}^2 (или D_F), определяющая ту долю дисперсии i -го признака, которая вызвана влиянием на него j -го фактора. Надо заметить, что, несмотря на независимость факторов друг от друга, один и тот же признак может испытывать влияние нескольких факторов.

Способ главных компонент исходит из того, что сумма влияний всех m факторов должна полностью исчерпывать дисперсию каждого признака и всей системы в целом: $\sum_{j=1}^m R_{ji}^2 = 1$.

Последовательность действий, выполняемых ЭВМ, выглядит следующим образом. Первая главная компонента (первый фактор F_1) определяется из полной исходной корреляционной матрицы. Для этого машина находит такую линейную комбинацию для k признаков, при которой величина $\sum_{j=1}^k R_{ji}^2$ оказывается максимальной, т. е. извлекает из полной матрицы максимум дисперсии. Затем все корреляции, обусловленные воздействием первого фактора, из исходной матрицы вычитаются, в результате чего получается остаточная матрица. Операция повторяется: для получения второго фактора (F_2) из остаточной матрицы снова извлекается максимальная дисперсия, после чего из этой матрицы вычитаются корреляции, обусловленные воздействием второго фактора, и т. д. — до тех пор, пока исходная матрица не окажется разделенной на m остаточных матриц, т. е. пока не будет выделено m факторов.

Практически довольствуются выделением 3—5 главных компонент (факторов). Дело в том, что объем информации, извлекаемый каждым последующим фактором, быстро убывает, а биологическая интерпретация факторов становится все более затруднительной. В обычных исследованиях можно ограничиться тем числом факторов, которое обеспечивает извлечение 80—90 % общей дисперсии признаков: гораздо важнее правильно истолковать смысл небольшого числа ведущих факторов, чем пытаться интерпретировать последующие факторы, исходя из ничтожной доли привносимой ими информации.

Из сказанного видно, что смысл способа главных компонент состоит в последовательном устранении влияния каждого выделенного фактора на систему связей между признаками. Это напоминает процедуру нахождения частных (парциальных) корреляций для пары признаков (раздел 4.2.2) с той, однако, принципиальной разницей, что в факторном анализе речь идет о системе корреляций между множеством признаков.

Бывают случаи, когда эта система достаточно «контрастна», т. е. значения коэффициентов корреляции в исходной матрице резко распадаются на «сильные» и «слабые». В таких случаях дисперсия каждого признака в значительной мере зависит от воздействия одного фактора, и получаемые в итоге факторные плеяды могут совпасть с корреляционными плеядами. Но чаще факторные плеяды, имея определенное сходство с корреляционными плеядами, отличаются от последних своим составом и строением. Во всяком случае факторный анализ всегда дает новую информацию, вскрывает новые особенности корреляционной структуры признаков и тем самым существенно дополняет метод корреляционных плеяд.

В качестве примера рассмотрим факторный анализ структуры признаков *O. litoralis*.¹ Исходная корреляционная матрица была приведена в табл. 61, а результаты факторного анализа сведены в табл. 65. Из этой таблицы видно, что первый фактор (F_1) определяет почти половину (44,9 %) общей дисперсии признаков. Заметим, что максимальные нагрузки по этому фактору имеют признаки *A*, *D*, *I*, *K* и *M*, определяющие размер растения и его основных органов. Это — фактор общих размеров растения. Со вторым фактором (F_2) сильнее других связаны генеративные признаки *O*, *P*, *Q* и *R*, и его можно назвать фактором размеров цветка. При этом обращает на себя внимание отрицательная связь с этим фактором числа междоузлий (признак *B*). С третьим фактором (F_3) сильно связан только один признак *C* — число укороченных междоузлий, и в соответствии с биологическим значением этого признака данный фактор следует считать фактором скороспелости. Четвертый фактор (F_4) определяет число цветков (признак *N*), а пятый (F_5) — длину венчика (признак *O*). В сумме первые пять факторов исчерпывают 80,6 % общей дисперсии признаков.

Графически факторную структуру признаков полезно представить в виде простой схемы (рис. 36, а). Однако более наглядным приемом выявления факторных плеяд является построение графика расположения признаков в пространстве двух или более факторов (последнее весьма затруднительно и применяется редко). Наибольшую информацию дают графики, отражающие положение признаков в системе I и II (иногда также I и III или II и III) факторов (рис. 36, б).

¹ Расчеты выполнены на ЭВМ М-222 Вычислительного центра ЛГУ по программе, составленной С. Ф. Колодяжным.

Таблица 65. Результаты анализа факторной структуры признаков *Odontites litoralis* Fr.

Признаки	F_1		F_2		F_3		F_4		F_5		$D_{F_1-F_5}$
	R	D_F	R	D_F	R	D_F	R	D_F	R	D_F	
A	+0,918	0,843	-0,170	0,029	-0,139	0,019	-0,062	0,004	+0,030	0,001	0,896
B	+0,548	0,300	-0,514	0,264	+0,396	0,157	-0,307	0,094	-0,156	0,024	0,840
C	+0,097	0,009	-0,431	0,186	+0,675	0,455	+0,221	0,049	+0,398	0,158	0,857
D	+0,781	0,610	+0,116	0,014	-0,336	0,113	-0,117	0,014	+0,027	0,001	0,751
I	+0,877	0,770	-0,166	0,028	-0,146	0,021	-0,195	0,038	-0,080	0,006	0,863
K	+0,826	0,682	-0,154	0,024	-0,111	0,012	+0,040	0,002	-0,110	0,012	0,732
L	+0,609	0,370	-0,410	0,168	+0,068	0,005	-0,253	0,064	-0,135	0,018	0,625
M	+0,869	0,756	-0,006	0,000	-0,134	0,018	+0,310	0,096	+0,209	0,044	0,913
N	+0,597	0,357	-0,230	0,053	-0,186	0,034	+0,587	0,344	+0,135	0,018	0,807
O	+0,399	0,159	+0,451	0,204	-0,056	0,003	-0,385	0,148	+0,635	0,403	0,917
P	+0,627	0,393	+0,500	0,250	+0,353	0,124	-0,090	0,008	-0,111	0,012	0,788
Q	+0,564	0,318	+0,548	0,300	+0,400	0,160	+0,090	0,008	-0,058	0,003	0,789
R	+0,520	0,271	+0,482	0,233	+0,172	0,020	+0,244	0,059	-0,331	0,110	0,702
D, %	44,9		13,5		8,9		7,1		6,2		89,6

Из этого рисунка видно, что признаки *O. litoralis* отчетливо группируются в три факторные плеяды, которые интересно сравнить с ранее полученными для того же объекта корреляционными плеядами (см. табл. 63 и 64; рис. 33).

Первая факторная плеяда (*ADIKM*) полностью совпадает с корреляционной плеядой роста. Вторая факторная плеяда (*OPQR*) по сравнению с соответствующей ей корреляционной плеядой чашечки-1 (см. табл. 64) содержит дополнительный признак (*O* — длина венчика): это свидетельствует о том, что корреляционный анализ не отразил той большей целостности

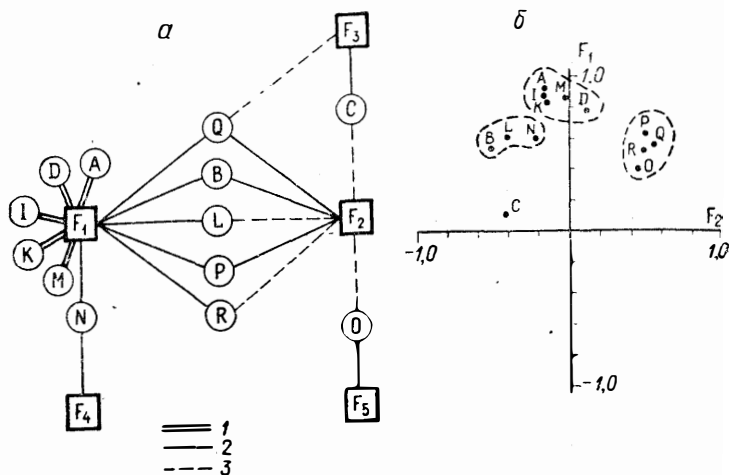


Рис. 36. Факторная структура признаков *Odontites litoralis* Fr.
 а — схема: 1 — $|R| \geq 0,7$; 2 — $0,5 < |R| < 0,7$; 3 — $0,4 < |R| < 0,5$; б — положение признаков и факторных плеяд в системе первых двух факторов (F_1 и F_2).

плеяды генеративных признаков, которая имеет место в действительности. Третья факторная плеяда (*BLN*) не имеет аналога среди выделенных ранее корреляционных плеяд признаков данного вида. Ее биологическая интерпретация несколько затруднительна, но обращает на себя внимание тот факт, что эта плеяда включает в себя «счетные» признаки (число междоузлий, число зубцов листа и число цветков); к ней же по фактору F_2 примыкает признак *C* — число укороченных междоузлий, чем и исчерпываются учтенные в данном анализе «счетные» признаки.

Таким образом, факторный анализ даже в рассмотренном частном примере расширил наши представления о корреляционной структуре признаков данного объекта. Следует добавить, что нередко факторный анализ дает и более ощутимые новые результаты.

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В СРАВНИТЕЛЬНОЙ ФЛОРИСТИКЕ

8.1. ОБЕСПЕЧЕНИЕ БИОЛОГО-СТАТИСТИЧЕСКОЙ СОПОСТАВИМОСТИ ФЛОР

Первейшим условием научного сравнения любых объектов является обеспечение их сопоставимости в той мере и в такой форме, которые диктуются задачами исследования. Применительно к сравнению флор обеспечение их биолого-статистической сопоставимости означает создание условий для корректного сравнения флористических показателей на основе одновременного соблюдения статистических и ботанико-географических требований.

Исторически сложилось так, что большинство флористических работ, содержащих важный фактический материал, затрагивает территории, весьма неравноценные либо по своим размерам, либо по принципу их выделения. Это обстоятельство, не снижая познавательной ценности флористических сводок, сильно затрудняет сравнение самих флор и выявление закономерностей изменчивости флористических показателей на поверхности земного шара. Поэтому проблема достижения биолого-статистической сопоставимости флор имеет важное значение в сравнительной флористике.

Вопросы сравнительного изучения количественных характеристик, в частности видового богатства флор и флористических областей, привлекали внимание исследователей уже в первой половине прошлого века. Основоположителем сравнительной флористики считается Альфонс Декандоль [De Candolle, 1855], впервые показавший, что флористическое богатство различных районов земного шара может и должно служить объектом сравнительно-флористического исследования. Декандоль придавал понятию равноценности флор топографический оттенок, указывая, что «было бы абсурдно сравнивать, с точки зрения общего числа видов, области, которые не являлись бы заметно одинаковыми по своей площади» [De Candolle, 1855, p. 1270].

А. И. Толмачевым [1941] было показано, что приблизительное равенство сравниваемых территорий является важным, но не единственным условием сопоставимости флор. Если эти территории велики, то в их различных частях местные условия среды по-разному влияют на богатство и состав флоры. При сокращении площади пестрота условий среды уменьшается и подлежащая изучению территория выглядит более однородной. Поэтому Толмачев сформулировал второе условие сопоставимости флор — ограниченность занимаемых ими территорий.

Третье условие, указанное тем же автором, требует высокой степени изученности сравниваемых флор. Надо только заметить, что чем мельче флора, тем легче выполняется это условие.

Итак, наиболее общие и основополагающие принципы количественного сравнения флор, обеспечивающие их сопоставимость, формулируются следующим образом:

1) приблизительное равенство площадей сравниваемых территорий;

2) ограниченный размер этих территорий;

3) высокая степень их флористической изученности.

По поводу первого принципа надо подчеркнуть, что на практике речь может идти, действительно, лишь о приблизительном равенстве сравниваемых территорий. Излишне жесткие в этом отношении требования технически трудно выполнимы и биологически мало оправданы, так как лишают нас возможности сравнения достаточно равноценных в ботанико-географическом отношении территорий. Более того, при решении частных задач биологически правомерными могут быть сравнения сильно отличающихся по своему объему флор (например, сопоставление систематической структуры флористической области со структурой флор слагающих ее районов).

Необходимо также отметить, что разные флористические показатели неодинаково тесно связаны с площадью флоры и общим числом свойственных ей таксонов и, следовательно, в разной степени чувствительны к разнице объемов сравниваемых флор (см. раздел 8.3).

Мало кто из флористов задавался целью изучения видового состава флор одинаковых по своей площади участков суши. Именно поэтому имели место попытки [Гроссгейм, 1936; Lebrun, 1960, и др.] пересчета количества видов на единицу площади. Авторы надеялись таким путем устранить влияние размера территории на число видов и получить сопоставимые коэффициенты флористического богатства.

Однако, как отмечает А. И. Толмачев [1941], операция пересчета числа видов на единицу площади имеет биологический смысл лишь тогда, когда изменение флористического богатства при наращивании территории в пределах данного района происходит равномерно и одинаково во всех его частях. В противном случае (а чаще всего так и бывает) полученный для района коэффициент оказывается абстрактной средней величиной, резко отличающейся от коэффициентов, свойственных отдельным частям того же района. Ясно, что в таких условиях сравнение средних показателей разновеликих районов создает лишь «иллюзию ботанической осмысленности пересчета» [Толмачев, 1941, с. 12], ибо простое разделение большего района на части, близкие по размерам меньшему району, может привести к со-

вершено иным выводам о степени их флористического сходства.

Более эффективный способ получения сопоставимых данных использовали Р. Бешел [1969] и Л. И. Малышев [1969]. Будучи, вообще говоря, сторонниками сравнения строго одинаковых по площади участков местности, в цитированных работах эти авторы сравнивали весьма разновеликие территории: первый — от 800 до 330 200 кв. км., а второй — от 5800 до 4 381 400 кв. км. Учитывая, что число видов зависит не только от площади, но и от многих других (климатических, физико-географических, исторических) факторов, Бешел и Малышев обратились к методам множественной регрессии.

Следует, однако, заметить, что как при простом пересчете числа видов на единицу площади, так и в случае регрессионного анализа факторов, влияющих на число видов в сильно отличающихся по своей площади флорах, ботанический смысл сравнения полученных показателей недостаточно ясен, так как фактически сравниваются несопоставимые по своему рангу площади или подразделения флоры.

Определенные возможности для произвольного истолкования оставляет и второй из перечисленных ранее принципов — ограниченный размер сравниваемых территорий: разные исследователи отдают предпочтение разным по площади эталонным участкам местности.

Так, Кайё [Cailleux, 1953] предложил сравнивать флористическое богатство территорий площадью 10 000 кв. км, а Моно [Monod, 1955] увеличил в аридных условиях эту площадь до 100 000 кв. км. Р. Бешел [1969], Л. И. Малышев [1969] и Б. А. Юрцев [1975] при изучении арктических и высокогорных флор сокращают размер учетной площади до 100 кв. км. Эти взаимоисключающие точки зрения в значительной мере обусловлены тем, что указанные авторы имели дело с флорами совершенно разных природных зон земного шара, но понимание этого не продвигает нас на пути решения проблемы биолого-статистической сопоставимости флор.

Можно признать, что формально-статистическая равноценность флор достигается путем сравнения топографически равных территорий или путем пересчета числа видов на единицу площади. Однако эти условия сами по себе не обеспечивают сопоставимости флор в ботанико-географическом отношении, если, конечно, не исходить из крайних представлений о флористическом континууме. Флора одновременно и непрерывна и дискретна (иначе мы не могли бы подразделять ее на категории различного ранга). Поэтому, если добиваться сравнения участков строго определенной площади или приведенных к единице площади количественных показателей, не обусловливая правомерность такого действия соображениями биологического характера, то при любой самой точной обработке данных эта

процедура останется сравнением безликих площадей или абстрактных цифр, а не реально существующих в природе подразделений флоры.

Правильнее стремиться к прямому сравнению флористических категорий одинакового ранга, биологическая сопоставимость которых, невзирая на некоторую (в пределах разумного допустимого) разницу в площадях, обеспечена их одинаковым положением в системе флористической иерархии. Конечно, при этом нельзя пренебрегать статистической стороной дела (именно поэтому мы и говорим о двойном условии биолого-статистической сопоставимости флор!) и сравнивать слишком разноразличные категории, даже если их ранг одинаков.

Указанному условию наилучшим образом соответствует представление о конкретной (элементарной) флоре, являющейся низшим флористическим подразделением.

8.2. МЕТОД КОНКРЕТНЫХ ФЛОР А. И. ТОЛМАЧЕВА

Метод конкретных флор, предложенный А. И. Толмачевым [1931, 1932, 1935, 1941, 1970], за последние 10—15 лет получил дальнейшее теоретическое развитие и применение в трудах многих советских ботаников [Малышев, 1972а, 1975, 1976; Ребристая, Шмидт, 1972; Шмидт, 1972, 1974, 1975, 1976, 1980; Юрцев, 1974, 1975, 1982; Ребристая, 1977; Шеляг-Сосонко, 1980, и др.].

Согласно первоначальному определению совокупность видов, слагающих конкретную флору, должна быть распространена на всей территории занимаемого флорой района, образуя лишь различные группировки в зависимости от чисто местных условий обитания. При этом участкам, одинаковым по условиям произрастания, соответствуют одинаковые комбинации видов, т. е. каждое сообщество характеризуется некоторым постоянством флористического состава. Флористические различия между отдельными участками конкретной флоры вызваны разницей в условиях обитания, тогда как другие факторы (географический, исторический и т. д.) играют в распределении видов ничтожную роль.

В соответствии с этим район конкретной флоры «должен быть достаточно мал для того, чтобы обеспечить действительную универсальность для него данной флоры, вместе с тем достаточно велик для того, чтобы охватить все возможные в его условиях станции, так как в противном случае отличия систематического состава разных сообществ затемнят чисто флористические особенности» [Толмачев, 1941, с. 15]. К этому определению А. И. Толмачев добавляет в цитированной работе еще одно требование, имеющее не только теоретический, но и практический характер: в силу слепого случая набор видов в одинаковых станциях может быть несколько различен, поэтому не-

обходима некоторая повторность изучения одинаковых местобитаний.

В процессе теоретического развития и практического использования метода конкретных флор оформились два самостоятельных направления [Юрцев, 1974, 1975; Шмидт, 1976, 1980].

Одно из них сложилось в условиях флористических исследований различных районов Советской Арктики и высокогорной Субарктики. Оно основано на понятии о «флористическом континууме», согласно которому «флористический градиент», т. е. непрерывное и постепенное изменение состава флоры, наблюдается на весьма обширных территориях. Это направление сужает объем понятия «конкретная флора», считая таковой «флору географического пункта», «пробу флористической ситуации в данном районе» или просто «пробу флоры», которая может быть малой (10 кв. км), средней (100 кв. км) и большой (1000 кв. км). Несмотря на признание реальности конкретных флор «как естественных флористических выделов» [Юрцев, 1975, с. 82], данное направление рассматривает конкретную флору лишь в качестве статистической выборки из флоры более крупного района. Считается, что методом конкретных флор «фактически изучаются эталонные флористические участки, а не флоры в целом» [Малышев, 1975, с. 1545—1546].

Такой подход к делу, заимствованный из геоботаники, в которой исследователи стремятся к сравнению большого числа пробных площадок, действительно представляющих собой всего лишь выборки из фитоценозов, понятен с чисто статистической точки зрения, но он недостаточен во флористическом отношении, ибо не отражает полностью биологической сущности конкретных флор как элементарных подразделений флоры. Мы не можем безоговорочно переносить положения, справедливые для фитоценологии, в область флористики. Между пробными площадками геоботаников и участками выявления конкретных флор (даже если понимать их как «пробы флоры») имеется не только количественная разница. Распределение видов в фитоценозах и флорах обусловлено воздействием разных факторов. Пробная площадка должна обеспечить выявление серии фитоценологических показателей, а участок конкретной флоры — только достаточную для данной местности полноту инвентаризации видового состава.

Между понятиями «пробная площадка» и «конкретная флора», действительно, имеется нечто общее: оба они в узко определенном смысле соответствуют представлению о статистической выборке, позволяя на основе частных наблюдений составить мнение о фитоценозе или флоре более крупного района. Но если роль пробной площадки этим и ограничивается, то конкретная флора является в отличие от нее естественным и целостным природным образованием.

Другое направление, развиваемое на кафедре ботаники Ленинградского государственного университета, исходит из представления о том, что наряду с чертами непрерывности флорам свойственны и черты дискретности, без которых было бы немислимым выделение в природе флористических подразделений различного ранга. Это направление оформилось в ходе флористических работ в пределах лесной зоны северо-запада РСФСР. Его важнейшим положением служит признание за конкретной флорой статуса наименьшей естественной флористической единицы. Таким образом, второе направление, отнюдь не отрицая статистического аспекта понятия о конкретной флоре, вкладывает в это понятие ботанико-географический смысл.

Конкретная флора — низшая, но специфическая ступень в иерархии флористических подразделений, не предназначенная, в частности, для картографического отражения занимаемой ею территории. Наименьшей единицей флористического районирования надо считать флористический район (или подрайон), который в общем случае значительно крупнее и слагается серией конкретных флор, переходящих в друг друга.

Конкретную флору следует рассматривать, с одной стороны, как элементарную естественную флористическую единицу (т. е. как генеральную совокупность произрастающих на ее территории видов растений), а с другой стороны, как выборку из состава флоры более крупного района, позволяющую с довольно высокой степенью точности судить о видовом составе последней. В такой трактовке статистической сущности понятия о конкретной флоре нет ничего противоречивого: в зависимости от поставленной задачи один и тот же массив исходных данных может представлять собой в одном случае всю генеральную совокупность, а в другом — всего лишь выборку из генеральной совокупности. Объем генеральной совокупности задается исследователем в соответствии с целями работы. Для сравнительно-флористических целей конкретная флора предстает перед нами как генеральная совокупность (т. е. как целостная единица флоры), а при необходимости приблизительно характеризовать флору более крупного района конкретная флора служит выборкой (моделью).

Конкретная флора обладает рядом свойств, некоторые из которых выгодно отличают ее от флористических категорий иного ранга и являются перспективными для решения многих задач сравнительной флористики и ботанической географии. Рассмотрим кратко наиболее важные из этих свойств.

Конкретная флора принципиально неделима на чисто флористических основаниях. Внутри нее можно, конечно, выделить «флоры» ландшафтных участков, урочищ, экотопов («парциальные флоры» в смысле Б. А. Юрцева). Изучение их представляет несомненный, но совершенно особый интерес, ибо подобного рода «флоры», будучи выделены на основе ландшафто-

ведческих, экологических или геоботанических критериев, не являются флорами в строгом смысле этого слова.

Основной структурной ячейкой конкретной флоры служит фитоценоз. Конкретная флора состоит из набора некоторого количества фитоценозов, причем однотипные из них могут повторяться внутри нее, а само количество фитоценозов зависит от степени пестроты условий среды. Поэтому любая попытка «дробления» конкретной флоры уводит нас из области флористики прежде всего в область геоботаники, а также в область смежных дисциплин (например, экологии и ландшафтоведения).

Конкретная флора — реально существующая флористическая единица. Доказательством этого служит то, что конкретным флорам присущи черты, в общей форме свойственные флористическим подразделениям любого ранга: принципиальная (но практически не обязательно реализуемая) возможность территориального отграничения; определенный уровень флористического богатства; определенное систематическое строение; характерное для данной области или района соотношение генетических, географических и других элементов флоры. В то же время конкретная флора обладает специфическими особенностями.

Одна из главных особенностей конкретной флоры состоит в том, что это понятие не требует привязки к определенной площади, хотя эта площадь и может быть установлена. Границы конкретной флоры обусловлены не квадратными километрами, а биотическими и историческими факторами, действовавшими в данном районе. Наиболее важным следствием этого является высокая степень сопоставимости конкретных флор, обеспечивающая возможность их сравнения в чрезвычайно широком географическом диапазоне.

Конкретные флоры обладают высокой степенью репрезентативности и в значительной мере могут служить своего рода моделями флоры того флористического района, к которому они относятся. По данным Е. В. Барановой [1973, 1977], сводный список 10 конкретных флор содержит 94,8 % видового состава флоры Псковской области, причем каждая из них включает в себя 75—90 % видов соответствующего флористического района. 13 конкретных флор, изученных О. В. Ребрисой [1977] в Большеземельской тундре, содержат 96,4 % видового состава флоры всей территории.

Из предыдущего свойства вытекает еще одна важная особенность метода конкретных флор — его эффективность или экономичность по сравнению с традиционным маршрутным способом изучения флор.

Остановимся теперь на вопросе выбора участка для изучения конкретной флоры, который не так прост, как это может показаться. Если исходить из представления о флористическом

континууме, полагая, что на достаточно обширной территории имеет место непрерывный флористический градиент, то в принципе любая, взятая наугад в пределах этой территории точка может служить как бы центром своей конкретной флоры.

Однако уже говорилось, что наряду с чертами непрерывности флоре свойственны и черты дискретности. В зависимости от условий флорогенеза может преобладать та или иная черта, что и характеризует одну из сторон диалектической сущности исторического развития флор. Если наличие флористического континуума допускает случайный выбор участков для изучения конкретных флор, то дискретность флор ограничивает подобного рода действия.

Для целей сравнительной флористики важно, чтобы через участок конкретной флоры не проходили ботанико-географические рубежи и, по возможности, границы ареалов видов, имеющих разные исторически сложившиеся области своего распространения. В этом отношении конкретная флора должна быть достаточно гомогенной. Однако при решении специальных задач флористического районирования (в целях уточнения положения флористических рубежей) возможно умышленное заложение конкретных флор на предполагаемых стыках флористических, геоботанических или ландшафтных районов. Сказанное определяет необходимость сознательного, а не случайного выбора участка конкретной флоры на местности. В основу такого выбора должны быть положены все имеющиеся данные флористического, геоботанического и геоморфологического характера.

8.2.1. Полевая методика изучения конкретных флор. В непосредственной близости от выбранной исходной точки (базы) обычно удается собрать (или зарегистрировать) не менее 50 % видов, присущих конкретной флоре. В большинстве своем это будут тривиальные виды растений.

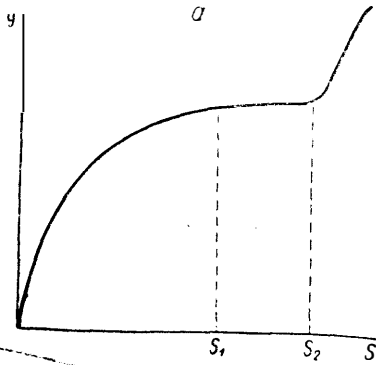
После этого во всех направлениях предпринимаются радиальные маршруты постепенно увеличивающейся дальности, прокладываемые с таким расчетом, чтобы в некоторой повторности и в разные сроки вегетационного периода обследовать все свойственные данной местности типы местообитаний и формы рельефа.

Список регистрируемых видов вначале быстро возрастает; затем этот процесс замедляется и, наконец, практически приостанавливается. Это свидетельствует о достижении границ минимум-ареала конкретной флоры или площади ее выявления.

Дальнейшее увеличение радиусов обследования поначалу ничего или почти ничего не добавляет к имеющемуся списку видов: исследователь находится в пределах так называемой «мертвой зоны». Но затем процесс прироста списка возобновляется с новой силой за счет обнаружения в знакомых типах

местообитаний видов растений, которые ранее в них не встречались. Это сигнал о достижении максимум-ареала конкретной флоры, свидетельствующий о том, что пересечены границы ареалов видов, не свойственных данной конкретной флоре. Другим указанием на это может служить исчезновение в соответствующих типах местообитаний тех видов, которые до сих пор регулярно в них встречались.

Указанные соотношения можно изобразить графически (рис. 37, а) или в виде схемы (рис. 38), оговорив при этом, что конфигурация территории конкретной флоры в каждом отдельном случае определяется местными условиями, т. е. не обязательно должна иметь форму круга.



б

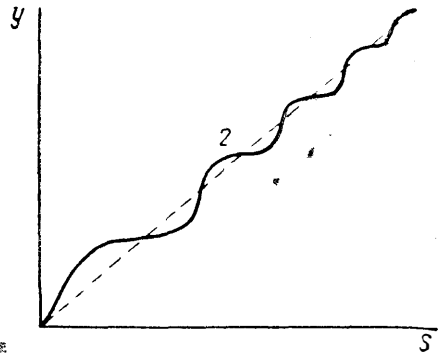
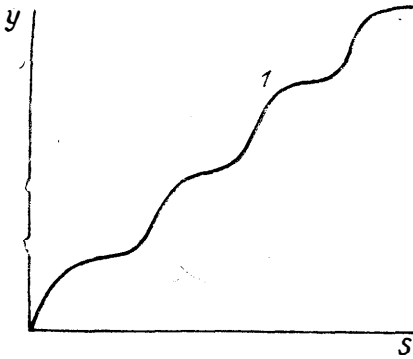


Рис. 37. Схема зависимости числа видов (y) от площади (S) при изучении конкретной флоры (а) и при изучении фитоценозов, слагающих конкретную флору (б).

S_1 — минимум-ареал; S_2 — максимум-ареал. 1 — нарастание числа видов по мере включения в исследование новых фитоценозов (каждая «ступень» символизирует ход кривой в пределах отдельного фитоценоза); 2 — «затухание» ступенчатой кривой при ранжировании фитоценозов по занимаемой ими площади.

Строго говоря, для инвентаризации видового состава конкретной флоры необходимо и достаточно изучить ее минимум-ареал. Однако определение его границ на местности иногда вызывает затруднения. Поэтому для гарантии полноты выявления видового состава рекомендуется продолжать обследование «мертвой зоны» вплоть до выхода на рубеж максимум-ареала

конкретной флоры, получая таким образом не одну, а две опорные точки на каждом радиусе. Как уже говорилось, количество зарегистрированных видов от этого практически не изменяется, но зато появляется уверенность в правильном ограничении территории конкретной флоры.

8.2.2. О площади конкретной флоры. Предварительное знание размера минимум-ареала конкретной флоры освободило бы флориста от непроизводительной затраты времени и сил на исследование «мертвой зоны». Поскольку нет, однако, никаких оснований полагать, что в разных природных условиях конкретные флоры имеют одинаковую площадь выявления, то проблема приобретает теоретический характер: важно знать, в каких условиях какие размеры развивают конкретные флоры. Соотношение минимум- и максимум-ареалов и, в частности, ширина «мертвой зоны» также не могут быть постоянными. Сама площадь конкретной флоры может рассматриваться как интересный флористический параметр, варьирующий в зависимости от условий среды.

Проблема площади выявления конкретной флоры в сравнительной флористике во многом аналогична проблеме площади выявления фитоценоза, представляющей собой один из самых спорных и еще не разрешенных вопросов геоботаники. Основная трудность в обоих случаях состоит в точном описании зависимости числа видов от площади и установлении размера минимум-ареала, если понимать его как площадь выявления конкретной флоры или фитоценоза.

Одна из первых формул, отражающих указанную зависимость, была предложена в начале 1920-х годов О. Аррениусом:

$$\frac{x}{x_1} = \left(\frac{y}{y_1} \right)^n,$$

где y и y_1 — количество видов на площадях x и x_1 ; n — константа, различная для разных ассоциаций.

Если условиться, что y_1 — количество видов на единице площади x_1 , равное a , то уравнение Аррениуса принимает вид экспоненциальной функции $y = ax^z$, где y — искомое число видов на данной площади, z — константа разнообразия растительного сообщества ($z = 1/n$; $0 < z < 1$). Из этой формулы следует, что при бесконечном наращивании площади число видов увеличивается также до бесконечности.

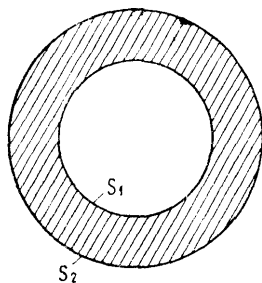


Рис. 38. Схематическое изображение границ минимум-ареала (S_1) и максимум-ареала (S_2) конкретной флоры. Заштрихована «мертвая зона».

Чуть позже Аррениуса другую формулу предложил Х. Глизон:

$$\frac{\log B - \log A}{\log C - \log A} = \frac{b - a}{c - a},$$

где A — площадь одного квадрата (пробной площадки); B — площадь всех квадратов; C — общий размер площади исследования; a — среднее число видов в одном квадрате; b — общее число видов во всех квадратах; c — общее число видов на всей исследованной площади.

После некоторых преобразований с целью линеаризации указанные уравнения можно представить в следующей сопоставимой форме:

$$\text{уравнение Аррениуса: } \log y = \log a + z \log x,$$

$$\text{уравнение Глисона: } y = a + b \log x,$$

где y — количество видов на площади размером x ; z и b — константы, отражающие флористическое разнообразие участка; a — количество видов, свойственных единице площади.

Некоторое внешнее сходство уравнений Аррениуса и Глисона в этой записи является обманчивым. Это видно уже из того, что для получения на графике прямой линии, отражающей зависимость «число видов — площадь», в первом случае приходится градуировать оси координат в логарифмическом, а во втором — в полулогарифмическом масштабах. Все дело в том, что в эти уравнения заложен существенно различный в биологическом отношении смысл: формула Аррениуса предполагает, что при бесконечном наращивании площади число видов будет увеличиваться также до бесконечности, тогда как по уравнению Глисона число видов по мере увеличения площади асимптотически затухает, стремясь к некоторому определенному пределу.

Аррениус имел дело с площадями, измеряемыми квадратными дециметрами, Глизон работал на площадях от 1 до 256 кв. м, и оба уравнения предназначались для применения в условиях достаточно гомогенной среды одного или ряда сходных фитоценозов. Но число видов в фитоценозе не бесконечно, и в этом смысле формула Аррениуса менее правильна. С другой стороны, уравнение Глисона (см. вторую его запись) также обладает недостатком: при нулевой площади ($x=0$) число видов не равно нулю ($y=a$).

Л. И. Малышев [1972а] считает допустимым использование в сравнительной флористике обоих уравнений, но отдает некоторое предпочтение формуле Глисона. Биологический смысл константы z из уравнения Аррениуса и константы b из уравнения Глисона он видит в том, что обе они оценивают про-

странственное разнообразие флоры, измеряя прирост числа видов при увеличении участка на единицу логарифма площади. Для их вычисления необходимо знать количество видов, присутствующее району в целом (y_2 на площади x_2), и количество видов, свойственное наиболее характерной части этого района (y_1 на площади x_1):

$$z = \frac{\log y_2 - \log y_1}{\log x_2 - \log x_1}; \quad b = \frac{y_2 - y_1}{\log x_2 - \log x_1}.$$

Таким же образом можно вычислять величины z и b для флор двух разновеликих территорий (площадью соответственно x_1 и x_2). Вычисленные Малышевым значения z для 404 материковых флор, их частей и флор крупных островов варьируют в пределах от 0,038 до 0,240, причем в направлении с севера на юг пространственное разнообразие флор возрастает, что представляется вполне естественным. Надо заметить, однако, что использование z -константы из уравнения Аррениуса и b -константы из уравнения Глисона предполагает два условия: а) наличие флористического континуума и б) допущение о случайном распределении видов. Эти условия далеко не всегда могут считаться приемлемыми в сравнительно-флористических исследованиях.

Для изучения зависимости числа видов от площади геоботаниками было предложено множество других формул, причем некоторые из авторов пытались определять размеры минимум-ареала исходя из кривых параболического типа, которые выводятся на основе предположения о непрерывности растительного покрова.

Как впервые совершенно верно подметил В. И. Василевич [1969] и на что обратил внимание Л. И. Малышев [1972а], большинство геоботаников, хотели они того или нет, фактически руководствовались стремлением получить единое уравнение, пригодное для описания интересующей нас зависимости в широчайшем диапазоне площадей — от пробной площадки до флоры континента и всего земного шара. Такое стремление из-за одной только неоднородности растительного покрова заранее обречено на неудачу. Единого уравнения, описывающего зависимость числа видов от площади на территориях любого размера, быть не может: это соотношение в пределах фитоценоза и в пределах флоры подчиняется разным законам. Необходимы поиски формул, описывающих соотношение «число видов — площадь» на разных уровнях геоботанической и флористической иерархии.

Обращаясь к конкретным флорам, следует напомнить, что они слагаются фитоценозами. По мере увеличения площади в пределах одного растительного сообщества число видов, как известно, вначале нарастает быстро, но затем замедляется и при-

останавливается. При выходе за пределы данного сообщества процесс прироста числа видов возобновляется. Это означает, что по мере обследования новых экотопов кривая на графике должна приобрести ступенчатую форму (см. рис. 37, б). Таким образом, если при изучении конкретной флоры вести учет видов по фитоценозам, а затем ранжировать их по мере уменьшения числа свойственных им видов (или по мере уменьшения их площади), то кривая приобретет характер затухающих колебаний (см. рис. 37, б).

В качестве первого приближения можно принять допущение о равномерном затухании периода и амплитуды этих колебаний. Тогда искомая эмпирическая формула может иметь вид

$$y = ax + e^{-bx} \cdot \sin(x^c),$$

где a — тангенс угла наклона «равнодействующей» к оси абсцисс; b — параметр, характеризующий затухание амплитуды; c — параметр, характеризующий затухание периода.

Изложенные выше соображения могут быть положены в основу специального эксперимента на местности. Но из опыта работы флористов уже сейчас достаточно ясно, что если в условиях Арктики и высокогорий площадь 100 кв. км обычно вполне гарантирует выявление конкретной флоры, то в условиях лесной зоны северо-запада РСФСР приходится изучать площади порядка 700—750 кв. км. Латвийские ботаники [Табака и др., 1977] экспериментальным путем (посредством последовательного наращивания квадратов определенного размера) пришли к близким цифрам: оказалось, что минимум-ареал конкретных флор в Западной Латвии составляет 600, а максимум-ареал — 650—700 кв. км. С учетом всего сказанного можно заключить, что конкретные флоры указанного размера, расположенные в Арктике и на северо-западе РСФСР, вполне сопоставимы друг с другом, несмотря на семикратную разницу в площадях.

8.3. КОЛИЧЕСТВЕННЫЕ ПОКАЗАТЕЛИ (ПАРАМЕТРЫ) ФЛОР, ИХ СВОЙСТВА И СПОСОБЫ СРАВНЕНИЯ

8.3.1. Показатели флористического богатства и систематического разнообразия. Такие показатели, как число видов, родов и семейств, свойственных данной флоре, а также число (или процент) этих таксонов в составе более крупных систематических групп, характеризуют флористическое богатство любой территории и тем самым представляют самостоятельную научную ценность. Если эти показатели относятся к конкретным флорам, то их прямое сравнение представляется вполне правомерным. В остальных случаях оно допустимо лишь при усло-

вии приблизительно территориальной равноценности сопоставляемых флор. Дело не столько в том, что показатели флористического богатства зависят от площади флоры, сколько в том, что эта зависимость в разных условиях может иметь различный характер и разную степень выраженности — от очень высокой до статистически недостоверной. Неоднозначность ука-

Таблица 66. Число видов в четырех разновеликих флорах

Территория	Площадь, кв. км	Число видов
Средняя полоса европейской части СССР	1 207 000	2330
Польша	313 000	2250
Белоруссия	208 000	1530
Ленинградская обл.	86 000	1283

Таблица 67. Показатели флористического богатства и систематического разнообразия флор южного Средиземноморья

Флоры	Площадь, кв. км	Число			Пропорции флоры		
		видов	родов	семейств	в/с	р/с	в/р
Алжир	2 382 000	3139	912	133	23,6	6,9	3,4
Ливия	1 760 000	1297	484	98	13,2	4,9	2,7
Египет	994 300	1911	689	105	18,2	6,6	2,8
Марокко	444 000	3494	806	132	26,5	6,1	4,3
Сирия	190 000	2478	658	125	19,8	5,3	3,8
Тунис	160 000	2158	750	119	18,1	6,3	2,9
Палестина	14 000	2321	711	141	16,5	5,0	3,3
Ливан	10 400	2109	750	129	16,3	5,8	2,8
Канарские о-ва	7 167	1574	539	119	13,2	4,5	2,9
О-в Мадейра	797	735	390	95	7,7	4,1	1,9

Примечание. в/с — среднее число видов в семействе; р/с — среднее число родов в семействе; в/р — среднее число видов в роде.

занной связи и является основным препятствием для сравнения флористического богатства разновеликих флор.

Приведем некоторые примеры. Из табл. 66 видно, что число видов в четырех разновеликих флорах явно зависит от их площади (коэффициент ранговой корреляции Спирмена,¹ рассчитанный для данной таблицы, составляет $\rho_s = 1$), однако эта зависимость не является пропорциональной. Совершенно очевид-

¹ Этот показатель подробно рассматривается в разделе 8.3.2.

но, что на числе видов в указанных флорах сказывается не только их площадь, но и зональное положение.

К иному результату приводит изучение зависимости показателей флористического богатства от площади в серии региональных флор южного Средиземноморья (табл. 67). Расчеты показывают, что коэффициент ранговой корреляции между числом видов и площадью этих флор составляет $\rho_s=0,43$, между числом родов и площадью — $\rho_s=0,41$ и между числом семейств и площадью — $\rho_s=0,18$, причем указанные значения связи статистически недостоверны при 5%-ном уровне существ-

Таблица 68. Количество видов, зарегистрированных в окрестностях некоторых населенных пунктов Швеции

Населенные пункты	Площадь, кв. км	Число видов
Боллебюгд	180	583
Симонсторп	162	581
Квиллинге	128	746
Трене	102	751
Крукек	93	763
Веддиге	81	585
Кварсебу	70	652
Фегре	68	551
Биллингс	30	563
Дала	25	587
Харстена	21	459

Примечание. Данные взяты из ряда работ шведских авторов.

венности (Приложение XI). Действительно, флора Марокко, уступающего по площади Алжиру, Ливии и Египту, заметно богаче их по числу видов, а флора Канарских о-вов по всем трем показателям флористического богатства богаче флоры Ливии, территория которой в 245 раз крупнее. Нарушение обычной зависимости этих показателей от площади в рассматриваемом случае можно объяснить тем, что огромные пространства некоторых стран заняты пустыней, мало что добавляющей к списку растений.

Причины нарушения обсуждаемой зависимости могут быть и другими. В табл. 68 приведены данные о числе видов в 11 небольших районах Швеции, расположенных в ограниченном ши-

ротном интервале. Связь между числом видов и площадью невелика ($\rho_s=0,41$) и недостоверна. Это можно объяснить тем, что перечисленные районы по своей площади малы (не достигают размеров минимум-ареала конкретной флоры), а уровень видового богатства в таких условиях сильнее, чем от площади, зависит от локальных особенностей рельефа, специфики и ограниченности набора местообитаний, свойственных каждому такому участку.

Другая особенность показателей флористического богатства состоит в том, что они довольно сильно коррелируют друг

Таблица 69. Связь между показателями флористического богатства в сериях конкретных флор

Конкретные флоры	Значения коэффициента ранговой корреляции ρ_s Спирмена между		
	числом видов и числом родов	числом видов и числом семейств	числом родов и числом семейств
17 конкретных флор Большеземельской тундры и некоторых более северных районов	0,90	0,81	0,94
11 флор п-ва Канин	0,96	0,90	0,93
12 флор Коми АССР	0,93	0,67	0,70
20 флор Ленинградской, Псковской и Новгородской областей	0,96	0,73	0,72

Примечание. Исходные данные для расчета коэффициентов взяты из работ О. В. Ребристой, В. Г. Сергиенко, В. А. Мартыненко и В. М. Шмидта. Значения коэффициентов 0,67 и 0,70 достоверны при 5%-ном, а все остальные — при 1%-ном уровне существенности.

с другом. Расчеты показывают, что во флорах южного Средиземноморья (табл. 67) связь между числом видов и родов составляет $\rho_s=0,83$, между числом видов и семейств — $\rho_s=0,84$, а между числом родов и семейств — $\rho_s=0,77$ (все значения достоверны, причем первые два из них — на 1 %-ном уровне существенности). Такая же закономерность справедлива и для конкретных флор, что видно из табл. 69. Заметим также, что связь между показателями флористического богатства является нелинейной: при увеличении площади число родов и особенно семейств нарастает значительно медленнее числа видов.

Полезные результаты дает анализ процентного распределения количества видов, родов и семейств по крупным группам растений в сравниваемых флорах. Например, во флорах южного Средиземноморья (табл. 70) прослеживается тенденция усиления относительной роли видов, родов и семейств Pteridophyta в направлении с востока на запад, причем происходит

Таблица 70. Процентное распределение видов, родов и семейств по более крупным группам растений в региональных флорах южного Средиземноморья

Группы растений	Сирия	Ливан	Палестина	Египет	Ливия	Тунис	Алжир	Марокко	Канарские о-ва	Мальдивы
Количество видов										
Pteridophyta	0,8	1,0	0,6	0,5	0,7	1,1	1,4	1,3	3,0	5,8
Gymnospermatophyta	0,4	0,5	0,4	0,2	0,4	0,5	0,5	0,5	0,6	0,5
Anthophyta	98,8	98,5	99,0	99,3	98,9	98,4	98,1	98,2	96,4	93,7
В том числе:										
Monocotyledoneae	14,6	16,1	18,0	20,9	19,4	19,8	18,9	17,8	16,0	17,3
Dicotyledoneae . .	85,4	83,9	82,0	79,1	80,6	80,2	81,1	82,2	84,0	82,7
В том числе:										
Choripetalae . .	57,4	61,1	60,1	60,9	62,2	61,3	58,3	61,0	57,5	57,7
Sympetalae . . .	42,6	38,9	39,9	39,1	37,8	38,7	41,7	39,0	42,5	42,3
Количество родов										
Pteridophyta	1,8	2,0	1,4	1,1	1,6	2,1	2,5	2,7	3,5	5,7
Gymnospermatophyta	0,9	0,5	0,6	0,3	0,8	0,5	1,0	1,0	0,5	0,8
Anthophyta	98,3	97,5	98,0	98,6	97,6	97,4	96,5	96,3	96,0	93,5
В том числе:										
Monocotyledoneae	17,4	21,2	19,9	23,3	22,4	23,1	22,7	17,4	20,6	19,4
Dicotyledoneae . .	82,6	78,8	80,1	76,7	77,6	76,9	77,3	82,6	79,4	80,6
В том числе:										
Choripetalae . .	59,3	56,6	59,5	68,2	56,1	60,1	60,2	60,2	56,3	52,8
Sympetalae . . .	40,7	43,4	40,5	31,8	43,9	39,9	39,8	39,8	43,7	47,2
Количество семейств										
Pteridophyta	5,5	6,2	4,9	5,7	5,1	7,7	8,3	8,3	8,4	9,6
Gymnospermatophyta	3,2	2,3	2,1	1,9	4,0	2,5	3,0	3,1	2,6	2,1
Anthophyta	91,3	91,5	93,0	92,4	90,9	89,8	88,7	88,6	89,0	88,3
В том числе:										
Monocotyledoneae	13,1	12,7	13,8	16,4	17,1	17,5	18,8	16,2	14,2	13,3
Dicotyledoneae . .	86,9	87,3	86,2	83,6	82,9	82,5	81,2	83,8	85,8	83,7
В том числе:										
Choripetalae . .	69,5	70,8	70,7	70,3	67,2	70,6	71,5	72,4	70,3	62,5
Sympetalae . . .	30,5	29,2	29,3	29,7	32,8	29,4	28,5	27,6	29,7	37,5

это за счет Anthophyta. Ливия выделяется наибольшим числом семейств Gymnospermatophyta, но по числу родов и видов голосеменных уступает некоторым другим странам. Египет наиболее беден видами и родами Pteridophyta, но по числу семейств папоротникообразных превосходит Палестину и Ливию.

К группе показателей систематического разнообразия относятся прежде всего так называемые «пропорции флоры»: среднее число видов в семействе (в/с), среднее число родов в семействе (р/с) и среднее число видов в роде (в/р). Вполне понятно, что «пропорции флоры», будучи простыми отношениями

Таблица 71. Показатели систематического разнообразия («пропорции флоры») ряда разновеликих флор СССР

Флоры	Площадь, тыс. кв. км	Пропорции флоры		
		в/с	р/с	в/р
П-ов Камчатка	472	13,0	4,2	3,1
Коми АССР	416	11,9	4,2	2,8
Белоруссия	208	13,3	5,0	2,7
Карельская АССР	172	10,6	4,3	2,5
Башкирская АССР	144	14,0	5,0	2,8
Ленинградская обл.	86	12,5	4,6	2,7
Псковская обл.	55	10,8	4,3	2,4
Ульяновская обл.	34	12,2	4,6	2,6
Орловская обл.	25	10,3	4,3	2,4
Р-н Пиедисярви-Лиекса (сев. ч. КАССР)	3,5	6,8	3,4	2,0

Примечание. Исходные данные для расчета показателей взяты из ряда отечественных литературных источников.

показателей флористического богатства, коррелируют с последними. Известно, что более богатые, а также территориально более крупные флоры отличаются повышенными значениями этих показателей, однако зависимость «пропорций флор» от площади изучена еще недостаточно. По-видимому, эта зависимость по сравнению с показателями флористического богатства в целом выражена слабее, но расхождения в силе связи в разных условиях сравнения флор могут быть весьма значительными. Это касается и показателя в/р (так называемого «родового коэффициента»), который обычно считается наиболее независимым от площади отношением.

Подтвердим сказанное двумя примерами. Первый из них касается серии региональных флор южного Средиземноморья (табл. 67). Расчеты показывают, что в этом случае достоверной

оказывается связь с площадью только показателя r/c ($\rho_s = 0,66$), тогда как показатели v/c и v/p достоверной связи с площадью не обнаруживают ($\rho_s = 0,63$ и $\rho_s = 0,31$ соответственно), причем наиболее слабой зависимостью отличается, действительно, «родовой коэффициент».

Если же мы возьмем для сравнения ряд разновеликих флор, расположенных в пределах Бореальной (отчасти Средневропейской) флористической области на территории СССР

Таблица 72. Отношение числа видов сем. Asteraceae к числу видов сем. Fabaceae (A/F) и к числу видов сем. Сурегасеae (A/C) во флорах разных флористических областей

Флористические области	Флоры	A/F	A/C
Арктическая	Новая Земля	4,3	0,9
	Вост. часть Чукотского п-ова	3,7	0,8
	Вост. часть Большеземельской тундры	3,2	0,9
	Арктическая Якутия	2,6	1,2
	Канадский арктический архипелаг	1,8	0,6
Бореальная	Ньюфаундленд	5,0	1,1
	Южная половина Коми АССР	3,8	1,3
	Башкирская АССР	3,3	1,8
	Карельская АССР	2,8	1,0
	Белорусская ССР	2,5	1,1
	Литовская ССР	2,5	1,1
	Центральная Якутия	2,1	1,2
Средиземноморская	Марокко	1,3	6,9
	Канарские о-ва	1,3	7,8
	Алжир	1,2	7,2
	Тунис	1,2	7,1
	Ливия	1,1	7,7
	О-в Мадейра	1,1	7,4
	Сирия	1,0	12,0
	Палестина	1,0	5,7
	Египет	0,9	3,6
	Ливан	0,9	6,7

(табл. 71), то получим противоположные результаты: наиболее сильно ($\rho_s=0,83$) и достоверно при 1 %-ном уровне существенности с площадью флор связан показатель v/p , тогда как зависимость от площади двух других показателей (для v/c $\rho_s=0,60$; для p/c $\rho_s=0,05$) недостоверна, причем показатель p/c практически вообще независим от площади.

Таким образом, сравнение «пропорций флоры», как и показателей флористического богатства, требует приблизительного равенства площадей сопоставляемых флор.

Редко употребляемыми, но весьма содержательными по своему смыслу являются индексы (отношения) видовой или родовой численности отдельных (специально подобранных для отражения той или иной закономерности) пар семейств. Такие индексы могут служить своеобразными «индикаторами» крупных флористических подразделений (например, областей), а полученные для серии более мелких (в том числе конкретных) флор — свидетельствовать об изменении относительной роли соответствующих семейств в экологическом или географическом аспекте.

Из данных табл. 72 видно, что при продвижении из Арктики в Бореальную и далее в Средиземноморскую флористическую область роль сем. Asteraceae относительно сем. Fabaceae падает, а относительно сем. Сурегасеае, наоборот, повышается. Подобного рода индексы, отражающие различия зонального характера и особенности исторического развития флор, служат полезными характеристиками при их анализе и сравнении. Надо также заметить, что эти индексы в меньшей степени, чем «пропорции флоры», зависят от площади.

8.3.2. Показатели сходства систематической структуры флор.

Свойственное каждой флоре распределение видов между систематическими категориями высшего ранга (включая показатели, рассмотренные в предыдущем разделе) А. И. Толмачев [1974] назвал систематической структурой флоры. В связи с необходимостью более четкой классификации признаков флор, по которым возможно их сравнение, понятие о систематической структуре трактуется нами [Шмидт, 1980] более узко. Используя этот термин, мы будем иметь в виду только так называемые *флористические спектры*, которые могут быть составлены по разным признакам. Важнейшими являются спектры, отражающие состав и последовательность расположения: 1) семейств по числу видов; 2) семейств по числу родов и 3) родов по числу видов.

Полный флористический спектр представляет собой ранжированный по одному из указанных признаков ряд семейств или родов. В сравнительных целях обычно используется не весь спектр, а лишь его головная часть, состоящая из 10—15 наиболее крупных («ведущих») семейств или родов. Опыт показывает, что набор из 10 ведущих по числу видов семейств вклю-

чает в себя более половины видового состава каждой флоры (в высоких широтах значительно больше, в тропиках чуть меньше). Такой набор весьма четко характеризует принадлежность флоры к определенной флористической области [Толмачев, 1974]. Положение в полном спектре других семейств может сильно варьировать в зависимости от ряда современных и исторических факторов, указывая на особенности той или иной флоры в пределах данной флористической области. Заменяют спектр семейства, представленные каждое одним видом (такие семейства играют ничтожную роль в сложении систематической структуры флоры, но могут представлять значительный интерес для понимания процессов флорогенеза).

Систематическая структура флоры в меньшей степени, чем другие флористические показатели, зависит от площади; на ней слабее сказывается также возможная неполнота инвентаризации флоры. Обладая известной консервативностью, систематическая структура достаточно крупных разновеликих флор в пределах одной и той же флористической области имеет сходный характер. Мелкие (в том числе конкретные) флоры, сохраняя характерные для флористической области особенности систематической структуры, в целом отличаются большой изменчивостью флористических спектров.

При сравнении систематической структуры флор исходные данные представлены *ранжированными рядами* ведущих семейств или родов (порядковый номер таксона в ряду представляет собой его ранг). Поэтому в качестве показателя сходства систематической структуры двух флор логично использовать коэффициенты ранговой корреляции τ Кендэла или ρ_s Спирмена [Бернштейн, 1968; Малышев, 1972б; Ребриская, Шмидт, 1972; Зайцев, 1973; Кендэл, 1975; Терентьев, Ростова, 1977; Шмидт, 1980, 1981 в]. Прежде чем переходить к их рассмотрению, необходимо обратить внимание на следующее.

Определяя ранг семейства (или рода) в головной части флористического спектра, исследователь может столкнуться с двумя случаями. В первом и более простом из них все семейства или роды обладают разным числом видов и занимают места (получают ранги) в порядке чисел натурального ряда, т. е. 1, 2, 3, ..., n . Это — случай *несовпадающих (несвязанных)* рангов, не вызывающий затруднений при вычислении коэффициентов ранговой корреляции.

Часто бывает, однако, что несколько семейств или родов, включая в себя одинаковое число видов, делят между собой два или несколько мест. Иногда, несмотря на это, удается отдать предпочтение некоторым из них перед остальными, выстроив их в порядке натурального ряда. Например, при равном числе видов можно повысить ранг того семейства, которое содержит большее число родов или играет во флоре более видную ценотическую роль. Но чаще отдать предпочтение тому или

ному таксону без потери чувства объективности оценки оказывается невозможным. В этом случае не остается ничего иного, как зафиксировать «дележ» нескольких мест группой семейств (или родов) и каждому из них придать одинаковый ранг, равный среднему арифметическому из суммы занимаемых ими мест. Например, если 4 семейства делят 5—8-е места, то каждому из них приписывается ранг, равный 6, 5 и т. д. Это — случай *усредненных (связанных)* рангов, требующий введения в формулы коэффициентов ранговой корреляции особым образом рассчитываемых поправок. В этой ситуации предпочтительнее использовать коэффициент ρ_s Спирмена, поправки для которого рассчитываются проще, чем для τ .

Таблица 73. Ведущие по числу видов (%) семейства и их ранги во флорах Египта и Сирии

Семейства	Египет		Сирия	
	%	ранг	%	ранг
Fabaceae	11,7	1	12,8	1
Poaceae	10,8	2	5,8	6
Asteraceae	10,7	3	12,5	2
Brassicaceae	5,1	4	6,0	5
Caryophyllaceae	4,6	5	4,1	8
Liliaceae	3,6	6	4,6	7
Chenopodiaceae	3,5	7	—	(11)
Lamiaceae	2,8	8	7,0	3
Boraginaceae	2,7	9	3,4	10
Scrophulariaceae	2,6	10	3,5	9
Apiaceae	—	(11)	6,1	4

Коэффициенты ранговой корреляции τ и ρ_s неадекватны по степени выражения связи. В конкретном исследовании можно пользоваться любым из них, причем в случае несвязанных рангов первый рассчитывается несколько проще. Оба они имеют одинаковые пределы, варьируя от +1 (полное совпадение рангов двух рядов) до -1 (полное несовпадение, обратная последовательность рангов), но отрицательные значения коэффициентов могут иметь место лишь при сопоставлении очень далеких флор.

Коэффициент Кендэла, который рекомендуется использовать в случае несвязанных рангов, имеет вид

$$\tau = \frac{2s}{n(n-1)},$$

где s — сумма рангов, вычисляемая особым, показанным ниже способом; n — число пар сравниваемых рангов (т. е. число пар семейств или родов, по которым производится сравнение систематической структуры флор).

Сравним, к примеру, систематическую структуру флор Египта и Сирии (табл. 73). Для сравнения взяты наборы 10 ведущих семейств, но поскольку эти наборы отличаются на одно семейство, то из математических соображений таблицу следует «уравнять», т. е. независимо от истинного места указанных семейств придать им ранг, равный 11 (указан в скобках), и вести сравнение уже не по 10, а по 11 семействам ($n=11$).

Выпишем из табл. 73 в две строки соответствующие ранги семейств, расположив какой-либо ряд (например, для Египта) в ранжированном порядке:

Египет:	1	2	3	4	5	6	7	8	9	10	11
Сирия:	1	6	2	5	8	7	11	3	10	9	4

Далее оперируем только со вторым рядом (Сирия). Находим ранг, равный 1, и сосчитываем количество рангов (цифр), стоящих справа и слева от него. Число первых берем со знаком плюс, а число вторых — со знаком минус: $+10-0=+10$. Вычеркиваем использованный ранг 1, находим ранг 2 и повторяем операцию: $+8-1=+7$ (вычеркнутые ранги из расчетов каждый раз исключаются). В конечном счете получаем нужную нам сумму рангов: $s = (10-0) + (8-1) + (3-5) + (0-7) + (5-1) + (5-0) + (3-1) + (3-0) + (0-2) + (0-1) = +19$. Подставляя величины s и n в формулу коэффициента Кендэла, находим

$$\tau = \frac{2 \cdot 19}{11 \cdot 10} = +0,35.$$

При наличии связанных рангов в формулу коэффициента Кендэла необходимо ввести поправки (U_x и U_y) на усредненные рангов в обоих (x и y) рядах:

$$\tau = \frac{s}{\sqrt{\left[\frac{n(n-1)}{2} - U_x \right] \left[\frac{n(n-1)}{2} - U_y \right]}}$$

причем поправки находят по формуле

$$U_{x, y} = \frac{\sum t(t-1)}{2},$$

где t — число объединенных рангов в каждой группе, а расчет величины s по сравнению с предыдущим случаем усложняется необходимостью исключения совпадающих рангов [Зайцев, 1973; Кендэл, 1975], что и заставляет считать коэффициент ρ_s Спирмена в этом случае более предпочтительным.

Коэффициент ρ_s Спирмена при отсутствии связанных рангов вычисляется по формуле

$$\rho_s = 1 - \frac{6 \sum d^2}{n^3 - n}$$

где $\sum d^2$ — сумма квадратов разностей ($d=x-y$) между соответствующими рангами (x и y) двух рядов; n — число пар сравниваемых рангов.

Так, для предыдущего примера, в котором сравнивается систематическая структура флор Египта (x) и Сирии (y), имеем:

x :	1	2	3	4	5	6	7	8	9	10	11
y :	1	6	2	5	8	7	11	3	10	9	4
d :	0	4	1	1	3	1	4	5	1	1	7
d^2 :	0	16	1	1	9	1	16	25	1	1	49

откуда $\sum d^2 = 120$ и

$$\rho_s = 1 - \frac{6 \cdot 120}{11^3 - 11} = 0,45.$$

В случае связанных рангов поправки (U_x и U_y) для коэффициента Спирмена рассчитывают по формуле

$$U_{x, y} = \frac{\sum (t^3 - t)}{12},$$

где t — число членов в каждой группе совпадающих рангов. Тогда формулу самого коэффициента можно записать следующим образом:

$$\rho_s = \frac{\frac{n^3 - n}{6} - (\sum d^2 + \sum U)}{\sqrt{\left(\frac{n^3 - n}{6} - 2U_x\right)\left(\frac{n^3 - n}{6} - 2U_y\right)}}$$

где $\sum U = U_x + U_y$.

В качестве примера приведем расчет коэффициента Спирмена для конкретных флор Ухты (x) и Кослана (y), расположенных в подзоне средней тайги Коми АССР (табл. 74):

x :	1	2	3	4	5	6	7	8	9	10,5	10,5	12
y :	1	2	4	6	3	6	6	10	12	8	10	10
d :	0	0	1	2	2	0	1	2	3	2,5	0,5	2
d^2 :	0	0	1	4	4	0	1	4	8	6,25	0,25	4

Подсчитываем нужные величины: $\sum d^2 = 33,5$; $(n^3 - n) : 6 = (12^3 - 12) : 6 = 286$. В ряду x имеет место одна группа из

двух совпадающих рангов, составляющих каждый 10,5. Следовательно, поправка для этого ряда

$$U_x = \frac{2^3 - 2}{12} = 0,5.$$

Ряд y содержит 2 группы по 3 совпадающих ранга каждая (ранги 6 и 10), поэтому

$$U_y = \frac{2(3^3 - 3)}{12} = 4,0$$

и $\Sigma U = 0,5 + 4,0 = 4,5.$

Подставляя полученные значения в формулу коэффициента Спирмена, имеем

$$r_s = \frac{286 - (33,5 + 4,5)}{(283 - 2 \cdot 0,5)(286 - 2 \cdot 4)} = 0,88.$$

Для сравнения систематической структуры флор можно пользоваться и обычным коэффициентом корреляции r Бравэ—Пирсона (раздел 4.2.1). Поскольку в рассматриваемом случае

Таблица 74. Ведущие по числу родов семейства и их ранги в двух конкретных флорах Коми АССР

Семейства	Ухта		Костан	
	число родов	ранг	число родов	ранг
Asteraceae	27	1	25	1
Рoaceae	23	2	19	2
Brassicaceae	16	3	12	4
Ranunculaceae	13	4	9	6
Rosaceae	12	5	13	3
Apiaceae	10	6	9	6
Caryophyllaceae	9	7	9	6
Lamiaceae	8	8	6	10
Orchidaceae	7	9	5	12
Scrophulariaceae	6	10,5	7	8
Fabaceae	6	10,5	6	10
Polypodiaceae	3	12	6	10

$n < 20$, то применять следует рабочую формулу этого коэффициента, предназначенную для малых выборок:

$$r = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

где x и y — значения рангов одноименных таксонов в структуре двух сравниваемых флор; n — число пар рангов.

Следует подчеркнуть, что эта формула дает результат, тождественный коэффициенту Спирмена, но обладает тем преимуществом, что не требует введения поправок в случае связанных рангов [Шмидт, 1981 в]. Проверим это на ранее рассмотренных примерах, выполнив предварительные расчеты для получения требуемых сумм ($\sum x$, $\sum y$, $\sum x^2$, $\sum y^2$, $\sum xy$).

При сравнении систематической структуры флор Египта и Сирии (случай несвязанных рангов, табл. 73) имеем:

x :	1	2	3	4	5	6	7	8	9	10	11
y :	1	6	2	5	8	7	11	3	10	9	4
x^2 :	1	4	9	16	25	36	49	64	81	100	121
y^2 :	1	36	4	25	64	49	121	9	100	81	16
xy :	1	12	6	20	40	42	77	24	90	90	44

Суммируя по строкам, находим: $\sum x = 66$, $\sum y = 66$, $\sum x^2 = 506$, $\sum y^2 = 506$, $\sum xy = 446$. Отметив, что в случае несвязанных рангов всегда соблюдаются равенства $\sum x = \sum y$ и $\sum x^2 = \sum y^2$, подставим полученные суммы в формулу коэффициента Бравэ—Пирсона:

$$r = \frac{446 - \frac{66 \cdot 66}{11}}{\sqrt{\left(506 - \frac{66^2}{11}\right)\left(506 - \frac{66^2}{11}\right)}} = 0,45.$$

Полученная величина совпадает с ранее рассчитанным для этого примера значением коэффициента Спирмена.

При сравнении систематической структуры двух конкретных флор Коми АССР (случай связанных рангов, табл. 74) предварительные расчеты выглядят так:

x :	1	2	3	4	5	6	7	8	9	10,5	10,5	12
y :	1	2	4	6	3	6	6	10	12	8	10	10
x^2 :	1	4	9	16	25	36	49	64	81	110,25	110,25	144
y^2 :	1	4	16	36	9	36	36	100	144	64	100	100
xy :	1	4	12	24	15	36	42	80	108	84	105	120

Суммируя по строкам, находим: $\Sigma x = 78$, $\Sigma y = 78$, $\Sigma x^2 = 649,5$, $\Sigma y^2 = 646$, $\Sigma xy = 631$. Отметив, что в случае связанных рангов $\Sigma x^2 \neq \Sigma y^2$, подставим полученные суммы в формулу коэффициента Бравэ—Пирсона:

$$r = \frac{631 - \frac{78 \cdot 78}{12}}{\sqrt{\left(649,5 - \frac{78^2}{12}\right)\left(646 - \frac{78^2}{12}\right)}} = 0,88.$$

Таким образом, и в случае связанных рангов $r = \rho_s$.

Адекватное измерение связи коэффициентами Бравэ—Пирсона и Спирмена при сравнении систематической структуры флор обусловлено спецификой исходного материала: сумма рангов двух рядов здесь всегда одинакова ($\Sigma x = \Sigma y$). Для коэффициента r это представляет собой частный случай, ибо при использовании его для изучения связи между количественными признаками в малых выборках указанное равенство, как правило, не имеет места.

В случае несвязанных рангов существует еще одно равенство, а именно $\Sigma x^2 = \Sigma y^2$, причем $\Sigma d^2 = 2(\Sigma x^2 - \Sigma xy)$. При наличии связанных рангов это равенство нарушается: $\Sigma x^2 \neq \Sigma y^2$, а величина $\Sigma d^2 = \Sigma x^2 + \Sigma y^2 - 2\Sigma xy$.

С учетом сказанного путем алгебраических преобразований [Шмидт, 1981] были получены новые, удобные в вычислительном отношении, формулы для расчета коэффициентов Бравэ—Пирсона и Спирмена: в случае несвязанных рангов

$$r = \rho_s = \frac{3[4 \Sigma xy - n(n+1)^2]}{n^3 - n},$$

а в случае связанных рангов

$$r = \rho_s = \frac{4 \Sigma xy - n(n+1)^2}{\sqrt{[4 \Sigma x^2 - n(n+1)^2][4 \Sigma y^2 - n(n+1)^2]}}$$

где все обозначения те же самые, что и ранее.

Предоставляя читателю возможность убедиться в правильности этих формул на рассмотренных выше примерах, уделим внимание вопросу о достоверности перечисленных коэффициентов сходства систематической структуры сравниваемых флор.

Достоверность коэффициента τ Кендэла при $n \leq 10$ оценивается по специальной таблице [Бернштейн, 1968; Терентьев, Ростова, 1977], а при $n > 10$ — по таблице стандартных значений критерия t (Приложение III), эмпирическая величина которого рассчитывается по формуле

$$t = \frac{\tau}{\frac{2(2n+5)}{9n(n-1)}}$$

где знаменатель представляет собой единую для данного числа пар рангов (n) ошибку репрезентативности (m -) коэффициента Кендэла. Эта ошибка может быть заранее рассчитана и табулирована [Зайцев, 1973, с. 217; Терентьев, Ростова, 1977, с. 150].

Для оценки достоверности коэффициента ρ_s Спирмена следует использовать Приложение XI, в котором приведены критические значения этого показателя при двух уровнях существенности.

Оценке достоверности коэффициента r Бравэ—Пирсона должно предшествовать z -преобразование Фишера (см. раздел 4.2.1), после чего обращаются к таблице стандартных значений критерия t (Приложение III).

Таблица 75. Места (ранги) ведущих по числу видов семейств во флорах Курильских о-вов

Семейства	Южные Курилы				Средние Курилы				Северные Курилы	
	1	2	3	4	5	6	7	8	9	10
	Шикотан	Кунашир	Итуруп	Уруп	Симушир	Кетон	Распуа	Магуа	Парамушир	Шумшу
Cyperaceae	1		3	2	2	2		4	2	2
Poaceae		2	2				3	1		3
Asteraceae	3	3		3	3	3	2	2	3	1
Rosaceae		4		4	4	4	4	3	5	8
Polypodiaceae s. l.	5	6	9	(11)	(16)	(12)	(14)	(11)	(14)	
Orchidaceae	6	7	6	7	9		6	(15)	(15)	(16)
Ranunculaceae	7	9	7	6	(12)	8	(11)	(14)	6	6
Caryophyllaceae	8	(14)	(13)	10	10	10	8	5	8	10
Polygonaceae	9	(15)	(16)	(14)	(15)	(16)	(16)	(16)	(16)	(15)
Liliaceae	10	5	5	5	5	5	5	6	9	5
Apiaceae	(11)	8	(11)	(12)	(13)	9	9	10	(13)	(13)
Fabaceae	(12)	(13)	8	(13)	6	(15)	(15)	(12)	(12)	(12)
Juncaceae	(13)	(12)	(14)	(15)	7	(13)	(13)	(13)	4	9
Lamiaceae	(14)	10	(15)	(17)	(17)	(17)	(17)	(17)	(17)	(17)
Scrophulariaceae	(15)	(11)	10	8	(11)	(11)	10	9	7	4
Brassicaceae	(16)	(17)	(17)	(16)	(14)	(14)	(12)	8	10	7
Ericaceae	(17)	(16)	(12)	9	8	6	7	7	(11)	(11)

Примечание. В скобках указаны ранги семейств, не входящих в ведущую десятку семейств по числу видов в соответствующих флорах, учитывать которых, однако, необходимо из математических соображений.

Все это справедливо, однако, для тех случаев, когда ошибка репрезентативности имеет реальный смысл, будучи сформированной за счет естественной изменчивости признака в генеральной совокупности, т. е. когда исследователь имеет дело с выборочным показателем, являющимся приближенной оценкой генерального параметра.

Главная часть флористического спектра, с которой мы имеем дело в рассматриваемой ситуации, не является выборкой из генеральной совокупности (т. е. полного флористического спектра): положение (ранг) семейства или рода в данное время и в

Таблица 76. Матрица коэффициентов ранговой корреляции τ Кендэла, характеризующих степень сходства структуры ведущих по числу видов семейств флор Курильских о-вов

Флоры	1	2	3	4	5	6	7	8	9	10
1	—	0,68	0,59	0,59	0,38	0,50	0,44	0,29	0,35	0,22
2	0,68	—	0,71	0,62	0,42	0,59	0,54	0,32	0,35	0,31
3	0,59	0,71	—	0,76	0,59	0,68	0,59	0,38	0,41	0,40
4	0,59	0,62	0,76	—	0,62	0,82	0,73	0,69	0,60	0,49
5	0,38	0,42	0,59	0,62	—	0,68	0,68	0,59	0,61	0,52
6	0,50	0,59	0,68	0,82	0,68	—	0,85	0,62	0,56	0,49
7	0,44	0,54	0,59	0,73	0,68	0,85	—	0,65	0,50	0,54
8	0,29	0,32	0,38	0,60	0,59	0,62	0,65	—	0,62	0,57
9	0,35	0,35	0,41	0,60	0,61	0,56	0,50	0,62	—	0,75
10	0,22	0,31	0,40	0,49	0,52	0,49	0,54	0,57	0,75	—

Примечание. Нумерация флор дана в соответствии с табл. 75.

данной флоре строго фиксировано и не варьирует в точном значении этого слова, ибо исторические сдвиги в составе флоры или неполнота сведений, из-за которой ранг таксона может быть определен неверно, к формированию ошибки репрезентативности отношения не имеют. Фактически сравниваемые наборы ведущих семейств или родов представляют каждый генеральную совокупность. Поэтому вычисление ошибок репрезентативности в данном случае лишено смысла, а рассмотренные показатели оценке на достоверность не подлежат.

Перейдем к вопросу о группировке флор по признакам систематической структуры. Исходным материалом теперь будет служить матрица рассчитанных коэффициентов (τ , ρ_s или r). Удобно использовать для ее анализа построение дендрита способом «максимального корреляционного пути» с последую-

щим выделением корреляционных плеяд наиболее сходных по своей систематической структуре флор. Техника этой работы была описана в разделе 7.3.2.

Остается привести соответствующий пример. Исследовалась систематическая структура флор 10 Курильских о-вов [Шмидт, 1980], в частности структура ведущих по числу видов семейств (табл. 75). Рассчитанные значения коэффициентов ранговой корреляции τ Кендэла сведены в матрицу (табл. 76). Полученные дендрит и корреляционные плеяды показаны на рис. 39. Как видно, этот дендрит связывает флоры Курильских о-вов в порядке их расположения с юга на север, что говорит о постепенном изменении систематической структуры в этом направлении. Исключение составляет флора о-ва Симушир, которая несколько изолирована по рассматриваемому признаку от флор других средних Курильских о-вов. При повышении уров-

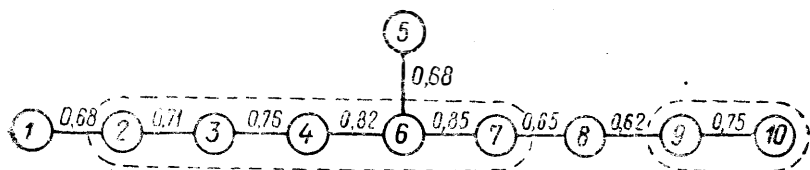


Рис. 39. Дендрит и корреляционные плеяды, отражающие степень сходства структуры ведущих по числу видов семейств во флорах Курильских о-вов.

Цифры у линий — коэффициенты ранговой корреляции Кендэла. Нумерация флор дана в табл. 75.

ня связи в дендрите до $\tau \geq 0,63$ вычленяется плеяда, состоящая из двух северных (Парамушир и Шумшу) островов и наиболее обособленная от других островов по структуре ведущих по числу видов семейств. На уровне $\tau \geq 0,66$ отщепляется Матуа, а при $\tau \geq 0,69$ — Шикотан и Симушир. Эти острова, хотя и примыкают по данному признаку к своим группам, но обладают наименьшим сходством с ними. После их отщепления помимо указанной выше северной плеяды остается смешанная плеяда, состоящая из трех южных (Кунашир, Итуруп, Уруп) и двух средних (Кетой, Расшуа) островов, дальнейший распад которой при повышении уровня связи в дендрите происходит только путем вычленения отдельных (сначала южных) островов.

Нет надобности доказывать, что подобного рода анализ, выполненный по всем признакам систематической структуры сравниваемых флор, приводит к интересным и важным выводам [Шмидт, 1980].

8.3.3. Информационные индексы сложности систематической структуры флор. Изложенные в разделе 8.3.2 методы, основанные на сравнении головной части флористических спектров, хорошо улавливают основные черты сходства или различия систематической структуры сравниваемых флор, но не принимают во внимание те семейства (или роды), которые представлены во флорах всего одним или несколькими видами. Ясно, что при таком подходе какая-то часть информации о систематической структуре флор теряется. Это видно также из следующих сообщений.

Во-первых, даже содержащие всего один вид семейства нередко свидетельствуют о тех или иных особенностях флорогенеза и могут быть далеко не безразличны для флориста. Во-вторых, одно и то же число видов при одном и том же числе семейств в сравниваемых флорах может быть распределено между семействами по-разному.¹ Это распределение теоретически может быть равномерным (все семейства флоры содержат одинаковое число видов), но на практике оно всегда неравномерно, причем чаще всего имеет место резкая асимметрия такого распределения, а именно: число бедных видами семейств намного превышает число семейств, богатых видами (т. е. головную часть флористического спектра). Соотношение «бедных» и «богатых» видами семейств во флорах может быть различным, но это также не учитывается изложенными выше приемами.

Приступая к рассмотрению метода оценки степени сходства полных флористических спектров, необходимо ввести новое понятие о *сложности систематической структуры* (или сложности флористического спектра), которое включает в себя как уровень флористического богатства (число таксонов того или иного ранга во флоре), так и степень выравненности распределения таксонов низшего ранга по таксонам высшего ранга. Иначе говоря, «сложность систематической структуры» мы будем рассматривать как функцию флористического богатства и выравненности распределения.

В геоботанике и экологии уже применяются показатели, учитывающие признаки, аналогичные тем, которые входят в состав сформулированного выше понятия. Речь идет об использовании в интересующих нас целях одной из информационных мер разнообразия — известной функции Шеннона—Уивера (или H -функции), имеющей вид

$$H = - \sum p_i \log_2 p_i,$$

причем, применительно к рассматриваемой ситуации, за вероятность p_i принимается наблюдаемая встречаемость данного

¹ Разумеется, то же самое может быть сказано о распределении числа родов по семействам и числа видов по родам.

числа таксонов низшего ранга в составе каждого таксона высшего ранга, т. е.

$$p_i = \frac{n_i}{n},$$

где n_i — число таксонов низшего ранга в составе таксона высшего ранга; n — общее число таксонов низшего ранга во флоре. Для расчета H -функции можно использовать логарифмы с любым основанием, но для удобства рекомендуем пользоваться логарифмами с основанием 2 (Приложение X).

Теоретически H -функция принимает максимальное значение тогда, когда имеет место полная выравненность распределения, что соответствует наибольшей сложности систематической структуры (все семейства флоры содержат одинаковое число видов). Для каждого числа таксонов высшего ранга, свойственных данной флоре, существует свое значение H_{\max} , достигаемое при равномерном распределении любого числа таксонов низшего ранга. Короче говоря, значение H_{\max} зависит от числа таксонов высшего ранга и математически определяется формулой $H_{\max} = \log_2 N$, где N — число таксонов высшего ранга.

Отношение

$$H' = \frac{H}{H_{\max}}$$

в известной мере может служить показателем приближения распределения к максимальной выравненности. Значение $H=0$ практически не имеет места, ибо при этом все виды флоры должны были бы относиться к одному семейству (или роду). Минимальное значение H -функции следует определять по формуле

$$H_{\min} = \frac{1}{n} [n \log_2 n - (n - N + 1) \log_2 (n - N + 1)],$$

где n — число таксонов низшего, а N — высшего ранга. Эта формула соответствует тому крайнему случаю, когда, например, одно семейство содержит $n - N + 1$ видов, а все остальные $N - 1$ семейств — по одному виду каждое. Таким образом, чем «контрастнее» распределение, тем меньше значение принимает H -функция.

Итак, H -функцию можно считать показателем сложности систематической структуры флор, отражающей как уровень флористического богатства, так и степень выравненности распределения таксонов низшего ранга по таксонам высшего ранга.

Для сравнения этих показателей, рассчитанных для двух флор, можно использовать t -критерий Стьюдента:

$$t = \frac{H_1 - H_2}{\sqrt{\text{var } H_1 + \text{var } H_2}},$$

где $\text{var } H_1$ и $\text{var } H_2$ — дисперсии H , получаемые по формуле

$$\text{var } H = \frac{1}{N} (\sum p_i \cdot \log^2 p_i - H^2).$$

Значимость различий оценивается при числе степеней свободы

$$v = \frac{(\text{var } H_1 + \text{var } H_2)^2}{\frac{\text{var } H_1^2}{N_1} + \frac{\text{var } H_2^2}{N_2}},$$

где H_1 и H_2 — функции, рассчитанные для первой и второй флор; $\text{var } H_1$ и $\text{var } H_2$ — их дисперсии; N_1 и N_2 — число таксонов более высокого ранга в первой и второй флорах.

Перейдем к рассмотрению фактического примера сравнения полных флористических спектров посредством интересующей

Таблица 77. Число таксонов в конкретных флорах

Флоры	Число		
	видов	родов	семейств
Бухта М. Прончищевой	96	47	17
Шойна	230	137	55
Кривые Озера	435	215	65
Любытино	538	296	86

нас методики. Исходным материалом послужили литературные данные о видовом составе четырех конкретных флор, расположенных в разных природных зонах и образующих широтный ряд в направлении с севера на юг: 1) окрестности бухты Марии Прончищевой — северо-восточный Таймыр, Арктическая флористическая область, подзона арктических тундр; 2) окрестности пос. Шойна — п-ов Канин, граница между Арктической и Бореальной флористическими областями, подзона кустарниковых тундр; 3) окрестности пос. Кривые Озера — Архангельская область, Бореальная флористическая область, подзона северной тайги; 4) окрестности пос. Любытино — Новгородская область, Среднеевропейская флористическая область, подзона южной тайги.

Как видно из табл. 77, число видов S , родов G и семейств F по мере продвижения от самой северной к самой южной конкретной флоре увеличивается в несколько раз, и все флоры достаточно сильно отличаются друг от друга по этим показателям. Распределение таксонов более низкого ранга по таксонам более высокого ранга в конкретных флорах показано в табл. 78.

Таблица 78. Распределение числа таксонов более низкого ранга (n_i) по таксонам более высокого ранга в конкретных флорах

n_i	Число семейств с n_i видами				Число семейств с n_i родами				Число родов с n_i видами			
	Бухта М. Проничевой	Шойна	Кривые Озера	Любытино	Бухта М. Проничевой	Шойна	Кривые Озера	Любытино	Бухта М. Проничевой	Шойна	Кривые Озера	Любытино
1	2	20	19	29	4	32	36	45	27	94	137	266
2	3	13	14	17	8	11	10	19	12	27	39	35
3	4	5	7	5	2	1	4	3	4	5		22
4	2	1	—	8	—	1		4	1	4	11	13
5	1	5	4	6	2	6	3	2	—	4	5	7
6	—	2	2	1	—	—	1	2	1	1	4	5
7	1	—	3	3	—	1	1	2	—	—	2	3
8	—	—	2	1	—	1	3	—	—	—	2	2
9	—	5	1	1	—	—	1	—	—	1	1	1
10	—	—	2	1	—	—	—	1	1	—	—	1
11	1	—	1	2	1	—	2	1	—	—	—	—
12	—	—	—	1	—	—	1	2	—	—	1	—
13	—	1	1	1	—	—	—	—	1	—	—	—
14	1	—	—	—	—	1	—	—	—	—	—	—
15	1	—	—	—	—	—	—	2	—	—	1	—
16	1	—	2	—	—	—	—	—	—	1	—	—
17	—	—	—	—	—	1	—	—	—	—	—	—
18	—	—	—	1	—	—	—	1	—	—	—	—
19	—	—	1	1	—	—	1	—	—	—	—	—
21	—	1	1	—	—	—	—	—	—	—	—	—
22	—	1	—	2	—	—	—	—	—	—	—	—
23	—	—	1	1	—	—	—	—	—	—	—	1
25	—	—	1	—	—	—	—	1	—	—	—	—
26	—	—	—	1	—	—	—	—	—	—	—	—
27	—	1	—	—	—	—	—	—	—	—	—	—
28	—	—	—	—	—	—	—	—	—	—	1	—
29	—	—	—	—	—	—	1	—	—	—	—	—
32	—	—	—	2	—	—	—	—	—	—	—	—
34	—	—	—	—	—	—	—	1	—	—	—	—
36	—	—	1	—	—	—	—	—	—	—	—	—
40	—	—	1	—	—	—	—	—	—	—	—	—
41	—	—	—	1	—	—	—	—	—	—	—	—
49	—	—	1	—	—	—	—	—	—	—	—	—
62	—	—	—	1	—	—	—	—	—	—	—	—

Таблица 79. Алгоритм для расчета H -функций и их дисперсий на примере флористического спектра структуры семейств по числу видов во флоре окрестностей бухты М. Прончищевой

Число видов в семействе, n_i	Число семейств с n_i видами, N_i	$p_i = \frac{n_i}{n}$	Расчет H_{SF}			Расчет $\text{var } H_{SF}$		
			$-p_i \log_2 p_i$	$N_i(-p_i \log_2 p_i)$	$\log_2 p_i$	$\log_2^2 p_i$	$p_i \log_2^2 p_i$	$N_i(p_i \log_2^2 p_i)$
1	2	0,0104	0,0685	0,1370	6,5805	43,3825	0,4512	0,9024
2	3	0,0208	0,1162	0,3486	5,5865	31,2094	0,6492	1,9475
3	4	0,0312	0,1561	0,6244	5,0032	25,0321	0,7810	3,1240
4	2	0,0417	0,1913	0,3826	4,5875	21,0454	0,8776	1,7552
5	1	0,0521	0,2221	0,2221	4,2630	18,1728	0,9468	0,9468
7	1	0,0729	0,2754	0,2754	3,7778	14,2716	1,0404	1,0404
	1	0,1146	0,3581	0,3581	3,1249	9,7643	1,1190	1,1190
14	1	0,1458	0,4050	0,4050	2,7778	7,7100	1,1250	1,1250
15	1	0,1562	0,4184	0,4184	2,6786	7,1750	1,1207	1,1207
16	1	0,1667	0,4309	0,4309	2,5849	6,6816	1,1138	1,1138
$n=96$	$N=17$	—	—	$H_{SF}=3,6025$	—	—	—	$\Sigma = 14,1948$

Примечание. После расчетов, выполненных в этой таблице, дисперсия показателя H_{SF} находится по формуле $\text{var } H_{SF} = \frac{1}{17} (14,1948 - 3,6025^2) = 0,0716$.

Для каждой из сравниваемых флор были рассчитаны следующие информационные индексы: H_{SF} — показатель сложности структуры семейств по числу видов, H_{GF} — показатель сложности структуры семейств по числу родов и H_{GS} — показатель сложности структуры родов по числу видов. Порядок расчета величин H и их дисперсий $\text{var } H$ показан в табл. 79.

Рассчитанные показатели сведены в табл. 80. Значения t -критерия Стьюдента с соответствующими числами степеней свободы, необходимые для оценки достоверности различий H -функций, приведены в табл. 81.

Таблица 80. Информационные показатели сложности систематической структуры изученных конкретных флор, пределы их изменчивости и дисперсии

Флоры	Индексы	H	H_{\max}	H_{\min}	$H' = \frac{H}{H_{\max}}$	$\text{var } H$
Бухта М. Прончищевой	H_{SF}	3,6025	4,0877	1,3169	0,8813	0,0716
	H_{GF}	3,7095	4,0877	2,2871	0,9075	0,0702
	H_{SG}	5,0402	5,5551	3,6459	0,9073	0,0384
Шойна	H_{SF}	4,9838	5,7821	2,1376	0,8619	0,0406
	H_{GF}	5,1210	5,7821	3,0964	0,8857	0,0378
	H_{SG}	6,6674	7,0987	5,1671	0,9392	0,0126
Кривые Озера	H_{SF}	5,0375	6,0229	1,4855	0,8364	0,0373
	H_{GF}	5,1049	6,0229	2,6646	0,8476	0,0370
	H_{SG}	7,1748	7,7488	4,8088	0,9259	0,0095
Любытино	H_{SF}	5,3468	6,4259	1,6423	0,8321	0,0322
	H_{GF}	5,4566	6,4259	2,7058	0,8492	0,0610
	H_{SG}	7,7985	8,2103	5,4927	0,9498	0,0036

Из полученных данных можно извлечь следующие основные выводы.

1. Информационные индексы довольно удачно отражают сложность систематической структуры флор, принимая наименьшие значения у наиболее бедной таксонами конкретной флоры окрестностей бухты М. Прончищевой, обладающей меньшей выравненностью распределения таксонов низшего ранга по таксонам высшего ранга (табл. 77, 78, 80). По мере продвижения к югу значения индексов последовательно возрастают и оказываются наибольшими для самой южной и наиболее богатой таксонами флоры окрестностей пос. Любытино, которая характеризуется и большей выравненностью флористических спектров.

2. Пределы изменчивости информационных индексов (табл. 80) расширяются и сдвигаются в большую сторону также в направлении с севера на юг. В том же порядке уменьшаются дисперсии индексов H_{SF} и H_{SG} , но дисперсия $\text{var } H_{GF}$ ведет себя более неопределенно, и закономерность ее изменения в географическом аспекте нуждается в дальнейшем изучении.

3. По значениям индексов H_{SF} и H_{GF} арктическая флора окрестностей бухты М. Прончищевой достоверно отличается от всех других флор (табл. 81). Это связано не только с есте-

Таблица 81. Оценка достоверности различия H -функций с помощью t -критерия Стьюдента

Флоры	1	2	3	4
Сравнение H_{SF}				
1. Бухта М. Прончищевой	—	38	37	34
2. Шойна	4,12	—	118	123
3. Кривые Озера	4,35	0,19*	—	145
4. Любытино	5,41	1,35*	1,17*	—
Сравнение H_{GF}				
1. Бухта М. Прончищевой	—	37	37	52
2. Шойна	4,30	—	119	141
3. Кривые Озера	4,26	0,06*	—	150
4. Любытино	4,82	1,07*	1,12*	—
Сравнение H_{SG}				
1. Бухта М. Прончищевой	—	80	72	56
2. Шойна	7,21	—	325	238
3. Кривые Озера	9,75	3,41	—	430
4. Любытино	13,46	8,89	5,45	—

Примечание. Слева от диагонали помещены рассчитанные значения t -критерия, справа — соответствующие числа степеней свободы. Случаи недостоверного различия H -функций при 0,1%-ном уровне существенности отмечены звездочкой.

ственной обедненностью таксономического состава этой флоры, но и с существенными отличиями ее флористических спектров. Например, в число пяти ведущих по числу видов семейств этой флоры входят сем. Brassicaceae (1-е место) и сем. Saxifragaceae (3-е место), не попадающие в пятерку наиболее крупных семейств во всех остальных флорах. Напротив, сем. Asteraceae и Сурегасеae, занимающие 1—3-е места в других флорах, не попадают здесь в ведущую пятерку.

Остальные флоры (Шойна, Кривые Озера, Любытино) по значениям информационных индексов H_{SF} и H_{GF} достоверно

друг от друга не отличаются (табл. 81). Обусловлено это прежде всего, незначительными различиями состава их флористических спектров. Однако более северная из них флора окрестностей пос. Шойна несколько отличается порядком расположения ведущих семейств и значительно беднее видовым составом, что находит свое выражение в значениях упомянутых индексов (табл. 80). Что же касается конкретных флор окрестностей пос. Кривые Озера и Любытино, то их флористические спектры наиболее сходны. Так, первые четыре места по числу видов в обеих флорах в одинаковом порядке занимают сем. *Asteraceae*, *Rosaceae*, *Syringaceae* и *Rosaceae*. Эти же флоры менее других отличаются друг от друга показателями флористического богатства (табл. 77). Сказанному соответствуют и наиболее близкие значения индексов H_{SF} и H_{GF} , рассчитанные для этих флор (табл. 80).

4. По значениям информационного индекса H_{SG} все четыре флоры достоверно отличаются друг от друга (табл. 81). Повидимому, это свидетельствует о большей оригинальности родового флористического спектра по сравнению со спектрами семейств. Возможно, что этот показатель несет больше информации об индивидуальных особенностях флор, связанных с историей флорогенеза на той или иной территории.

8.3.4. Коэффициенты сходства систематического состава флор (флористических списков). Таких коэффициентов, многие из которых находят широкое применение в геоботанике и зоогеографии, существует несколько. В. И. Василевичем [1969] показано, что каждый из них имеет свои преимущества и недостатки, а цифры, получаемые с их помощью, несопоставимы, так как математическая основа этих коэффициентов различна. Мы остановимся на рассмотрении только некоторых коэффициентов,¹ заслуживающих наибольшего внимания при сравнении флор.

Все коэффициенты сходства прямо или косвенно включают в себя число таксонов (будем говорить о видах) в сравниваемых флорах. Поскольку число видов зависит от площади, то и в данном случае следует придерживаться принципа приблизительного равенства территорий.

Одним из наиболее простых и понятных показателей флористического сходства является коэффициент Жаккара, предложенный им в 1901 г.:

$$K_j = \frac{c}{a + b - c},$$

где a — число видов в одной флоре; b — число видов в другой

¹ Все они пригодны для анализа как видового, так и родового сходства флор, а также для сравнения флор по географическим элементам. Для краткости изложения мы будем везде говорить о сравнении видового состава.

флоре; c — число видов, общих для двух флор. Пределы этого коэффициента от 0 до 1, причем $K_j=1$ означает полное сходство флор (абсолютное совпадение списков), а $K_j=0$ означает, что флоры не имеют ни одного общего вида.

Более употребительной (особенно в геоботанике) является в настоящее время другая формула, называемая иногда коэффициентом Сёрнсена, но открытая еще в 1913 г. Чекановским, почему правильнее именовать ее коэффициентом Сёрнсена — Чекановского:

$$K_{sc} = \frac{2c}{a+b},$$

где обозначения и пределы те же самые, что и в формуле Жакара.

Далее следует упомянуть предложенную в 1940 г. Экманом «зоогеографическую формулу»:

$$\mathcal{E} = \frac{A+B}{C},$$

где A — число видов, встречающихся в первой флоре, но отсутствующих во второй; B — число видов, встречающихся во второй флоре, но отсутствующих в первой; C — число видов, встречающихся в обеих флорах. Неудобство этого коэффициента состоит в том, что при полном сходстве двух флор он имеет своим нижним пределом 0, но не имеет верхнего предела, так как зависит от числа видов во флорах.

Стугрен и Радулеску в 1961 г. попытались освободить формулу Экмана от ее основного недостатка — отсутствия верхнего предела — и предложили свой коэффициент:

$$\rho_{sr} = \frac{X+Y-Z}{X+Y+Z}.$$

где обозначения X , Y , Z соответствуют обозначениям A , B и C в формуле Экмана. Данный коэффициент обладает следующей особенностью: варьируя в пределах от -1 до $+1$, он в пределах от -1 до 0 указывает на сходство, а в пределах от 0 до $+1$ — на различие (дискриминацию) флор. Следовательно, если при сравнении серии флор все полученные значения $\rho_{sr} < 0$, то за меру сходства можно брать абсолютные значения коэффициента.

Более сложен показатель Престона, предложенный им в 1962 г.:

$$N_{a+b}^{\frac{1}{z}} = N_a^{\frac{1}{z}} + N_b^{\frac{1}{z}},$$

где N_a и N_b — числа видов в первой и второй флорах; N_{a+b} — общее число видов в двух флорах (т. е. число видов в объединенном списке); z — константа, выведенная Престоном для логарифмически нормального распределения. Поскольку значения

N_a , N_b и N_{a+b} находятся при составлении флористических списков, то вся задача сводится к нахождению величины z , которая варьирует от 0 до 1.

Престон показал, что при $z=0,27$ флоры (фауны) статистически однородны. Чем меньше эмпирическое значение z относительно критической величины 0,27, тем однороднее флоры, чем больше — тем сильнее их различие. Поэтому величина z может считаться мерой различия, а величина $1-z$ — мерой сходства двух флор. При пользовании этим коэффициентом достаточно подсчитать

$$\frac{N_a}{N_{a+b}} \quad \text{и} \quad \frac{N_b}{N_{a+b}},$$

после чего величину z находят по специальной таблице [Preston, 1962; Старобогатов, 1970].

В. И. Василевич [1969] для измерения степени различия флористических списков двух растительных сообществ предложил использовать «расстояние» между ними в двумерной системе координат, задаваемое формулой

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

где x_1 и x_2 — значения одного, а y_1 и y_2 — другого геоботанического признака в двух описаниях. Если учитывается только один признак (наличие или отсутствие вида в описании), то подкоренное выражение в приведенной выше формуле представляет собой просто число видов, по которому описания различаются. Применительно к сравнению флор это означает, что

$$D = \sqrt{a + b - 2c},$$

где a — число видов в одной флоре; b — число видов в другой флоре; c — число общих для двух флор видов. Нижним пределом этого показателя служит 0, означающий полное совпадение двух флористических списков, а верхний предел зависит от числа видов во флорах, что вызывает некоторое неудобство при использовании данного коэффициента в его «чистом» виде.

Можно предложить следующее. Максимальное различие достигается в том случае, когда флоры не совпадают ни по одному виду:

$$D_{\max} = \sqrt{a + b}.$$

Если теперь значение D выразить в долях от D_{\max} и отнять от единицы, то мы получим меру сходства двух флор в виде выражения

$$D' = 1 - \frac{D}{D_{\max}} = 1 - \frac{\sqrt{a + b - 2c}}{\sqrt{a + b}} = 1 - \sqrt{1 - \frac{2c}{a + b}}.$$

Эту величину, варьирующую в пределах от 0 до 1, значительно

удобнее использовать для дальнейших сопоставлений при сравнении флор.

Возникает естественный вопрос: каким же из рассмотренных коэффициентов сходства предпочтительнее пользоваться при сравнении флористических списков? В категорической форме на этот вопрос в настоящее время ответить трудно. Рассмотрим, однако, конкретный пример, касающийся сравнения 7 флористических районов, выделенных на территории Псковской области [Баранова, Миняев, Шмидт, 1971].

Исходные данные для расчета коэффициентов сходства между флорами этих районов приведены в табл. 82, а сами коэффициенты — в табл. 83—85.

Таблица 82. Число общих видов (*c*) и численность суммарного видового списка (*d*) для флористических районов Псковской обл.

Показатели и номера районов		<i>d</i>						
		1	2	3	4	5	6	7
<i>c</i>	1	853	1043	880	866	907	937	942
	2	813	1003	1021	1012	1027	1050	1054
	3	689	698	716	749	821	865	861
	4	673	677	653	686	809	860	855
	5	720	750	669	651	774	881	881
	6	757	794	692	667	734	841	930
	7	749	787	693	669	731	749	838

Примечание. По диагонали выделены числа видов в каждом районе.

Несмотря на то, что коэффициенты различны по своей величине, на их основе (за исключением коэффициента Престона) получаются совершенно одинаковые дендриты и корреляционные плеяды (рис. 40), что свидетельствует о допустимости применения любого из них. Однако разрешающая способность коэффициентов неодинакова, в чем можно убедиться, сопоставив разницу их максимальных значений в табл. 83—85. По всей видимости, наибольшим дифференцирующим свойством обладает коэффициент Стургена—Радулеску (указанная разница в нашем примере составляет 0,40), которому и можно отдать предпочтение при сравнении серии весьма близких флор, когда все значения этого показателя заведомо меньше нуля.

Адекватным этому коэффициенту (отличающимся только масштабом) является линейно связанный с ним коэффициент Жаккара. С точки зрения теории множеств именно коэффициент Жаккара обладает наибольшей математической корректностью [Тамарин, Шмидт, 1975]. Применение коэффициента Сёренса-

Таблица 83. Коэффициенты сходства Жаккара (K_j) и Сёрнсена—Чекановского (K_{sc}), рассчитанные для флористических районов Псковской обл.

Показатели и номера районов		K_j						
		1	2	3	4	5	6	7
K_{sc}	1	—	0,78	0,78	0,78	0,79	0,81	0,80
	2	0,88	—	0,68	0,67	0,73	0,76	0,75
	3	0,88	0,81	—	0,87	0,81	0,80	0,80
	4	0,87	0,80	0,93	—	0,80	0,78	0,78
	5	0,89	0,84	0,90	0,89	—	0,83	0,83
	6	0,89	0,86	0,89	0,87	0,91	—	0,81
	7	0,89	0,85	0,89	0,88	0,91	0,89	—

Таблица 84. Коэффициенты сходства Экмана (1—Э) и Стургена—Радулеску (ρ_{sr}), рассчитанные для флористических районов Псковской обл.

Показатели и номера районов		1—Э						
		1	2	3	4	5	6	7
$ \rho_{sr} $	1	—	0,72	0,72	0,71	0,74	0,76	0,74
	2	0,56	—	0,54	0,51	0,63	0,68	0,66
	3	0,57	0,37	—	0,85	0,77	0,75	0,76
	4	0,55	0,34	0,74	—	0,76	0,71	0,72
	5	0,59	0,46	0,63	0,61	—	0,80	0,79
	6	0,62	0,51	0,60	0,55	0,67	—	0,76
	7	0,59	0,49	0,61	0,56	0,66	0,61	—

Таблица 85. Коэффициенты Престона (1—z) и расстояние Василевича (D'), рассчитанные для флористических районов Псковской обл.

Показатели и номера районов		D'						
		1	2	3	4	5	6	7
1—z	1	—	64,8	65,1	64,6	66,1	67,4	66,2
	2	0,86	—	56,7	55,5	60,5	62,9	61,9
	3	0,87	0,83	—	73,8	68,1	66,6	67,1
	4	0,88	0,85	0,91	—	67,1	64,5	65,1
	5	0,85	0,86	0,86	0,87	—	69,8	69,5
	6	0,86	0,85	0,87	0,88	0,88	—	67,2
	7	0,84	0,83	0,88	0,88	0,88	0,85	—

на — Чекановского оправдывается опытом, но он не имеет каких-либо преимуществ перед коэффициентом Жаккара. Что же касается показателя Престона, то по сравнению с другими коэффициентами он обладает слабыми дифференцирующими возможностями (диапазон его варьирования в табл. 85 составляет всего 0,08) и не может быть рекомендован к использованию при сравнении близких по составу флор.

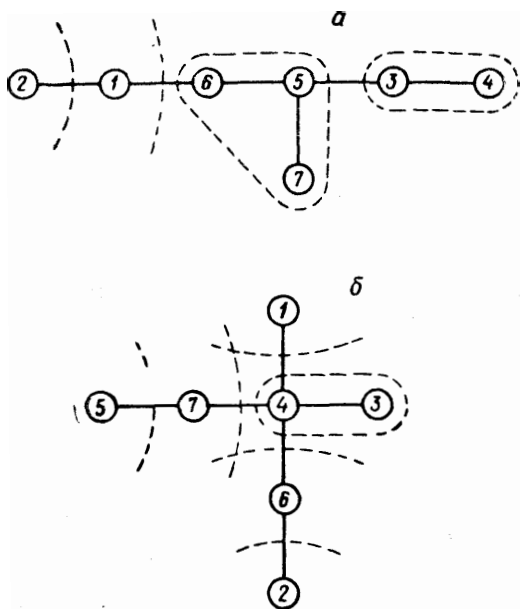


Рис. 40. Дендриты и корреляционные плеяды, отражающие степень сходства видового состава флористических районов Псковской области. *a* — на основе использования коэффициентов Жаккара, Сёрнсена—Чекановского, Экмана, Сугрена—Радулеску и Василевича; *б* — на основе показателя Престона. 1—7 — флористические районы.

Поскольку рассматриваемые коэффициенты несопоставимы, но используются разными авторами, возникает вопрос о возможности их перевода друг в друга. Было показано [Тамарин, Шмидт, 1975], что уравнение связи между интересующими нас коэффициентами имеет вид

$$\frac{1}{K_j} + 1 = \frac{2}{K_{sc}} = \mathcal{E} + 2 = \frac{2}{1 - \rho_{sr}} + 1 = \frac{2}{1 - (1 - D')^2},$$

откуда легко получить точные формулы для перевода одних коэффициентов в другие:

$$K_j = \frac{K_{sc}}{2 - K_{sc}} = \frac{1}{\mathcal{E} + 1} = \frac{1 - \rho_{sr}}{2} = \frac{1 - (1 - D')^2}{1 + (1 - D')^2};$$

$$K_{sc} = \frac{2K_j}{1+K_j} = \frac{\vartheta+2}{\vartheta+2} = \frac{2(1-\rho_{sr})}{3-\rho_{sr}} = 1 - (1-D')^2;$$

$$\vartheta = \frac{1-K_j}{K_j} = \frac{2(1-K_{sc})}{K_{sc}} = \frac{1+\rho_{sr}}{1-\rho_{sr}} \frac{2(1-D')^2}{1-(1-D')^2};$$

$$\rho_{sr} = 1 - 2K_j = \frac{\vartheta-1}{\vartheta+1} = \frac{2-3K_{sc}}{2-K_{sc}} = \frac{1-3(1-D')^2}{1+(1-D')^2};$$

$$D' = 1 - \sqrt{\frac{1-K_j}{1+K_j}} = 1 - \sqrt{1-K_{sc}} =$$

$$= 1 - \sqrt{\frac{\vartheta}{\vartheta+2}} = 1 - \sqrt{\frac{1+\rho_{sr}}{3-\rho_{sr}}}.$$

Из этих формул видно также, что связь между всеми коэффициентами сходства, кроме K_j и ρ_{sr} , является нелинейной.

В заключение раздела следует напомнить, что все перечисленные коэффициенты сходства, базирующиеся на абсолютном числе видов в сравниваемых флорах, не имеют ошибок репрезентативности и оценке на достоверность не подлежат.

8.4. ГЕОГРАФИЧЕСКАЯ ИЗМЕНЧИВОСТЬ ФЛОРИСТИЧЕСКИХ ПОКАЗАТЕЛЕЙ И ВОЗМОЖНОСТЬ ИХ ПРОГНОЗИРОВАНИЯ

В общей форме некоторые закономерности географической изменчивости флористических показателей неоднократно отмечались исследователями, начиная с первой четверти прошлого века. Давно известно, например, что по мере продвижения от полюсов к экватору возрастает богатство флор, причем число видов увеличивается значительно быстрее, нежели число родов и особенно семейств. А. Декандоль [De Candolle, 1855] указывал на снижение «удельного веса» однодольных в северном полушарии в направлении с севера на юг, а также в странах с более континентальным климатом. Горные флоры тропических областей и флоры изолированных океанических островов характеризуются повышенной ролью спайнолепестных [Толмачев, 1974], тогда как в Арктике эта группа представлена слабо. Однако в целом сведений об изменчивости этих и других флористических показателей накоплено еще мало.

Словесное описание подобного рода закономерностей на современном уровне развития науки не может быть признано достаточным. Любая закономерность должна быть проверена и описана математически, что, кстати, открывает возможность прогнозирования тех или иных показателей в неизученных районах.

Наиболее объективным образом указанные задачи могут быть решены на основе изучения конкретных флор. В качестве примера рассмотрим последовательность действий и биометри-

ческие приемы, использованные при изучении географической изменчивости 10 показателей 102 конкретных и близких к ним флор на территории европейской части СССР [Шмидт, 1980].

Подготовительный этап работы состоял в унификации флористических списков в отношении трактовки объема таксонов. Эта операция совершенно необходима для обеспечения сопоставимости флористических списков, опубликованных разными авторами в разное время. После этого для каждой флоры были рассчитаны интересующие нас флористические показатели (параметры флор), представляющие собой исходные данные для

Таблица 86. Пределы изменчивости параметров конкретных и близких к ним флор на территории европейской части СССР (исключая Кавказ)

Параметры	Флористические области		
	Арктическая	Бореальная	Средне-европейская
Число видов	22—299	278—809	538—984
Число родов	17—152	142—320	294—451
Число семейств	9—57	50—96	75—103
Среднее число видов в семействе	3,3—6,2	5,6—8,1	6,3—10,5
Среднее число родов в семействе	1,9—3,0	2,8—3,9	3,4—4,7
Среднее число видов в роде . . .	1,7—2,1	1,8—2,3	1,8—2,3
Процент однодольных от числа цветковых	24—34	22—36	19—27
Процент спайнолепестных от числа двудольных	0—42	36—43	37—45
Процент видов в 10 ведущих семействах	57—100	51—60	49—63
Отношение числа видов семейств Asteraceae/Poaceae	0—1,2	0,6—1,5	0,9—1,5

дальнейшей работы. Эти данные, равно как и составленные на их основе карты с изолиниями (изофлорами), соединяющими пункты с одинаковыми значениями флористических параметров, приведены в другой работе [Шмидт, 1980]. Поэтому здесь мы ограничимся перечнем изученных флористических показателей с указанием пределов их изменчивости на территории европейской части СССР (табл. 86).

Основные тенденции географической изменчивости флористических показателей на указанной территории вкратце состоят в следующем.

Показатели флористического богатства (число видов, родов и семейств) обнаруживают четкую закономерность увеличения

в направлении с севера на юг. Показателям систематического разнообразия в/с и р/с свойственна такая же тенденция, но показатель в/р варьирует в пределах европейской части СССР слабо и неопределенным образом. Процент видов однодольных слабо уменьшается с севера на юг. Процент спайнолепестных во флорах к югу увеличивается, а процент видов, охватываемых 10 ведущими семействами, в том же направлении, наоборот, па-

Таблица 87. Результаты дисперсионного анализа влияния географической широты на количественные показатели конкретных и близких к ним флор европейской части СССР

Показатели	Число видов	Число родов	Число семейств	Среднее число видов в семействе (в с)	Среднее число родов в семействе (р/с)	Среднее число видов в роде (в/р)	Процент однодольных от числа цветковых	Процент спайнолепестных от числа двудольных	Процент видов в 10 ведущих семействах	Отношение числа видов семейств Asteraceae/Roaceae
$r-1$	6	6	6	6	6	6	6	5	5	5
$n-r$	95	81	81	84	83	84	81	83	83	83
C_y	5700918	955013	47422	200	30,0	2,01	992,5	3522	5970	8,74
C_x	4968730	844126	38853	145	23,7	0,41	353,1	2709	4557	6,11
C_z	732188	110887	8569	55	6,3	1,60	639,4	813	1413	2,6)
σ_x^2	828122	140688	6476	24,17	3,95	0,07	70,6	541,8	911,4	1,23
σ_z^2	7707	1369	106	0,66	0,08	0,02	7,9	9,8	17,0	0,03
η_{1x}^2	0,87	0,88	0,82	0,72	0,79	0,20	0,36	0,77	0,76	0,70
η_{1z}^2	0,13	0,12	0,18	0,28	0,21	0,80	0,64	0,23	0,24	0,30
F	107,5	102,8	61,1	36,6	49,4	3,5	8,9	55,3	53,6	41,0

Примечание. При обработке данных использованы схема вычислений и система обозначений, рекомендованные Н. А. Плохинским [1970]: r — число градаций фактора (широты) n — число флор; C_y , C_x , C_z — общая, факториальная и случайная дисперсии; σ_x^2 , σ_z^2 — факториальная и случайная вариации; η_{1x}^2 — показатель силы влияния фактора (широты) на параметры флор; η_{1z}^2 — показатель силы влияния остальных (не учтенных в анализе) факторов; F — критерий Фишера.

дает, однако изменения этих показателей наиболее четко выражены лишь в пределах Арктики. Отношение числа видов Asteraceae/Roaceae медленно повышается с севера на юг.

Добиться чего-либо сверх сказанного описательными методами, по-видимому, нельзя. Попробуем уточнить и углубить наши представления о намеченных закономерностях географической изменчивости флористических показателей посредством

использования приемов дисперсионного (глава 3) и регрессионного (глава 5) анализов.

Посредством дисперсионного анализа можно выделить ту долю изменчивости параметров флор, которая обусловлена различием в географической широте их местоположения. В табл. 87 представлены результаты анализа 10 (по числу учитываемых флористических показателей) однофакторных дисперсионных комплексов. Доля влияния фактора географической широты на флористические показатели определяется (в процентах) величиной η_x^2 , доля влияния остальных (не учтенных) факторов — величиной η_z^2 , а достоверность выводов оценивается критерием Фишера F (Приложение IV).

Таблица 88. Средние значения количественных показателей конкретных и близких к ним флор европейской части СССР в зависимости от географической широты (точки эмпирических линий регрессий)

Географическая широта, °с. ш.	Число видов	Число родов	Число семейств	Среднее число видов в семействе	Среднее число родов в семействе	Среднее число видов в роде	Процент однодольных от числа цветковых	Процент спайнолепестных от числа двудольных	Процент видов в 10 ведущих семействах	Отношение числа видов семейства Asteraceae/Poaceae
50—55	816	376	90	9,2	4,3	2,1	22,6	42,4	58,1	1,4
55—60	687	333	90	7,5	3,7	2,0	25,0	40,1	53,5	1,2
60—65	522	258	65	6,7	3,4	2,0	26,8	40,2	55,9	1,2
65—70	260	136	48	5,3	3,4	1,9	28,9	37,3	63,2	0,8
70—75	108	58	23	4,8	2,6	1,9	27,0	20,7	80,7	0,4
75—80	60	36	15	4,0	2,4	1,7	35,1	8,1	91,2	0,1
80—85	25	17	9	3,3	1,9	1,8	33,0	0	100,0	0

Как видно из табл. 87, сильнее всего от широты зависят показатели флористического богатства: число видов — на 87 %, число родов — на 88, число семейств — на 82 %. Количество спайнолепестных, процент видов в 10 ведущих семействах и отношение числа видов Asteraceae/Poaceae зависят от широты на 70—77 %. Из «пропорций флоры» сильно зависят от географической широты показатели в/с (на 72 %) и р/с (на 79 %), тогда как показатель в/р обнаруживает хотя и достоверную, но наиболее слабую (всего на 20 %) зависимость от этого фактора, что еще раз подчеркивает своеобразие этого показателя. Довольно слабо (на 36 %), но вполне достоверно зависит от широты и такой показатель, как процент однодольных в конкрет-

ных флорах. Таким образом, дисперсионный анализ не только подтверждает, но и уточняет высказанные ранее глазомерные представления о географической изменчивости флористических показателей в пределах рассматриваемой территории.

Переходя к регрессионному анализу, поставим своей задачей получение уравнений, моделирующих изменение флористических параметров по мере продвижения с севера на юг и позволяющих с известной степенью риска осуществлять приближенное прогнозирование величины этих параметров на той или иной географической широте.

В табл. 88 приведены средние значения флористических показателей в зависимости от географической широты (точки эмпирических линий регрессии). Для каждого показателя подбиралась наилучшая интерполяционная формула (см. раздел 5.4.2) и рассчитывалась теоретическая функция, моделирующая наблюдаемую закономерность.

Оказалось, что зависимость показателей флористического богатства (числа видов, родов и семейств) от географической широты хорошо интерполируется (рис. 41) обратной логистической функцией

$$y = \frac{A}{1 + 10^{a-bx}},$$

где y — число таксонов во флоре; x — географическая широта; A — верхняя асимптота, за которую принято наибольшее число таксонов, свойственное одной из самых южных флор (Беловежская Пуща); a и b — константы, расчет которых показан в разделе 6.2.1.

Логистическая кривая применяется главным образом для описания роста организмов и динамики численности популяций. Поэтому необходимо объяснить биологический смысл s -образной кривой в рассматриваемом случае.

Как известно, логистическая кривая имеет три точки перегиба («критические точки»): T_1 , T_2 и T_3 , характеризующие переломные моменты описываемого этой кривой процесса. Прос-

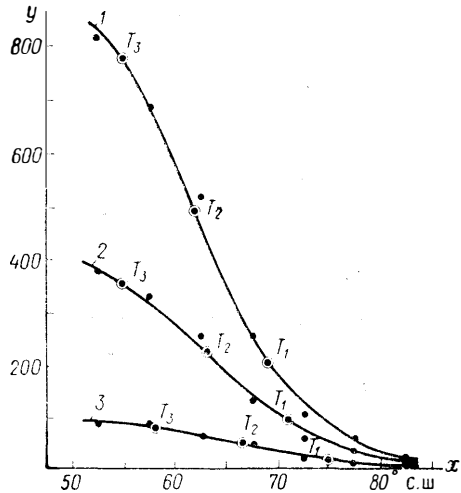


Рис. 41. Регрессия показателей флористического богатства (y) на географическую широту (x).

1 — число видов, 2 — число родов, 3 — число семейств; T_1 , T_2 и T_3 — критические точки логистических кривых.

тые формулы для определения координат указанных точек были даны в разделе 6.2.1. Применительно к нашему случаю эти формулы означают, что на той широте, под которой лежит точка T_2 , конкретные флоры содержат в среднем 50 % таксонов от их числа (A) в наиболее богатой флоре, а на широте точек T_1 и T_3 — около 21 и 79 % соответственно. Скорость и ускорение процесса, подчиняющегося логистической кривой, могут быть определены в любой ее точке посредством вычисления первой и второй производных, формулы которых также даны в разделе 6.2.1.

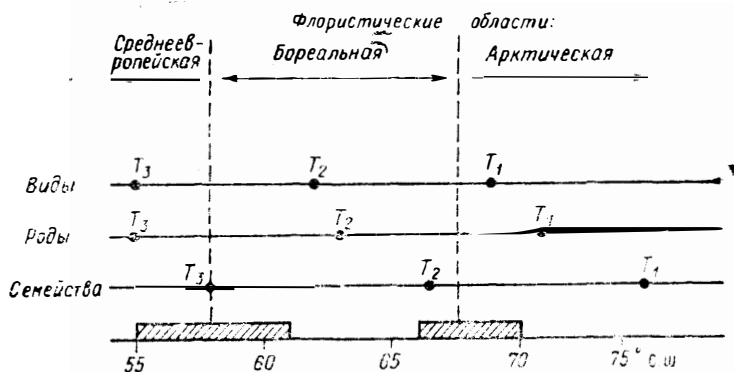


Рис. 42. Относительное положение критических точек T_1 , T_2 и T_3 логистических кривых.

Штриховкой показаны пределы, в которых на территории европейской части СССР проходят границы флористических областей по А. И. Толмачеву [1974].

S-образная форма кривых показателей флористического богатства в рассматриваемом примере определенно связана с тем, что, продвигаясь с севера на юг, мы последовательно пересекаем границы между Арктической и Бореальной, Бореальной и Среднеевропейской флористическими областями, т. е. переступаем важные флористические рубежи. Пояснить это можно простой схемой (рис. 42), на которой показано положение точек перегиба логистических кривых числа видов, родов и семейств в конкретных флорах. На схеме видно, что точки T_1 , в которых достигается максимальное ускорение процесса увеличения числа таксонов, лежат в пределах Арктики. Точки перегиба T_2 , характеризующие момент достижения максимальной скорости того же процесса, находятся в пределах Бореальной флористической области. Наконец, точки T_3 , соответствующие моменту максимального «торможения» прироста числа таксонов, лежат на территории Среднеевропейской флористической области.

Из рис. 42, а также из табл. 89, в которой приведены вычисленные методом наименьших квадратов уравнения логистических кривых и координаты точек перегиба, видно, что критические точки T_1 , T_2 и T_3 для числа семейств лежат севернее, чем соответствующие точки для числа родов, а последние, в свою очередь, севернее, чем для числа видов. Иными словами, по мере продвижения с севера на юг переломные моменты процесса увеличения числа семейств в конкретных флорах предвещают наступление аналогичных моментов для числа родов и видов.

Таблица 89. Ожидаемое число таксонов, максимальные скорость и ускорение увеличения их числа в критических точках логистических кривых

Флористические показатели	Интерполяционная формула	Координаты критических точек						Максимальная скорость увеличения числа таксонов на 1 градус широты в точке T_3	Максимальное ускорение увеличения числа таксонов в точке T_1
		T_1		T_2		T_3			
		x	y	x	y	x	y		
Число видов	$y = \frac{984}{1 + 10 - 4,7886 + 0,0773x}$	69	208	62	492	55	776	44	3,00
Число родов	$y = \frac{451}{1 + 10 - 4,6297 + 0,736x}$	71	95	63	226	55	356	19	1,24
Число семейств	$y = \frac{103}{1 + 10 - 4,5494 + 0,0684x}$	75	22	66,5	52	58	81	4	0,25

Примечание. Для удобства приводимые в таблице цифры даны с округлением; x — градусы северной широты, y — число таксонов.

Из данных табл. 89 следует также, что максимальная скорость увеличения числа таксонов на 1 градус широты составляет 44 вида (на широте 62°), 19 родов (на широте 63°) и 4 семейства (на широте $66,5^\circ$). Таким образом, максимум пополнения флор новыми семействами лежит на $3,5$ — $4,5$ градуса севернее максимума пополнения флор новыми родами и видами.

Следующий показатель — процент видов однодольных от числа цветковых — обнаруживает практически линейную зависимость от географической широты, и его изменение описывается уравнением прямой линии $y = 3,4409 + 0,3689x$ (рис. 43, 1). Величина коэффициента регрессии ($R_{y/x} = 0,3689$) из этого уравнения свидетельствует о том, что при продвижении с севера на юг процент однодольных в конкретных флорах падает с каждым градусом широты примерно на 0,37.

Регрессия трех показателей — процента спайлопестных от числа двудольных, процента видов в 10 ведущих семействах и отношения числа видов Asteraceae/Poaceae — является нели-

нейной и удовлетворительно интерполируется параболой второй степени. Процент спайнолепестных (рис. 43, 3) следует уравнению $y = -122,616 + 6,049x - 0,0554x^2$; процент видов в 10 ведущих семействах (рис. 43, 2) — уравнению $y = 328,0445 - 9,4389x + 0,0818x^2$, а отношение числа видов Asteraceae/Рoaceae (рис. 44) — формуле $y = -0,2164 + 0,0891x - 0,0011x^2$.

Из рисунков видно, что процент спайнолепестных и отношение Asteraceae/Рoaceae в направлении с севера на юг возрастают, причем увеличение первого показателя прекращается примерно на широте 55° . Противоположным образом изменяется процент видов в 10 ведущих семействах, падающий в указанном направлении примерно до $57,5^\circ$ с. ш., но затем, южнее, несколько возрастающий.

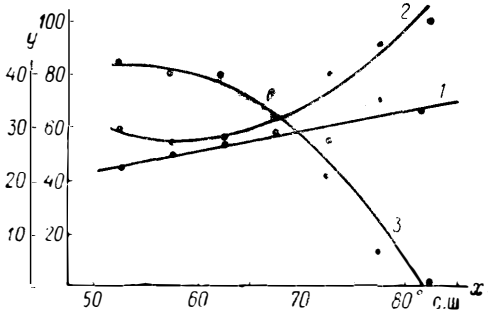


Рис. 43. Регрессия флористических показателей (y) на географическую широту (x). 1 — процент однодольных от числа цветковых; 2 — процент видов, охватываемых десятью ведущими семействами; 3 — процент спайнолепестных от числа двудольных.

Беря первую производную от исходной функции по формуле $y' = b + 2cx$, можно определить скорость изменения флористических показателей в любой точке параболы.

Для точек эмпирических линий регрессии трех упомянутых выше показателей такие данные приведены в табл. 90. Приравняв первую производную к нулю и получив тем самым формулу $x = -(b/2c)$, легко установить максимум или минимум функции. Расчёты показывают, что минимальный процент числа видов в 10 ведущих семействах наблюдается на широте $57,7^\circ$, а максимум процента спайнолепестных достигается под $54,6^\circ$ с. ш.

Изменение «пропорций флоры» в зависимости от географической широты имеет на территории европейской части СССР линейный характер (рис. 45). Среднее число видов в семействе связано с широтой функцией $y = 18,6536 - 0,1900x$; среднее число родов в семействе — уравнением $y = 8,2098 - 0,0757x$, а среднее число видов в роде — зависимостью $y = 2,6838 - 0,0114x$. Практически это означает, что при про-

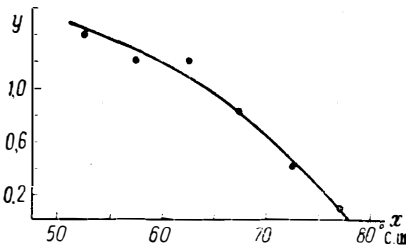


Рис. 44. Регрессия отношения числа видов Asteraceae/Рoaceae (y) на географическую широту (x).

Таблица 90. Скорость изменения количественных показателей (y) конкретных флор на 1 градус широты в точках эмпирических линий регрессии (x)

x	y		
	Процент спайнолепестных от числа двудольных	Процент видов в 10 ведущих семействах	Отношение числа видов Asteraceae/Poaceae
52,5	+0,23	-0,85	-0,03
57,5	-0,32	-0,03	-0,04
62,5	-0,88	+0,79	-0,05
67,5	-1,43	+1,60	-0,06
72,5	-1,98	+2,42	-0,07
77,5	-2,54	+3,24	-0,08
82,5	-3,09	+4,06	-0,09

движении с севера на юг на каждый градус широты показатель v/c возрастает в конкретных флорах примерно на 0,2, показатель p/c — на 0,1, а показатель v/p — всего лишь на 0,01. Тем самым окончательно подтверждается высказанное ранее мнение о слабой географической изменчивости этого показателя.

Выше отмечалось, что почти все важнейшие количественные показатели флор обнаруживают сильную зависимость от географической широты. Дело в том, что географическая широта местности, не являясь сама по себе экологическим фактором, служит тем не менее хорошим суммарным показателем, вбирающим в себя множество реальных условий внешней среды. С ней сильно коррелируют, например, такие важнейшие экологические факторы, как температурный режим и влажоснабжение. Все это позволяет использовать приведенные нами уравнения регрессии в качестве уравнений приближенного прогнозирования соответствующих показателей конкретных

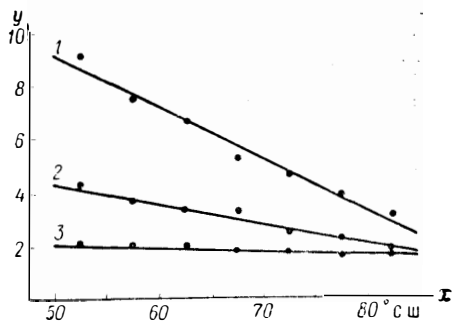


Рис. 45. Регрессия показателей систематического разнообразия (y) на географическую широту (x).

1 — среднее число видов в семействе; 2 — среднее число родов в семействе; 3 — среднее число видов в роде.

Таблица 91. Прогнозируемая величина (y') количественных показателей конкретных флор европейской части СССР (исключая Кавказ) в зависимости от географической широты (x); величина ошибки (m) уравнений регрессии и теоретически допустимая погрешность прогноза ($\pm tm$)

Показатель	Тип функции	Уравнение	m	$\pm tm$
Число видов	Обратная логистическая	$y' = \frac{984}{1+10^{-4,7886+0,0773x}}$	30,6	86
Число родов		$y' = \frac{451}{1+10^{-4,6297+0,0736x}}$	16,7	47
Число семейств		$y' = \frac{103}{1+10^{-4,5494+0,0684x}}$	5,0	14
Среднее число видов в семействе	Прямая линия	$y' = 18,6536 - 0,1900x$	0,4	1,0
Среднее число родов в семействе		$y' = 8,2098 - 0,0757x$	0,2	0,5
Среднее число видов в роде		$y' = 2,6838 - 0,0114x$	0,1	0,3
Процент однодольных	Парабола второго порядка	$y' = 3,4409 + 0,3689x$	2,1	5,4
Процент спайнолепестных		$y' = -122,6160 + 6,0490x - 0,0554x^2$	3,	8,6
Процент видов в 10 ведущих семействах		$y' = 328,0445 - 9,4389x + 0,0818x^2$	5,1	14,3
Отношение числа видов Asteraceae/Rosaceae		$y' = -0,2164 + 0,0891x - 0,0011x^2$	0,1	0,3

флор на территории европейской части СССР [Шмидт, 1981].¹

В табл. 91 помимо самих уравнений приведены их ошибки (m), вычисленные по формуле [Зайцев, 1973, с. 135—136]

$$m = \sqrt{\frac{(y - y')^2}{N - n}},$$

где y и y' — фактические и теоретические значения флористических показателей; N — число точек эмпирической линии регрессии, по которым вычислялся уравнение; n — число коэффициентов уравнения, включая свободный член. Там же показана теоретически допустимая погрешность прогноза ($\pm tm$) при 5%-ном уровне существенности, полученная путем умножения ошибки уравнения на табличное значение t -критерия (Приложение III) при $N - n$ степенях свободы.

Прогноз величины того или иного флористического показателя осуществляется посредством подстановки в соответствующее уравнение вместо неизвестного x координаты той широты, под которой расположен центр подлежащей изучению конкретной флоры (при этом минуты широты следует переводить в десятичную дробь). Пользование приведенными уравнениями, за исключением первых трех, никаких трудностей не вызывает. Что же касается логистических уравнений числа видов, родов и семейств, то для вычисления прогнозируемой величины следует использовать алгоритм, приведенный в табл. 92 и снабженный фактическим примером.

Для определения степени соответствия прогноза (y') полученным на практике данным (y) необходимо

¹ В перспективе имеется возможность выведения более точных уравнений множественной регрессии, основанных на учете ряда важнейших факторов среды.

Таблица 92. Алгоритм вычисления прогнозируемых значений показателей флористического богатства (y') на примере конкретной флоры окрестностей г. Себеж Псковской обл. (обратная логистическая функция, $x = 56,27^\circ$ с. ш.)

Флористические показатели	Параметры уравнений из табл. 91			bx	$a + bx = \lg\left(\frac{A}{y' - 1}\right)$	$\frac{A}{y' - 1}$	$\frac{A}{y'}$	Прогноз, y'	Фактически найдено, y
	А		b						
	a	b							
Число видов	984	-4,7886	0,0773	4,3497	-0,4389	1,5611	0,3640	721	723
Число родов	451	-4,6297	0,0736	4,1415	-0,4882	1,5118	0,3251	340	335
Число семейств	103	-4,5494	0,0684	3,8489	-0,7005	1,2995	0,1993	86	91

прежде всего посмотреть, укладывается ли разность между этими величинами в интервал $\pm tm$, указанный в табл. 91. Если эта разность превышает величину tm , следует считать, что прогноз не оправдался, и искать причину данного явления в локальных особенностях факторов среды рассматриваемой конкретной флоры. Для всех трех показателей, приведенных в табл. 92, прогноз следует считать оправдавшимся.

Однако и в пределах допустимой ошибки прогноза ($\pm tm$) степень его осуществления может быть различной. Следовательно, надо найти способ оценки точности прогноза и тогда, когда теоретически его можно считать оправдавшимся.

Абсолютную разность между фактическим (y) и прогнозированным значениями флористического показателя запишем в виде выражения $d = y - y'$. Отнеся эту разность к значению прогноза и выразив ее в процентах, получим относительную погрешность прогноза

$$m_{\text{отн}} = \frac{100d}{y'} \%$$

Относительной погрешностью прогноза, вообще говоря, уже можно пользоваться для оценки степени расхождения между фактическими и прогнозированными значениями флористических показателей. Гораздо удобнее, однако, использовать дополняющую ее до 100 % величину, которая и представляет собой процент осуществления (или точность) прогноза:

$$\omega_p = 100 - \frac{100d}{y'} = \frac{100(y' - d)}{y'} \%$$

Учитывая, что в случае $y' > y$ $d = y' - y$, а при $y' < y$ $d = y - y'$, приведенную выше формулу после некоторых преобразований можно представить в виде двух более удобных для практического использования вариантов, а именно:

при $y' > y$

$$\omega_p = \frac{100y}{y'} \%$$

при $y' < y$

$$\omega_p = \frac{100(2y' - y)}{y'} \%$$

Применительно к данным табл. 92 имеем:
для числа видов

$$\omega_p = \frac{100(2 \cdot 721 - 723)}{721} = 99,7 \%$$

для числа родов

$$\omega_p = \frac{100 \cdot 335}{340} = 98,5 \%$$

для числа семейств

$$\omega_p = \frac{100(2 \cdot 86 - 91)}{86} = 94,2\%.$$

Как видим, точность прогноза в этих случаях оказалась весьма высокой.

В табл. 93 показаны результаты прогнозирования 10 флористических показателей для 11 конкретных флор, расположенных в довольно широком географическом интервале (в пределах Арктической, Бореальной и Среднеевропейской флористических областей), изученных уже после того, как рекомендуемые уравнения регрессии были выведены.

Из 110 сделанных прогнозов не оправдалось ($d > tm$) 19, что составляет 17,3 %. При этом 14 случаев неоправдания прогноза приходится на конкретные флоры п-ова Канин и, за исключением одного случая, относятся лишь к трем показателям: среднему числу видов в семействе, среднему числу родов в семействе и проценту спайнолепестных.

Худшие результаты прогнозирования показателей в/с и р/с канинских флор формально связаны с тем, что флоры Канина обнаруживают «недобор» числа видов и родов, но «перебор» числа семейств: при расчете отношений в/с и р/с это обстоятельство усиливает отклонение от прогноза. Но за этими «внешними» причинами кроются глубокие ботанико-географические особенности флоры п-ова Канин, заключающиеся в проникновении здесь на север отдельных видов из таких семейств, которые в других районах обычно не имеют представительства в высоких широтах. Граница между Арктической и Бореальной флористическими областями на Канине имеет «размытый» характер, представляя собой широкую переходную зону, в которую довольно легко проникают виды, имеющие в целом более южное распространение. Этой же причиной объясняется и повышенный процент спайнолепестных в конкретных флорах Канина.

Из остальных 5 случаев невыполнения прогноза наиболее бросающимся в глаза является «недобор» числа видов в конкретной флоре Водосье Новгородской области: здесь вместо прогнозируемых 607 видов обнаружено только 507. Объяснить это можно тем, что значительная часть территории этой флоры занята регулярно затопляемыми, угнетенными и бедными по своему флористическому составу дубняками и мелколиственными лесами в пойме р. Волхов.

В 91 случае из 110 (что составляет 83,7 %) прогноз следует считать оправдавшимся ($d \leq tm$) причем в 55 случаях точность прогноза $\omega_p > 90\%$, в 38 случаях $\omega_p = 80-90\%$ и лишь в 17 случаях $\omega_p < 80\%$.

В среднем точность прогноза достаточно высока — 88,3 %. В табл. 93 подсчитан также средний процент осуществления

Таблица 93. Прогнозируемые (y'), фактически полученные (y) значения количественных показателей конкретных флор и процент осуществления прогноза (ω_p)

Район конкретной флоры	Градусы северной широты	Число видов			Число родов			Число семейств			Среднее число видов в семействе		
		y'	y	ω_p	y'	y	ω_p	y'	y	ω_p	y'	y	ω_p
П-ов Канин:													
Тарханово	68,50	234	197	84,2	126	118	93,7	43	49	86,0	5,6	4,0	71,4
Камбальница	68,33	239	185	77,4	129	114	88,4	44	48	90,9	5,7	3,9	68,4
Микулкин Нос	67,82	256	213	83,2	137	128	93,4	46	51	89,1	5,8	4,2	72,4
Голубница	67,07	282	209	74,1	149	126	84,6	49	51	95,9	5,9	4,1	69,5
Вижас	66,63	298	266	89,2	157	158	99,4	51	56	90,2	6,0	4,8	80,0
Ленингр. обл.:													
Пашозеро	69,14	571	526	92,1	277	245	88,4	75	75	100,0	7,2	7,0	97,2
Новгор. обл.:													
Водосье	59,28	607	507	83,5	293	285	97,3	78	86	89,7	7,4	5,9	79,7
Теребутенец	59,11	613	569	92,8	296	272	91,9	79	89	87,3	7,4	6,4	86,5
Сольцы	58,10	654	570	87,2	313	288	92,0	81	88	91,4	7,6	6,5	85,5
Латв. ССР:													
Салаца	57,89	662	678	97,6	316	339	92,7	82	93	83,6	7,6	7,3	96,1
Валка	57,68	670	673	99,6	319	342	92,8	82	94	85,4	7,7	7,2	93,5
$\bar{\omega}_p$	—	—	—	87,4	—	—	92,2	—	—	90,2	—	—	81,8

Район конкретной флоры	Граду-сы се-верной широты	Среднее число родов в семей-стве			Среднее число видов в роде			Процент однодольных			Процент спайнолепест-ных			Процент видов в 10 ведущих семействах			Отношение чис-ла видов Astera-сcae/Poaceae			Средний % осуществле-ния прогноза	
		у'	у	ω _p	у'	у	ω _p	у'	у	ω _p	у'	у	ω _p	у'	у	ω _p	у'	у	ω _p		
И-ов Канин:																					
Тарханово . . .	68,50	3,0	2,4	80,0	1,9	1,7	89,5	28,7	24,3	84,7	31,8	42,0	67,9	65,3	61,9	94,8	0,7	0,9	71,4	82,4	
Камбальница	68,33	3,0	2,4	80,0	1,9	1,6	84,2	28,6	27,2	95,1	32,0	39,0	78,1	65,0	61,5	94,6	0,7	0,9	71,4	82,9	
Микулкин Нос	67,82	3,1	2,5	80,6	1,9	1,7	89,5	28,5	28,0	98,2	32,8	42,5	70,4	64,1	60,1	93,8	0,8	0,9	87,5	85,8	
Голубница . . .	67,07	3,1	2,5	80,6	1,9	1,7	89,5	28,2	26,4	93,6	33,9	47,6	59,6	62,9	58,4	92,8	0,8	1,0	75,0	81,5	
Вижас	66,63	3,2	2,8	87,5	1,9	1,7	89,5	28,0	22,4	80,0	34,9	43,8	74,5	62,3	57,5	92,3	0,8	1,0	75,0	85,8	
Ленингр. обл.:																					
Пашозеро . . .	60,14	3,7	3,3	89,2	2,0	2,1	95,0	25,6	24,6	96,1	40,8	42,6	95,6	56,2	56,3	99,8	1,2	1,2	100,0	95,3	
Новгор. обл.:																					
Водосье	59,28	3,7	3,3	89,2	2,0	1,8	90,0	25,3	22,6	89,3	41,3	43,6	94,4	56,0	53,5	95,5	1,2	1,6	66,7	87,5	
Теребутенец . . .	59,11	3,7	3,1	83,8	2,0	2,1	95,0	25,2	26,6	94,4	41,4	44,8	91,8	55,9	54,1	96,8	1,2	1,3	91,7	91,2	
Сольцы	58,10	3,8	3,3	86,8	2,0	2,0	100,0	24,9	24,8	99,6	41,8	40,8	97,6	55,8	53,3	95,5	1,2	1,0	83,3	91,9	
Латв. ССР:																					
Салаца	57,89	3,8	3,6	94,7	2,0	2,0	100,0	24,8	26,3	94,0	41,9	37,0	88,3	55,8	54,3	97,3	1,3	1,2	92,3	94,0	
Валка	57,68	3,8	3,6	94,7	2,0	2,0	100,0	24,7	26,0	94,7	42,0	35,8	85,2	55,8	54,5	97,7	1,3	1,2	92,3	93,6	
ω _p		—	—	—	86,1	—	—	92,9	—	—	92,7	—	—	82,1	—	—	95,5	—	—	82,4	88,3

Примечание. Фактические значения количественных показателей флор приведены (или вычислены) на основании опубликованных работ В. Г. Сергиенко, Г. Б. Клявина и др., а также дипломных работ студентов кафедры ботаники ЛГУ С. А. Рукавишниковой, И. Я. Нейуминой, Л. Л. Малышева и А. Г. Андреева, выполненных в 1977–1980 гг. Подчеркнутые в таблице цифры – случаи неосуществления прогноза. Пояснения в тексте.

прогноза ($\bar{\omega}_p$) для каждой флоры по всем 10 показателям и для каждого показателя по всем изученным флорам. Как видно, для 5 флор $\bar{\omega}_p > 90\%$, а для остальных — более 80%. Что касается флористических показателей, то для половины из них точность прогноза превышает 90%. Наилучший результат прогнозирования ($\bar{\omega}_p = 95,5\%$) получен в среднем для такого показателя, как процент видов в 10 ведущих семействах, а наихудший, но также достаточно высокий ($\bar{\omega}_p = 81,8\%$) — по среднему числу видов в семействе.

Таким образом, использованные уравнения регрессии являются вполне приемлемыми для приближенного прогнозирования параметров конкретных флор в пределах европейской части СССР. Разумеется, по мере изучения новых конкретных флор эти уравнения (вернее, их коэффициенты) могут подвергаться уточнению.

8.5. СРАВНЕНИЕ ФЛОР ПО ИХ ИСТОРИКО-ГЕОГРАФИЧЕСКИМ СВЯЗЯМ С ПОМОЩЬЮ МЕТОДА ТАКСОНОМИЧЕСКОГО АНАЛИЗА Е. С. СМИРНОВА

Как известно, ботанико-географический анализ флоры может проводиться в двух основных аспектах: *географическом* и *генетическом* [Толмачев, 1974]. Собственно географический анализ флоры строится на изучении современного распространения видов растений: виды, обладающие сходным типом ареала, объединяются в тот или иной географический элемент флоры. В отличие от этого предметом генетического анализа флоры является выяснение того, откуда происходит данный вид, когда и как он попал в состав изучаемой флоры. Виды, обнаруживающие общие черты в отношении своего географического происхождения, объединяются в тот или иной генетический элемент флоры.

Из сказанного ясно, что географический и генетический элемент флоры — понятия существенно различные. Использование термина «элементы флоры» без уточняющего эпитета может привести к недоразумениям. Поэтому необходимо сразу же оговорить, что, используя для краткости этот термин, мы в дальнейшем будем иметь в виду *миграционно-генетический* (или, как его можно еще называть, историко-географический) элемент флоры, представление о котором введено и обосновано Н. А. Миняевым в ряде его работ (например, [Миняев, 1966]).

Миграционно-генетический элемент — это группа видов, имеющих общие пути и время проникновения на территорию данной флоры. При отнесении того или иного вида к определенному миграционно-генетическому элементу флоры учитывается, конечно, его современный ареал, но главное заключа-

ется в том, что на фоне данных палеоботаники, стратиграфии и геоморфологии прослеживается история заселения видом данной территории. Важность такого подхода к анализу флоры неоспорима уже потому, что виды, обладающие сходным типом современного ареала, по своему географическому происхождению, путям миграции и времени вхождения в состав флоры могут быть существенно различны. Может иметь место и обратное соотношение: расселение близких по своему географическому происхождению видов порой происходит с разной скоростью и разными путями. На основе понятия о миграционно-генетическом элементе оказывается возможным как сравнение флор по их историко-географическим связям, так и статистический подход к флористическому районированию.

На первый взгляд может показаться, что коль скоро установлен перечень элементов флоры и набор входящих в каждый из них видов растений, то для сравнения флор по этому признаку достаточно использовать коэффициенты сходства, рассмотренные в разделе 8.3.4. Но это не так. Определение степени сходства флор по их историко-географическим связям представляет собой задачу, значительно более тонкую и сложную, нежели сопоставление флористических списков.

Во-первых, здесь важно учесть не просто сходство флор по совпадению видов, а их сходство по числу видов каждого элемента флоры, независимо от того, одинаковые это виды или разные. Диктуется это тем, что одна и та же историко-географическая связь может проявляться на данной территории в форме присутствия далеких друг от друга в систематическом отношении видов, принадлежащих тем не менее к одному и тому же элементу флоры.

Во-вторых, кроме факта присутствия того или иного элемента, необходимо учесть его «вес» в составе флоры, т. е. степень выраженности той историко-географической связи, которую этот элемент представляет. С помощью обычных коэффициентов сходства эту задачу решить нельзя.

Наиболее эффективным средством достижения требуемого в данном случае результата является метод таксономического анализа Е. С. Смирнова, изложенный в разделе 7.2.2.

Напомним, что в соответствии с этим методом наибольший «вес» придается редким признакам. Применительно к нашей задаче это означает, что совпадение флор по наличию в их составе редких элементов мы должны оценивать выше, нежели совпадение по элементам, широко представленным в каждой флоре. Редкие элементы флоры чаще всего свидетельствуют о древних историко-географических связях, представляющих наибольший интерес, но иногда редкие элементы могут быть и «молодыми», указывая на новейшие тенденции изменения состава флор. Но в любом случае редкие элементы флоры заслуживают внимания исследователя больше, нежели элементы,

вполне обычные для изучаемой флоры. Из этого следует, что сущность метода таксономического анализа Е. С. Смирнова полностью соответствует содержанию интересующей нас задачи.

В качестве примера использования указанного метода рассмотрим сравнение 10 конкретных флор Псковской области по их историко-географическим связям [Шмидт, Баранова, 1975]. Виды, образующие в пределах Псковской области достаточно четкие границы своих ареалов, принадлежат к 12 элементам флоры. Распределение числа видов каждого элемента флоры показано в табл. 94.

Таблица 94. Количество видов различных элементов флоры в 10 конкретных флорах Псковской обл.

Районы и номера конкретных флор	Шифр элементов флоры												Всего
	A	B	C	D	E	F	G	H	I	K	L	M	
1. Плюсса	0	4	11	6		0	8	1	1	4	2	10	49
2. Середка	2	1	4	10		6	14	2	1	1	3	4	51
3. Псков	5	6	11	20		13	26	8	6	0	3	13	121
4. Качаново	1	4	12	16	3	7	20	5	3		0	9	83
5. Пушкинские Горы	0	2	2	9	2	6	6	1	0	3	0	3	34
6. Алоль	0	2	5	13		12	29	4	0	5	2	2	77
7. Себеж	0	1	1	0	4	11	12	0	0	0	2	1	32
8. Полибно	0	1	4	6	1			0	0	1	2	0	43
9. Жижица	1	2	2	8	1			0	0	1	0	2	43
10. Усвяты	0	0	0	0	0	6	0	0	0	0	0	0	6

Примечание. Элементы флоры: А—арктический и гипоарктический океанический, В—гипоарктический континентальный, С—евросибирский таежный, D—евросибирский подтаежный, E—евросибирский пустынно-степной, F—восточноевропейский широколиственный (сарматский), G—паневропейский широколиственный (неморальный), H—среднеевропейский широколиственный (неморальный), I—среднеевропейский горный (альпийский), K—среднеевропейский горный (карпатский), L—балтийский, M—расширенно-океанический (атлантический).

В таксономическом анализе Е. С. Смирнова решающее значение имеет кодирование признаков. Прямое кодирование в рассматриваемом примере ведет к ошибочному результату.

Материал, с которым мы имеем дело, больше всего соответствует ситуации с наличием градиента признака (с. 166). Хотя градиент в полном смысле этого слова и не наблюдается (количество видов данного элемента в каждой из конкретных флор не образует ряда с закономерно изменяющимся интервалом), решающим оказывается то обстоятельство, что сравниваемые флоры можно расположить в порядке нарастающего (или убывающего) ряда модальностей каждого элемента, на-

Таблица 95. Характеристика конкретных флор Псковской обл. по наличию в их составе видов различных элементов флоры

Номер флоры	A	B	C	D	
1	$a_1 a_2 a_3$	$B_1 B_2 B_4 b_6$	$C_1 C_2 C_4 C_5 C_{11} c_{12}$	$D_5 d_3 d_9 d_{10} d_{13} d_{16} d_{20}$	
2	$A_1 A_2 a_3$	$B_1 b_2 b_1 b_6$	$C_1 C_2 C_4 c_3 c_{11} c_{12}$	$D_5 D_8 D_9 D_{10} d_{13} d_{16} d_{20}$	
3	$A_1 A_2 A_3$	$B_1 B_2 B_4 B_6$	$C_1 C_2 C_4 C_5 C_{11} c_{12}$	$D_5 D_8 D_9 D_{10} D_{13} D_{16} d_{20}$	
4	$A_1 a_2 a_3$	$B_1 B_2 B_4 b_6$	$C_1 C_2 C_4 C_5 C_{11} c_{12}$	$D_5 D_8 D_9 D_{10} D_{13} D_{16} d_{20}$	
5	$a_1 a_2 a_3$	$B_1 B_2 b_4 b_6$	$C_1 C_2 c_4 c_3 c_{11} c_{12}$	$D_5 D_8 D_9 d_{10} d_{13} d_{16} d_{20}$	
6	$a_1 a_2 a_3$	$B_1 B_2 b_4 b_6$	$C_1 C_2 C_4 C_5 c_{11} c_{12}$	$D_3 D_8 D_9 D_{10} D_{13} d_{16} d_{20}$	
7	$a_1 a_2 a_3$	$B_1 b_2 b_4 b_6$	$C_1 c_2 c_4 c_3 c_{11} c_{12}$	$d_3 d_8 d_9 d_{10} d_{13} d_{16} d_{20}$	
8	$a_1 a_2 a_3$	$B_1 b_2 b_1 b_6$	$C_1 C_2 C_4 c_3 c_{11} c_{12}$	$D_5 d_8 d_9 d_{10} d_{13} d_{16} d_{20}$	
9	$A_1 a_2 a_3$	$B_1 B_2 b_4 b_6$	$C_1 C_2 c_4 c_3 c_{11} c_{12}$	$D_5 D_8 d_9 d_{10} d_{13} d_{16} d_{20}$	
10	$a_1 a_2 a_3$	$b_1 b_2 b_4 b_6$	$c_1 c_2 c_4 c_3 c_{11} c_{12}$	$d_3 d_8 d_9 d_{10} d_{13} d_{16} d_{20}$	
Номер флоры	E	F	G		
1	$E_1 E_2 e_3 e_4 e_{10}$	$f_6 f_7 f_{10} f_{11} f_{12} f_{13}$	$G_6 G_8 g_{12} g_{11} g_{16} g_{17} g_{20} g_{26} g_{29}$		
2	$E_1 E_2 E_3 e_4 e_{10}$	$F_6 f_7 f_{10} f_{11} f_{12} f_{13}$	$G_6 G_8 G_{12} G_{14} g_{16} g_{17} g_{20} g_{26} g_{29}$		
3	$E_1 E_2 E_3 E_4 E_{10}$	$F_6 F_7 F_{10} F_{11} F_{12} F_{13}$	$G_6 G_8 G_{12} G_{11} G_{16} G_{17} G_{20} G_{26} g_{29}$		
4	$E_1 E_2 E_3 e_4 e_{10}$	$F_6 F_7 f_{10} f_{11} f_{12} f_{13}$	$G_6 G_8 G_{12} G_{11} G_{16} G_{17} G_{20} g_{26} g_{29}$		
5	$E_1 E_2 e_3 e_4 e_{10}$	$F_6 f_7 f_{10} f_{11} f_{12} f_{13}$	$G_6 G_8 g_{12} g_{14} g_{16} g_{17} g_{20} g_{26} g_{29}$		
6	$E_1 E_2 E_3 e_4 e_{10}$	$F_6 F_7 F_{10} F_{11} F_{12} f_{13}$	$G_6 G_8 G_{12} G_{11} G_{16} G_{17} G_{20} G_{26} G_{29}$		
7	$E_1 E_2 E_3 E_4 e_{10}$	$F_6 F_7 F_{10} F_{11} f_{12} f_{13}$	$G_6 G_8 G_{12} g_{11} g_{16} g_{17} g_{20} g_{26} g_{29}$		
8	$E_1 e_2 e_3 e_4 e_{10}$	$F_6 F_7 F_{10} F_{11} f_{12} f_{13}$	$G_6 G_8 G_{12} G_{11} G_{16} G_{17} g_{20} g_{26} g_{29}$		
9	$E_1 e_2 e_3 e_4 e_{10}$	$F_6 F_7 F_{10} f_{11} f_{12} f_{13}$	$G_6 G_8 G_{12} G_{11} G_{16} g_{17} g_{20} g_{26} g_{29}$		
10	$e_1 e_2 e_3 e_4 e_{10}$	$F_6 f_7 f_{10} f_{11} f_{12} f_{13}$	$g_6 g_8 g_{12} g_{14} g_{16} g_{17} g_{20} g_{26} g_{29}$		
Номер флоры	H	I	K	L	M
1	$H_1 h_2 h_4 h_5 h_8$	$I_1 i_3 i_6$	$K_1 K_3 K_4 k_5$	$L_2 l_3$	$M_1 M_2 M_3 M_4 M_9 M_{10} m_{13}$
2	$H_1 H_2 h_4 h_5 h_8$	$I_1 i_3 i_6$	$K_1 k_3 k_4 k_5$	$L_2 L_3$	$M_1 M_2 M_3 M_4 m_9 m_{10} m_{13}$
3	$H_1 H_2 H_4 H_5 H_8$	$I_1 I_3 I_6$	$k_1 k_3 k_4 k_5$	$L_2 L_3$	$M_1 M_2 M_3 M_4 M_9 M_{10} M_{13}$
4	$H_1 H_2 H_4 H_5 h_8$	$I_1 I_3 i_6$	$K_1 K_3 k_4 k_5$	$l_2 l_3$	$M_1 M_2 M_3 M_4 M_9 m_{10} m_{13}$
5	$H_1 h_2 h_4 h_5 h_8$	$i_1 i_3 i_6$	$K_1 K_3 k_4 k_5$	$l_2 l_3$	$M_1 M_2 M_3 m_4 m_9 m_{10} m_{13}$
6	$H_1 H_2 H_4 h_5 h_8$	$i_1 i_3 i_6$	$K_1 K_3 K_4 K_5$	$L_2 l_3$	$M_1 M_2 m_3 m_4 m_9 m_{10} m_{13}$
7	$h_1 h_2 h_4 h_5 h_8$	$i_1 i_3 i_6$	$k_1 k_3 k_4 k_5$	$L_2 l_3$	$M_1 m_2 m_3 m_4 m_9 m_{10} m_{13}$
8	$h_1 h_2 h_4 h_5 h_8$	$i_1 i_3 i_6$	$K_1 k_3 k_4 k_5$	$L_2 l_3$	$m_1 m_2 m_3 m_4 m_9 m_{10} m_{13}$
9	$h_1 h_2 h_4 h_5 h_8$	$i_1 i_3 i_6$	$K_1 k_3 k_4 k_5$	$l_2 l_3$	$M_1 M_3 m_3 m_4 m_9 m_{10} m_{13}$
10	$h_1 h_2 h_4 h_5 h_8$	$i_1 i_3 i_6$	$k_1 k_3 k_4 k_5$	$l_2 l_3$	$m_1 m_2 m_3 m_4 m_9 m_{10} m_{13}$

Примечание. Нумерация флор и расшифровка элементов флоры даны в табл. 94.

пример, 0, 1, 2, 5 по элементу *A* или 0, 6, 8, 9, 10, 13, 16, 20 по элементу *D* (см. табл. 94).

При таком подходе мы должны принять во внимание, что обладание модальностью более высокого порядка означает одновременно обладание модальностями всех низших порядков. Например, если флоре 2 (табл. 94) присуще значение положительного признака C_4 (четыре вида евросибирского таежного элемента), то она тем более обладает положительными признаками C_2 , C_1 и C_0 (т. е. двумя, одним и нулем видов того же элемента флоры), но не обладает модальностями C_5 , C_{11}

Таблица 96. Таксономические отношения конкретных флор Псковской обл., показывающие степень их сходства по историко-географическим связям

Номер флоры	1	2	3	4	5	6	7	8	9	10
1	+0,903	-0,023	-0,233	+0,012	+0,139	-0,187	-0,165	-0,165	-0,161	-0,120
2	-0,023	+0,632	-0,050	+0,024	+0,036	-0,113	-0,121	-0,118	-0,047	-0,220
3	-0,233	-0,050	+2,460	+0,457	-0,578	+0,117	-0,421	-0,465	-0,462	-0,825
4	+0,012	+0,024	+0,457	+1,054	-0,162	+0,125	-0,456	-0,319	-0,183	-0,555
5	+0,139	+0,036	-0,578	-0,162	+0,575	-0,162	-0,033	-0,102	+0,106	+0,181
6	-0,187	-0,113	+0,117	+0,125	-0,162	+1,055	-0,188	-0,051	-0,117	-0,489
7	-0,165	-0,121	-0,421	-0,456	-0,033	-0,188	+0,742	+0,223	+0,080	+0,339
8	-0,165	-0,118	-0,465	-0,319	-0,102	-0,051	+0,223	+0,606	+0,223	+0,168
9	-0,161	-0,047	-0,462	-0,180	+0,106	-0,117	+0,080	+0,223	+0,499	+0,059
10	-0,120	-0,220	-0,825	-0,555	+0,181	-0,489	+0,339	+0,168	+0,059	+1,462

Примечания. 1. По диагонали таблицы даны таксономические отношения флор к самим себе, показывающие степень их оригинальности по историко-географическим связям. 2. Нумерация флор дана в соответствии с табл. 94.

и C_{12} , свойственными некоторым другим флорам. По этим модальностям для флоры 2 следует ввести отрицательные признаки, и в итоге ее характеристика по элементу *C* будет выглядеть следующим образом: $C_0 C_1 C_2 C_4 C_5 C_{11} C_{12}$. Для флоры 4, обладающей 12 видами элемента *C*, можно написать формулу $C_0 C_1 C_2 C_4 C_5 C_{11} C_{12}$, а для флоры 10, не обладающей ни одним видом из этой флористической группы, — выражение $C_0 C_1 C_2 C_4 C_5 C_{11} C_{12}$.

Составив подобным образом характеристики наших флор по всем 12 элементам флоры, приходим к табл. 95. Заметим при этом, что модальности нулевого порядка (A_0 , B_0 , C_0 и т. д.), а также модальности, реально не существующие в исходном (табл. 94) распределении (например, D_5 , D_7 или F_5 , F_8 и т. д.), в расчет не принимаются, ибо они совпадают у всех 10 флор и ничего не дают для таксономического анализа.

По данным табл. 95, исходя из вертикальных рядов значений признаков, легко составить их бимодальные распределения и вычислить таксономические веса, необходимые для расчета таксономических отношений (см. раздел 7.2.2).

Расчитанные таксономические отношения сведены в табл. 96, а полученный методом «максимального корреляционного пути» дендрит и выделенные корреляционные плеяды флор нанесены на карту Псковской области (рис. 46). Из рассмотрения этих данных можно сделать следующие основные выводы.

По степени сходства историко-географических связей конкретные флоры Псковской области образуют две плеяды: Западную (Псков, Качаново, Алоль) и Южную (Пушкинские Горы, Себеж, Полибино, Жижица, Усвяты). Две северные

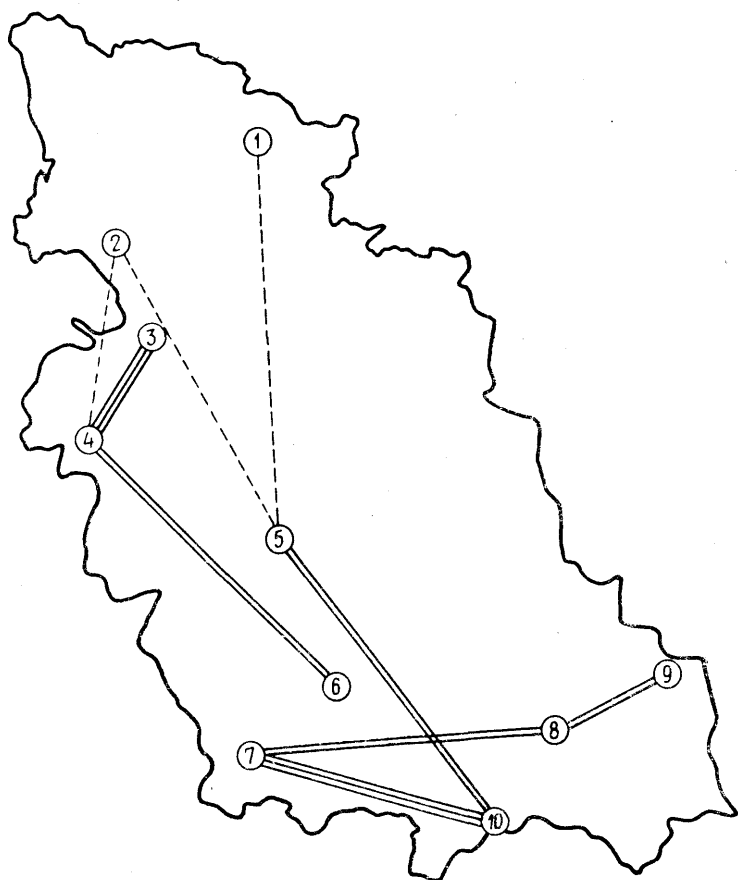


Рис. 46. Дендрит и корреляционные плеяды, отражающие степень сходства историко-географических связей конкретных флор Псковской области. Плеяды выделены двойными, а их ядра — тройными линиями. Нумерация флор дана в табл. 94.

флоры (Плюсса и Середка) выглядят обособленными и, имея слабые положительные связи с флорами Качанова и Пушкинских Гор, не образуют самостоятельной плеяды. Наибольшим сходством по анализируемому признаку обладают пары флор Псков — Качаново и Себеж — Усвяты, составляющие ядра своих плеяд.

Выделенные плеяды характеризуются преобладанием разных историко-географических связей входящих в их состав флор. Сравнивая плеяды с данными табл. 94, легко заметить, что если паневропейский элемент широко представлен во всех флорах, кроме Усвят, то флоры Западной плеяды выделяются обилием видов евросибирского подтаежного элемента, а флоры Южной плеяды — обилием сарматского элемента. В Западной плеяде довольно сильно представлены атлантический, среднеевропейский неморальный, гипоарктический континентальный и, отчасти, карпатский элементы флоры, которые в Южной плеяде представлены слабо, а в некоторых ее флорах отсутствуют вовсе. Отсутствует во флорах Южной плеяды и альпийский элемент.

Все это хорошо согласуется с ранее высказанными [Миняев, 1966; Баранова, Миняев, Шмидт, 1971] соображениями о роли и путях проникновения различных элементов флоры на территорию Псковской области. Таким образом, полученные методом таксономического анализа результаты подкрепляют сложившиеся представления об историко-географических связях флоры данной области.

В разделе 7.2.2 было уделено внимание понятию о таксономическом сходстве объекта с самим собой (t_{xx}) как мере его оригинальности в данной системе объектов. Применительно к нашему примеру с конкретными флорами Псковской области это понятие также имеет четкий биологический смысл. Как видно из табл. 96, наиболее оригинальной является флора окрестностей Пскова ($t_{xx}=2,460$). Действительно, эта флора выделяется среди других как своим богатством, так и наибольшим числом видов из разных флористических групп (табл. 94). Именно здесь находится средоточие многих из учтенных нами видов, проникающих сюда по различным направлениям и часто образующих границы своих ареалов в районе г. Пскова.

На втором месте по степени оригинальности стоит наиболее южная по своему положению флора окрестностей с. Усвяты ($t_{xx}=1,462$), но биологический смысл ее оригинальности противоположен предыдущему: данная флора характеризуется полным отсутствием всех учтенных элементов, кроме сарматского (см. табл. 94). Виды северных элементов флоры не достигают этой местности, а пути проникновения видов, принадлежащих к западным, южным и восточным элементам флоры, лежат в стороне, как бы охватывая ее клещами.

Следующими по степени оригинальности являются флоры окрестностей с. Алоль ($t_{xx}=1,065$) и с. Качаново ($t_{xx}=1,054$), самобытность которых хорошо известна исследователям флоры Псковской области.

Наименее оригинальны по своим историко-географическим связям флоры пос. Жижица ($t_{xx}=0,499$) и пос. Пушкинские Горы ($t_{xx}=0,575$), обладающие средним по числу и весу набором видов разных флористических групп. Разумеется, это не означает, что в составе данных флор нет уникальнейших для флоры Псковской области видов растений (таких, например, как *Galium vernum* близ Пушкинских Гор или *Pedicularis kaufmannii* недалеко от Жижицы).

8.6. ФЛОРИСТИЧЕСКОЕ РАЙОНИРОВАНИЕ НА СТАТИСТИЧЕСКОЙ ОСНОВЕ

Одна из наиболее общих закономерностей формирования растительного покрова состоит в сопряженности развития флоры и растительности. Эта сопряженность имеет сложный характер. Флорогенез и развитие определенного типа растительности протекают совместно на одной и той же территории, но подчиняются не только общим, но и специфическим для них закономерностям. Флоре, например, в гораздо большей степени свойственны черты исторического консерватизма, и ее изменение происходит значительно медленнее изменения растительных сообществ. Хорошо известен тот факт, что одинаковые типы растительности могут формироваться на совершенно разной флористической основе.

Современный растительный покров следует рассматривать как комплексный результат развития флоры и растительности, не адекватный результатам развития их в отдельности. Флора, растительность и растительный покров — явления разного порядка, относящиеся к различным уровням системной организации растительного мира и подлежащие изучению разными отраслями ботанической науки: флористикой, геоботаникой и ботанической географией.

В соответствии с этим ботаническое районирование территорий может быть флористическим, геоботаническим и ботанико-географическим (комплексным). Биологический смысл, цели, методы и критерии, на основе которых могут быть выполнены три типа районирования, различны, а получаемые ими единицы районирования, как правило, территориально не совпадают, хотя в отдельных случаях [Толмачев, 1974] такие совпадения и могут иметь место.

Флористическое районирование исходит прежде всего из степени общности систематического состава флор, исторически сложившихся на данной территории, причем особое внимание уделяется таким показателям, как географическое распростра-

нение и происхождение, а также эндемизм таксонов различного ранга.

Геоботаническое районирование, учитывая сходство систематического состава растительных сообществ, больше опирается на их ландшафтообразующую похожесть, уделяя главное внимание оценке роли ведущих (эдификаторных, доминантных, активных и т. д.) видов растений, играющих наиболее существенную роль в сложении фитоценозов.

Ботанико-географическое районирование стремится объединить в одной схеме флористический и геоботанический аспекты формирования растительного покрова. Оно тесно связано с проблемами биогеоценологии, имеет большое значение для изучения биосферы, а его комплексный характер подчеркивается необходимостью более полного использования данных геоморфологии и климатологии.

Таким образом, если флористическое районирование сильнее отражает историю формирования растительного покрова, а геоботаническое районирование больше говорит о современном воздействии на него экологических факторов, то районирование комплексное призвано объединить и картографически отобразить одновременно несколько аспектов становления растительного покрова. Задача эта чрезвычайно сложная, и на пути ее решения возникают некоторые тенденции, с которыми нельзя согласиться.

Одна из них состоит в том, что флористическому и геоботаническому районированию отводится роль всего лишь двух сторон районирования ботанико-географического, которое признается единственно «естественным» или «природным». Между тем флористическое и геоботаническое районирование являются ничуть не менее «естественными» или «природными», имеют каждое самостоятельное значение для своей науки, а их роль далеко не ограничивается обеспечением выполнения районирования комплексного.

Другая тенденция заключается в стремлении подменить флористическое районирование комплексным или «улучшить» его путем дополнительного введения геоботанических критериев, например критериев встречаемости или активности видов. Однако эти критерии, столь важные в геоботанике, не имеют такого же значения во флористике. Единичные находения редких видов могут решающим образом отразиться на флористическом районировании, не повлияв на районирование геоботаническое. Напротив, доминирующие в растительном покрове виды, привлекающие особое внимание геоботаника, не играют существенной роли при флористическом районировании.

Каждый тип районирования основывается на собственных критериях. Критериев флористического районирования существует несколько. Все они в конечном счете учитывают степень

сходства систематического состава флор, но отдают предпочтение разным аспектам этого сходства. Предложить единый «рецепт» на все случаи выделения флористических подразделений (фитохорий) различного ранга нельзя, ибо на разных уровнях флористической иерархии роль одних и тех же флористических показателей далеко не одинакова.

При выделении крупных фитохорий важная роль принадлежит критерию эндемизма [Тахтаджян, 1978]. Так, например, для флористических царств помимо высокого уровня видового и родового эндемизма характерно наличие эндемичных семейств; для флористических областей — эндемичных родов, тогда как для флористических провинций уровень родового эндемизма значительно ниже. С понижением ранга фитохорий падает и уровень видового эндемизма. С переходом к мелким флористическим подразделениям (округам, районам) критерий эндемизма утрачивает свое значение.

Важным критерием, выделяющим флористические области, является систематическая структура [Толмачев, 1974; Тахтаджян, 1978]. При этом систематическая структура более мелких (вплоть до конкретных) флор, сохраняя отличительные для своей области черты, в целом более изменчива, и этим обстоятельством можно пользоваться для детализации районирования относительно небольших территорий.

На низших уровнях флористического районирования (при выделении флористических округов и районов) решающее значение приобретает критерий сходства видового состава флор.

Используя этот критерий, мы будем исходить из следующих методических установок.

Виды, распространенные по всей территории, для ее флористического районирования никакого значения не имеют и из дальнейшего анализа исключаются. Оставшиеся виды, образующие на исследуемой территории границы своих ареалов, объединяются в группы, которые составляют каждая свой элемент флоры. Напомним, что под этим термином мы подразумеваем миграционно-генетический элемент флоры, т. е. группу видов, имеющих общие пути и время проникновения на данную территорию.

Анализ сходства различных частей районированной территории строится не на совпадении конкретных видов, а на соотношении числа видов каждого элемента флоры, т. е. на степени представительства разных элементов флоры в различных частях территории. Это обусловлено тем, что все виды, входящие в состав данного элемента флоры, свидетельствуют о наличии одной и той же исторической закономерности флорогенеза. Такой подход допускает формализацию исходных данных для математической обработки.

В качестве примера рассмотрим флористическое районирование Псковской области. Флора этой области, лежащей на

стыке Балтийского и Средне-Русского флористических округов, насчитывает в своем составе более 1115 видов высших растений (без мохообразных). Из этого числа было выделено 216 видов, образующих в пределах области границы своих ареалов и принадлежащих к 12 различным элементам флоры.

Характеристика элементов флоры Псковской области и перечень относящихся к ним видов даны в упомянутой ранее статье [Баранова, Миняев, Шмидт, 1971]. Поэтому здесь мы ограничимся перечислением учтенных элементов флоры, указав число входящих в них видов растений: 1) арктический и гипоарктический океанический — 5; 2) гипоарктический континентальный — 6; 3) евросибирский таежный — 19; 4) евросибирский подтаежный — 33; 5) евросибирский пустынно-степной — 18; 6) восточноевропейский широколиственный (сарматский) — 26; 7) паневропейский широколиственный (неморальный) — 47; 8) средневропейский широколиственный (неморальный) — 12; 9) средневропейский горный (альпийский) — 9; 10) средневропейский горный (карпатский) — 7; 11) балтийский — 6 и 12) расширенно-океанический (атлантический) — 28.

Для всех 216 видов, представляющих перечисленные элементы флоры и образующих в пределах Псковской области границы своих ареалов, были составлены точечные карты распространения их на этой территории.

Затем Псковская область, начиная от пересечения 29-го меридиана и 58-й параллели, была разбита на квадраты со стороной 31,5 км, т. е. площадью около 1000 кв. км. Выбор такой величины учетного квадрата был обусловлен необходимостью соблюдения двух противоречивых условий: квадрат должен быть достаточно мал, чтобы обеспечить точность районирования, и вместе с тем достаточно велик, чтобы свести к минимуму неизбежное (даже при хорошей изученности территории) влияние неполноты имеющихся сведений о местонахождениях видов растений.

Путем наложения кальки с сеткой квадратов на карты ареалов было учтено количество видов каждого элемента флоры, встречающихся в каждом квадрате. В качестве меры сходства между квадратами был использован коэффициент Сёренсена — Чекановского

$$K_{sc} = \frac{2c}{a + b},$$

где, применительно к нашей задаче, a и b — число видов всех 12 групп (элементов флоры) в первом и во втором квадратах; c — число общих для двух квадратов единиц всех элементов флоры (напомним, что это не обязательно должны быть одинаковые виды: важно, чтобы они (хотя бы и разные) относились к одному и тому же элементу флоры). Таким образом, величина c характеризует не совпадение одинаковых видов, а со-

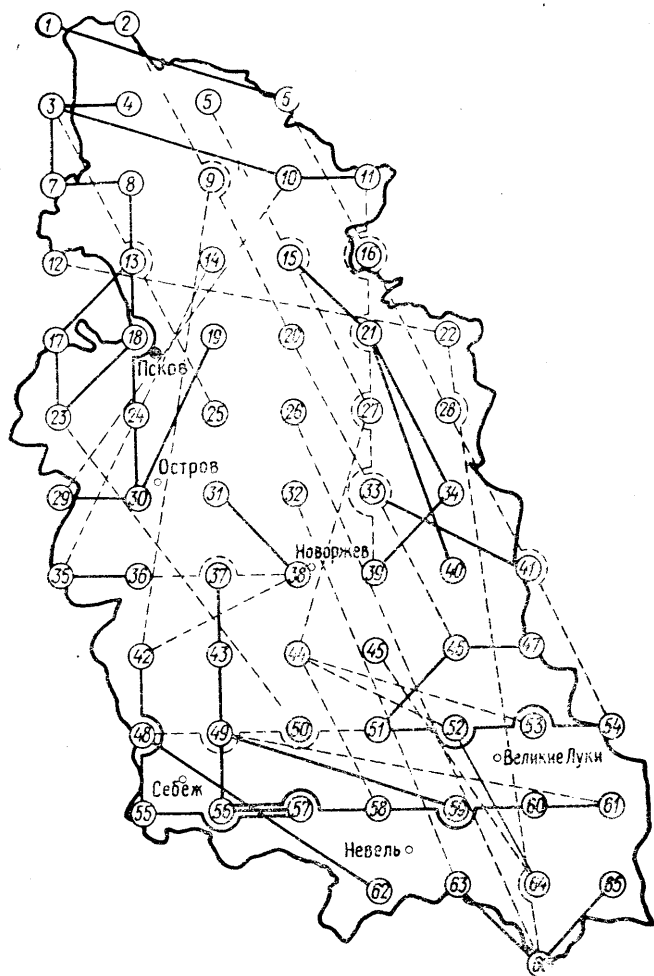


Рис. 47. Наложенный на карту Псковской области дендрит, соединяющий квадраты, наиболее сходные по наличию тех или иных элементов флоры. Цифры в кружках — центры соответствующих квадратов; сплошные линии — внутрирайонные, а пунктирные — межрайонные флористические связи.

впадение числа видов, относящихся к тому или иному флористическому элементу.

Этим способом было рассчитано 2016 коэффициентов сходства (корреляционная матрица не может быть здесь воспроизведена в силу ее внушительных размеров) и построен дендрит, нанесенный затем на карту Псковской области (рис. 47). Для выделения флористических районов дендрит сам по себе не нужен, так как главную роль играют полученные корреляционные плеяды. Однако построение дендрита в качестве промежуточного этапа работы представляется весьма полезным.

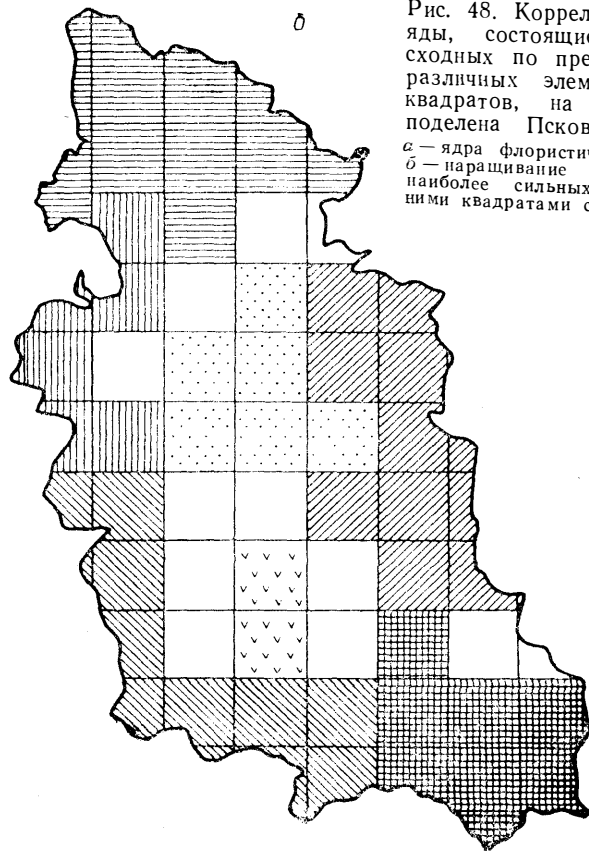
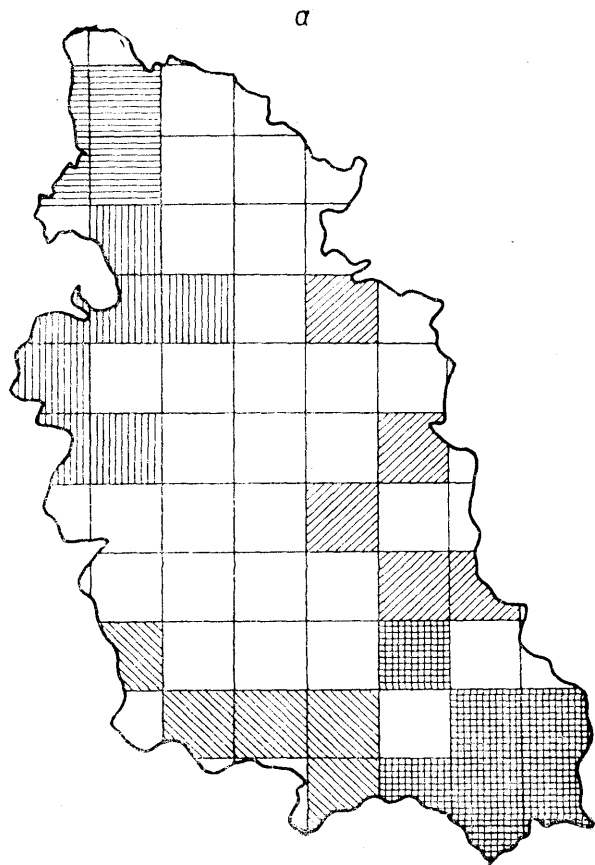


Рис. 48. Корреляционные плеяды, состоящие из наиболее сходных по представительству различных элементов флоры квадратов, на которые была поделена Псковская область.

а — ядра флористических районов;
 б — наращивание плеяд за счет наиболее сильных между соседними квадратами связей.

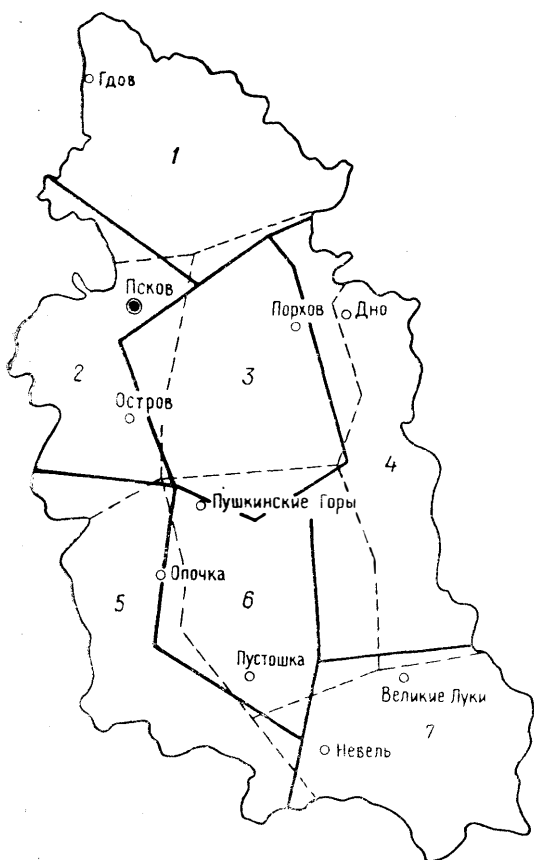


Рис. 49. Флористическое районирование Псковской области.

Районы: 1—Северный, 2—Западный, 3—Центрально-Судомский, 4—Восточный, 5—Юго-Западный, 6—Центрально-Бежаницкий, 7—Юго-Восточный. Сплошными линиями показано районирование, выполненное на статистической основе, а пунктирными — ранее и независимо от него произведенное глазомерное районирование.

Во-первых, при построении дендрита (рис. 47) обнаружилось расчленение территории Псковской области на две части по линии Плюсса—Остров—Невель, что соответствует представлению об упомянутом ранее стыке между Балтийским и Средне-Русским флористическими округами. Во-вторых, построение дендрита способом «максимального корреляционного пути» обеспечивает естественный ход выделения корреляционных плеяд (см. раздел 7.3.2).

Обычным путем, посредством повышения уровня связи в этом дендрите, были выделены корреляционные плеяды (рис. 48, а), состоящие из наиболее сходных по исследуемому признаку квадратов и представляющие собой ядра будущих флористических районов. В дальнейшем было произведено «наращивание» выделенных плеяд (рис. 48, б) за счет наиболее сильно связанных с ними смежных квадратов, в результате чего обнаружилось разделение Псковской области на 7 флористических районов, причем окончательные границы между

ними проводились с учетом преимущественных связей соседних квадратов на фоне ландшафтных условий местности.

Произведенное районирование в основном совпало с независимыми от него прежними глазомерными соображениями [Миняев, Шмидт, 1970], основанными на том же флористическом принципе, но содействовало уточнению границ флористических районов (рис. 49), сдвиг которых против ранее установленных составил местами 30—45 км (при максимальной ширине всей области 200—210 км).

Вычисление средних арифметических из внутрирайонных коэффициентов сходства показало, что флористически наиболее целостным является Западный район, а наименее — Центрально-Судомский. Это вполне соответствует сложившимся в ходе исследования флоры Псковской области представлениям о степени самобытности ее отдельных частей.

В заключение укажем, что при уточнении положения крупных флористических рубежей важная роль принадлежит критерию систематической структуры флор. Соответствующие примеры можно найти в специальной монографии [Шмидт, 1980].

ПРИЛОЖЕНИЯ¹

Приложение I. Случайные числа

3393	6270	4228	6069	9407	1855	8549	3217	2351	8410
9168	2330	2157	7416	0398	6173	1703	8132	9065	6717
7891	3590	2502	5945	3402	0491	4328	2365	6175	7695
9085	6307	6910	9174	1753	1797	9229	3422	9831	8357
2638	2908	6368	0398	5495	3283	0031	5955	6544	3883
1313	8338	0623	8600	4950	5414	7131	0134	7241	0651
3897	4202	3814	3575	1599	1649	2784	1994	5775	1406
4380	9543	1646	2859	8415	9120	8062	2421	6161	4634
1618	6309	7909	0874	0401	4301	4517	9197	3350	0434
4858	4676	7363	9141	6133	0549	1972	3461	7116	1496
5354	9142	0847	5393	5416	6505	7156	5634	9703	6221
0905	6986	9396	3975	9255	0537	2479	4589	0562	5345
1420	0470	8679	2328	3939	1292	0406	5428	3789	2882
3218	9080	6604	1813	8209	7039	2083	3369	4437	3798
9697	8431	4387	0622	6893	8788	2320	9358	5904	9539
0912	4964	0502	9683	4636	2861	2876	1273	7870	2030
4636	7072	4838	0601	3894	7182	8417	2367	7032	1003
2515	4734	9878	6761	5636	2949	3979	8650	3430	0635
5964	0412	5012	2369	6461	0678	3693	2928	3740	8947
7848	1523	7904	1521	1455	7089	8094	9872	0898	7174
5192	2571	3643	0707	3434	6818	5729	8615	4298	4129
8438	8325	9886	1805	0226	2310	3675	5058	2515	2388
8166	6349	0319	5436	6838	2460	6433	0644	7428	8556
9158	8263	6504	2562	1160	1526	1816	9690	1215	9590
6061	3525	4048	0382	4224	7148	8259	6526	5340	4034

¹ Вспомогательные таблицы I—XI заимствованы из руководств по биометрии, указанных в списке литературы.

Приложение II. Критерий τ_{st} для исключения крайних вариант

n	τ_{st}		n	τ_{st}		n	τ_{st}	
	0,05	0,01		0,05	0,01		0,05	0,01
5	1,92	1,97	21	2,80	3,11	80	3,33	3,70
6	2,70	2,16	22	2,82	3,13	90	3,37	3,74
7	2,18	2,31	23	2,84	3,16	100	3,40	3,77
8	2,27	2,43	24	2,86	3,18	120	3,46	3,83
9	2,35	2,53	25	2,88	3,20	150	3,53	3,90
10	2,41	2,62	26	2,90	3,22	200	3,61	3,98
11	2,47	2,69	27	2,91	3,24	300	3,73	4,09
12	2,52	2,75	28	2,93	3,26	400	3,80	4,17
13	2,56	2,81	29	2,94	3,28	500	3,87	4,24
14	2,60	2,86	30	2,96	3,29	600	3,92	4,28
15	2,64	2,90	35	3,02	3,36	700	3,96	4,32
16	2,67	2,94	40	3,08	3,42	800	3,99	4,35
17	2,70	2,98	45	3,12	3,48	900	4,02	4,38
18	2,73	3,02	50	3,16	3,52	1000	4,05	4,41
19	2,75	3,05	60	3,22	3,58	1500	4,14	4,50
20	2,78	3,08	70	3,28	3,64	2000	4,21	4,56

Приложение III. Стандартные значения критерия t_{st} Стьюдента

v	0,05	0,01	0,001	v	0,05	0,01	0,001
1	12,71	63,66	—	21	2,08	2,83	3,82
2	4,30	9,93	31,60	22	2,07	2,82	3,79
3	3,18	5,84	12,94	23	2,07	2,81	3,77
4	2,78	4,60	8,61	24	2,06	2,80	3,75
5	2,57	4,03	6,87	25	2,06	2,79	3,73
6	2,45	3,71	5,96	26	2,06	2,78	3,71
7	2,37	3,50	5,41	27	2,05	2,77	3,69
8	2,31	3,36	5,04	28	2,05	2,76	3,67
9	2,26	3,25	4,78	29	2,04	2,76	3,66
10	2,23	3,17	4,59	30	2,04	2,75	3,65
11	2,20	3,11	4,44	40	2,02	2,70	3,55
12	2,18	3,06	4,32	50	2,01	2,68	3,50
13	2,16	3,01	4,22	60	2,00	2,66	3,46
14	2,15	2,98	4,14	80	1,99	2,64	3,42
15	2,13	2,95	4,07	100	1,98	2,63	3,39
16	2,12	2,92	4,02	120	1,98	2,62	3,37
17	2,11	2,90	3,97	200	1,97	2,60	3,34
18	2,10	2,88	3,92	500	1,96	2,59	3,31
19	2,09	2,86	3,88	∞	1,96	2,58	3,29
20	2,09	2,85	3,85				

Примечание. Различие считается достоверным, если рассчитанное значение t превышает стандартное при выбранном уровне существенности.

Приложение IV. Критические значения F-критерия Фишера

α_1 α_2		Степени свободы для											
		1	2	3	4	5	6	7	8	9	10	11	12
Степени свободы для меньшей дисперсии	1	161 4052	200 4999	216 5403	225 5625	230 5764	234 5889	237 5928	239 5981	241 6022	242 6056	243 6082	244 6106
	2	18,51 98,49	19,00 99,01	19,16 99,17	19,25 99,25	19,30 99,30	19,33 99,33	19,36 99,34	19,37 99,36	19,38 99,38	19,39 99,40	19,40 99,41	19,41 99,42
	3	10,13 34,12	9,55 30,81	9,28 29,46	9,12 28,71	9,01 28,24	8,94 27,91	8,88 27,67	8,84 27,49	8,81 27,34	8,78 27,23	8,76 27,13	8,74 27,05
	4	7,71 21,20	6,94 18,00	6,59 16,69	6,39 15,98	6,26 15,52	6,16 15,21	6,09 14,98	6,04 14,80	6,00 14,66	5,96 14,54	5,93 14,45	5,91 14,37
	5	6,61 16,26	5,79 13,27	5,41 12,06	5,19 11,39	5,05 10,97	4,95 10,67	4,88 10,45	4,82 10,27	4,78 10,15	4,74 10,05	4,70 9,96	4,68 9,89
	6	5,99 13,74	5,14 10,92	4,76 9,78	4,53 9,15	4,39 8,75	4,28 8,47	4,21 8,26	4,15 8,10	4,10 7,98	4,06 7,87	4,03 7,79	4,00 7,72
	7	5,59 12,25	4,74 9,55	4,35 8,45	4,12 7,85	3,97 7,46	3,87 7,19	3,79 7,00	3,73 6,84	3,68 6,71	3,63 6,62	3,60 6,54	3,57 6,47
	8	5,32 11,26	4,46 8,65	4,07 7,59	3,84 7,01	3,69 6,63	3,58 6,37	3,50 6,19	3,44 6,03	3,39 5,91	3,34 5,82	3,31 5,74	3,28 5,67
	9	5,12 10,56	4,26 8,02	3,86 6,99	3,63 6,42	3,48 6,06	3,37 5,80	3,29 5,62	3,23 5,47	3,18 5,35	3,13 5,26	3,10 5,18	3,07 5,11
	10	4,96 10,04	4,10 7,55	3,71 6,55	3,48 5,99	3,33 5,64	3,22 5,39	3,14 5,21	3,07 5,06	3,02 4,95	2,97 4,85	2,94 4,78	2,91 4,71
	11	4,84 9,65	3,98 7,20	3,59 6,22	3,36 5,67	3,20 5,32	3,09 5,07	3,01 4,88	2,95 4,74	2,90 4,63	2,86 4,54	2,82 4,46	2,79 4,40
	12	4,75 9,33	3,88 6,93	3,49 5,95	3,26 5,41	3,11 5,06	3,00 4,82	2,92 4,65	2,85 4,50	2,80 4,39	2,76 4,30	2,72 4,22	2,69 4,16
	13	4,67 9,07	3,80 6,70	3,41 5,74	3,18 5,20	3,02 4,86	2,92 4,62	2,84 4,44	2,77 4,30	2,72 4,19	2,67 4,10	2,63 4,02	2,60 3,96
	14	4,60 8,86	3,74 6,51	3,34 5,56	3,11 5,03	2,96 4,69	2,85 4,46	2,77 4,28	2,70 4,14	2,65 4,03	2,60 3,94	2,56 3,86	2,53 3,80
	15	4,54 8,68	3,68 6,36	3,29 5,42	3,06 4,89	2,90 4,56	2,79 4,32	2,70 4,14	2,64 4,00	2,57 3,89	2,55 3,80	2,51 3,73	2,48 3,67
	16	4,49 8,53	3,63 6,23	3,24 5,29	3,01 4,77	2,85 4,44	2,74 4,20	2,66 4,03	2,50 3,89	2,54 3,78	2,49 3,69	2,45 3,61	2,42 3,55
	17	4,45 8,40	3,59 6,11	3,20 5,18	2,96 4,67	2,81 4,34	2,70 4,10	2,62 3,93	2,55 3,79	2,50 3,68	2,45 3,59	2,41 3,52	2,38 3,45

(F_{05} — верхние, F_{01} — нижние числа)

большой дисперсии													v_2
14	16	20	24	30	40	50	75	100	200	500	∞		
245	246	248	249	250	251	252	253	253	254	254	254	254	1
6142	6169	6208	6234	6258	6286	6302	6323	6334	6352	6361	6366		
19,42	19,43	19,44	19,45	19,46	19,47	19,47	19,48	19,49	19,49	19,50	19,50	19,50	2
99,43	99,44	99,45	99,46	99,47	99,48	99,48	99,49	99,49	99,49	99,50	99,50		
8,71	8,69	8,66	8,64	8,62	8,60	8,58	8,57	8,56	8,54	8,54	8,53	8,53	3
26,92	26,83	26,69	26,60	26,50	26,41	26,35	26,27	26,23	26,18	26,14	26,12		
5,87	5,84	5,80	5,77	5,74	5,71	5,70	5,68	5,66	5,65	5,64	5,63	5,63	4
14,24	14,15	14,02	13,93	13,83	13,74	13,69	13,61	13,57	13,52	13,48	13,46		
4,64	4,60	4,56	4,53	4,50	4,46	4,44	4,42	4,40	4,38	4,37	4,36	4,36	5
9,77	9,68	9,55	9,47	9,38	9,29	9,24	9,17	9,13	9,07	9,04	9,02		
3,96	3,92	3,87	3,84	3,81	3,77	3,75	3,72	3,71	3,69	3,68	3,67	3,67	6
7,60	7,52	7,39	7,31	7,23	7,14	7,09	7,02	6,99	6,94	6,90	6,88		
3,52	3,49	3,44	3,41	3,38	3,34	3,32	3,29	3,28	3,25	3,24	3,23	3,23	7
6,35	6,27	6,15	6,07	5,98	5,90	5,85	5,78	5,75	5,70	5,67	5,65		
3,23	3,20	3,15	3,12	3,08	3,05	3,03	3,00	2,98	2,96	2,94	2,93	2,93	8
5,56	5,48	5,36	5,28	5,20	5,11	5,06	5,00	4,96	4,91	4,88	4,86		
3,02	2,98	2,93	2,90	2,86	2,82	2,80	2,77	2,76	2,73	2,72	2,71	2,71	9
5,00	4,92	4,80	4,73	4,64	4,56	4,51	4,45	4,41	4,36	4,33	4,31		
2,86	2,82	2,77	2,74	2,70	2,67	2,64	2,61	2,59	2,56	2,55	2,54	2,54	10
4,60	4,52	4,41	4,33	4,25	4,17	4,12	4,05	4,01	3,96	3,93	3,91		
2,74	2,70	2,65	2,61	2,57	2,53	2,50	2,47	2,45	2,42	2,41	2,40	2,40	11
4,29	4,21	4,10	4,02	3,94	3,86	3,80	3,74	3,70	3,66	3,62	3,60		
2,64	2,60	2,54	2,50	2,46	2,42	2,40	2,36	2,35	2,32	2,31	2,30	2,30	12
4,05	3,98	3,86	3,78	3,70	3,61	3,56	3,49	3,46	3,41	3,38	3,36		
2,55	2,51	2,46	2,42	2,38	2,34	2,32	2,28	2,26	2,24	2,22	2,21	2,21	13
3,85	3,78	3,67	3,59	3,51	3,42	3,37	3,30	3,27	3,21	3,18	3,16		
2,48	2,44	2,39	2,35	2,31	2,27	2,24	2,21	2,19	2,16	2,14	2,13	2,13	14
3,70	3,62	3,51	3,43	3,34	3,26	3,21	3,14	3,11	3,06	3,02	3,00		
2,43	2,39	2,33	2,29	2,25	2,21	2,18	2,15	2,12	2,10	2,08	2,07	2,07	15
3,56	3,48	3,36	3,29	3,20	3,12	3,07	3,00	2,97	2,92	2,89	2,87		
2,37	2,33	2,28	2,24	2,20	2,16	2,13	2,09	2,07	2,04	2,02	2,01	2,01	16
3,45	3,37	3,25	3,18	3,10	3,01	2,96	2,89	2,86	2,80	2,77	2,75		
2,33	2,29	2,23	2,19	2,15	2,11	2,08	2,04	2,02	1,99	1,87	1,96	1,96	17
3,35	3,27	3,16	3,08	3,00	2,92	2,86	2,79	2,76	2,70	2,67	2,65		

$v_2 \backslash v_1$	Степени свободы для											
	1	2	3	4	5	6	7	8	9	10	11	12
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34
	8,28	6,01	5,09	4,58	4,25	4,01	3,85	3,71	3,60	3,51	3,44	3,37
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,34	2,31
	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,31	2,28
	8,10	5,85	4,94	4,43	4,10	3,87	3,71	3,56	3,45	3,37	3,30	3,23
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25
	8,02	5,78	4,87	4,37	4,04	3,81	3,65	3,51	3,40	3,31	3,24	3,17
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	2,26	2,23
	7,94	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28	2,24	2,20
	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07
24	4,26	3,40	3,01	2,72	2,62	2,51	2,43	2,36	2,30	2,26	2,22	2,18
	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,25	3,17	3,09	3,03
25	4,24	3,38	2,99	2,76	2,60	2,49	2,41	2,34	2,28	2,24	2,20	2,16
	7,77	5,57	4,68	4,18	3,86	3,63	3,46	3,32	3,21	3,13	3,05	2,99
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15
	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,17	3,09	3,02	2,96
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,30	2,25	2,20	2,16	2,13
	7,63	5,49	4,60	4,11	3,79	3,56	3,39	3,26	3,14	3,06	2,98	2,93
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,19	2,15	2,12
	7,64	5,45	4,57	4,07	3,76	3,53	3,36	3,23	3,11	3,03	2,95	2,90
29	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28	2,22	2,18	2,14	2,10
	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,08	3,00	2,92	2,87
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,16	2,12	2,09
	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,06	2,98	2,90	2,84
32	4,15	3,30	2,90	2,67	2,51	2,40	2,32	2,25	2,19	2,14	2,10	2,07
	7,50	5,34	4,46	3,97	3,66	3,42	3,25	3,12	3,01	2,94	2,86	2,80
34	4,13	3,28	2,88	2,65	2,49	2,38	2,30	2,23	2,17	2,12	2,08	2,05
	7,44	5,29	4,42	3,93	3,61	3,38	3,21	3,08	2,97	2,89	2,82	2,76
36	4,11	3,26	2,86	2,63	2,48	2,36	2,28	2,21	2,15	2,10	2,06	2,03
	7,39	5,25	4,38	3,89	3,58	3,35	3,18	3,04	2,94	2,86	2,78	2,72
38	4,10	3,25	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,05	2,02
	7,35	5,21	4,34	3,86	3,54	3,32	3,15	3,02	2,91	2,82	2,75	2,69

большей дисперсии												v_2
14	16	20	24	30	40	50	75	100	200	500	∞	
2,29	2,25	2,19	2,15	2,11	2,07	2,04	2,00	1,98	1,95	1,93	1,92	18
3,27	3,19	3,07	3,00	2,91	2,83	2,78	2,71	2,68	2,62	2,59	2,57	
2,26	2,21	2,15	2,11	2,07	2,02	2,00	1,96	1,94	1,91	1,90	1,88	19
3,19	3,12	3,00	2,92	2,84	2,76	2,70	2,63	2,60	2,54	2,51	2,49	
2,23	2,18	2,12	2,08	2,04	1,99	1,96	1,92	1,90	1,87	1,85	1,84	20
3,13	3,05	2,94	2,86	2,77	2,69	2,63	2,56	2,53	2,47	2,44	2,42	
2,20	2,15	2,09	2,05	2,00	1,96	1,93	1,89	1,87	1,84	1,82	1,81	21
3,07	2,99	2,88	2,80	2,72	2,63	2,58	2,51	2,47	2,42	2,38	2,36	
2,18	2,13	2,07	2,03	1,98	1,93	1,91	1,87	1,84	1,81	1,80	1,78	22
3,02	2,94	2,83	2,75	2,67	2,58	2,53	2,46	2,42	2,37	2,33	2,31	
2,14	2,10	2,04	2,00	1,96	1,91	1,88	1,84	1,82	1,79	1,77	1,76	23
2,97	2,89	2,78	2,70	2,62	2,53	2,48	2,41	2,37	2,32	2,28	2,26	
2,13	2,09	2,02	1,98	1,94	1,89	1,85	1,82	1,80	1,76	1,74	1,73	24
2,93	2,85	2,74	2,66	2,58	2,49	2,44	2,36	2,33	2,27	2,23	2,21	
2,11	2,06	2,00	1,96	1,92	1,87	1,84	1,80	1,77	1,74	1,72	1,71	25
2,89	2,81	2,70	2,62	2,54	2,45	2,40	2,32	2,29	2,23	2,19	2,17	
2,09	2,05	1,99	1,95	1,90	1,85	1,82	1,78	1,76	1,72	1,70	1,69	26
2,86	2,77	2,63	2,58	2,50	2,41	2,36	2,28	2,25	2,19	2,15	2,13	
2,08	2,03	1,97	1,93	1,88	1,84	1,80	1,76	1,74	1,71	1,68	1,67	27
2,83	2,74	2,63	2,55	2,47	2,38	2,33	2,25	2,21	2,16	2,12	2,10	
2,07	2,02	1,96	1,91	1,87	1,81	1,78	1,75	1,72	1,69	1,67	1,65	28
2,80	2,71	2,60	2,52	2,44	2,35	2,30	2,22	2,18	2,13	2,09	2,06	
2,05	2,00	1,94	1,90	1,85	1,80	1,77	1,73	1,71	1,68	1,65	1,64	29
2,77	2,68	2,57	2,49	2,41	2,32	2,27	2,19	2,15	2,10	2,06	2,03	
2,04	1,99	1,93	1,89	1,84	1,79	1,76	1,72	1,69	1,66	1,64	1,62	30
2,74	2,66	2,55	2,47	2,38	2,29	2,24	2,16	2,13	2,07	2,03	2,01	
2,02	1,97	1,91	1,86	1,82	1,76	1,74	1,69	1,67	1,64	1,61	1,59	32
2,70	2,62	2,51	2,42	2,34	2,25	2,20	2,12	2,08	2,02	1,98	1,96	
2,00	1,95	1,89	1,84	1,80	1,74	1,71	1,67	1,64	1,61	1,59	1,57	34
2,66	2,58	2,47	2,38	2,30	2,21	2,15	2,08	2,04	1,98	1,94	1,91	
1,98	1,93	1,87	1,82	1,78	1,72	1,69	1,65	1,62	1,59	1,56	1,55	36
2,62	2,54	2,43	2,35	2,26	2,17	2,12	2,04	2,00	1,94	1,90	1,87	
1,96	1,92	1,85	1,80	1,76	1,71	1,67	1,63	1,60	1,57	1,54	1,53	38
2,59	2,51	2,40	2,32	2,22	2,14	2,08	2,00	1,97	1,90	1,86	1,84	

		Степени свободы для											
		1	2	3	4	5	6	7	8	9	10	11	12
Степени свободы для меньшей дисперсии	40	4,08 7,31	3,23 5,18	2,84 4,31	2,61 3,83	2,45 3,51	2,34 3,29	2,25 3,12	2,18 2,99	2,12 2,88	2,07 2,80	2,04 2,73	2,00 2,66
	42	4,07 7,27	3,22 5,15	2,83 4,29	2,59 3,83	2,44 3,49	2,32 3,26	2,24 3,10	2,17 2,93	2,11 2,86	2,06 2,77	2,02 2,70	1,99 2,64
	44	4,06 7,24	3,21 5,12	2,82 4,26	2,58 3,78	2,43 3,46	2,31 3,24	2,23 3,07	2,16 2,94	2,10 2,84	2,05 2,75	2,01 2,68	1,98 2,62
	46	4,05 7,21	3,20 5,10	2,81 4,24	2,57 3,76	2,42 3,44	2,30 3,22	2,22 3,05	2,14 2,92	2,09 2,82	2,04 2,73	2,00 2,66	1,97 2,60
	48	4,04 7,19	3,19 5,08	2,81 4,22	2,56 3,74	2,41 3,42	2,30 3,20	2,21 3,04	2,14 2,90	2,08 2,80	2,03 2,71	1,99 2,64	1,96 2,58
	50	4,03 7,17	3,18 5,05	2,79 4,20	2,56 3,72	2,40 3,41	2,29 3,18	2,20 3,02	2,13 2,88	2,07 2,78	2,02 2,70	1,98 2,62	1,95 2,56
	55	4,02 7,12	3,17 5,01	2,78 4,16	2,54 3,68	2,38 3,37	2,27 3,15	2,18 2,98	2,11 2,85	2,05 2,75	2,00 2,66	1,97 2,59	1,93 2,53
	60	4,00 7,08	3,15 4,98	2,76 4,13	2,52 3,65	2,37 3,34	2,25 3,12	2,17 2,95	2,10 2,82	2,04 2,72	1,99 2,63	1,95 2,56	1,92 2,50
	65	3,99 7,04	3,14 4,95	2,75 4,10	2,51 3,62	2,36 3,31	2,24 3,09	2,15 2,93	2,08 2,79	2,01 2,70	1,98 2,61	1,94 2,54	1,90 2,47
	70	3,98 7,01	3,13 4,92	2,74 4,08	2,50 3,60	2,35 3,29	2,23 3,07	2,14 2,91	2,07 2,77	2,01 2,67	1,97 2,59	1,93 2,51	1,89 2,45
	80	3,96 6,96	3,11 4,88	2,72 4,04	2,48 3,56	2,33 3,25	2,21 3,04	2,12 2,87	2,05 2,74	1,99 2,64	1,95 2,55	1,91 2,48	1,89 2,41
	100	3,94 6,90	3,09 4,82	2,70 3,98	2,46 3,51	2,30 3,20	2,19 2,99	2,10 2,82	2,03 2,69	1,97 2,59	1,92 2,51	1,88 2,43	1,85 2,36
	125	3,92 6,84	3,07 4,78	2,68 3,94	2,44 3,47	2,29 3,17	2,17 2,95	2,08 2,79	2,01 2,65	1,95 2,56	1,90 2,37	1,86 2,40	1,83 2,33
	150	3,91 6,81	3,06 4,75	2,67 3,91	2,43 3,44	2,27 3,14	2,16 2,92	2,07 2,76	2,00 2,62	1,94 2,53	1,89 2,44	1,85 2,47	1,82 2,30
	200	3,89 6,76	3,04 4,71	2,65 3,88	2,41 3,41	2,26 3,11	2,14 2,90	2,05 2,73	1,98 2,60	1,92 2,50	1,87 2,41	1,83 2,34	1,80 2,28
	400	3,86 6,70	3,02 4,65	2,62 3,83	2,39 3,36	2,23 3,06	2,12 2,85	2,03 2,69	1,96 2,55	1,90 2,46	1,85 2,37	1,81 2,29	1,78 2,23
	1000	3,85 6,66	3,00 4,61	2,61 3,81	2,38 3,34	2,22 3,04	2,10 2,82	2,02 2,66	1,95 2,53	1,89 2,43	1,84 2,34	1,80 2,26	1,76 2,20
	∞	3,84 6,64	2,99 4,60	2,60 3,78	2,37 3,32	2,21 3,02	2,09 2,80	2,01 2,64	1,94 2,51	1,88 2,41	1,83 2,32	1,79 2,24	1,75 2,18

Примечание. Различие считается достоверным, если рассчитанное значение F число.)

большой дисперсии												ν_2
14	16	20	24	30	40	50	75	100	200	500	∞	
1.95 2.56	1.90 2.49	1.84 2.37	1.79 2.29	1.74 2.20	1.69 2.11	1.66 2.05	1.61 1.97	1.59 1.94	1.55 1.88	1.53 1.84	1.51 1.81	40
1.94 2.54	1.89 2.46	1.82 2.35	1.78 2.26	1.73 2.17	1.68 2.08	1.64 2.02	1.60 1.94	1.57 1.91	1.54 1.85	1.51 1.80	1.49 1.78	42
1.92 2.52	1.88 2.44	1.81 2.32	1.76 2.24	1.72 2.15	1.66 2.06	1.63 2.00	1.58 1.92	1.56 1.88	1.52 1.82	1.50 1.78	1.48 1.75	44
1.91 2.59	1.87 2.42	1.80 2.30	1.75 2.22	1.71 2.13	1.65 2.04	1.62 1.98	1.57 1.90	1.54 1.86	1.51 1.80	1.48 1.76	1.46 1.72	46
1.90 2.48	1.86 2.40	1.79 2.28	1.74 2.20	1.70 2.11	1.64 2.02	1.61 1.96	1.56 1.88	1.53 1.84	1.50 1.78	1.47 1.73	1.45 1.70	48
1.90 2.46	1.85 2.39	1.78 2.26	1.74 2.18	1.69 2.10	1.63 2.00	1.60 1.94	1.55 1.85	1.52 1.82	1.48 1.76	1.46 1.71	1.44 1.68	50
1.88 2.43	1.83 2.35	1.76 2.23	1.72 2.15	1.67 2.06	1.61 1.96	1.58 1.90	1.52 1.82	1.50 1.78	1.46 1.71	1.43 1.66	1.41 1.64	55
1.86 2.40	1.81 2.32	1.75 2.20	1.70 2.12	1.65 2.03	1.59 1.93	1.56 1.87	1.50 1.79	1.48 1.74	1.44 1.68	1.41 1.63	1.39 1.60	60
1.85 2.37	1.80 2.30	1.73 2.18	1.68 2.09	1.63 2.00	1.57 1.90	1.54 1.84	1.49 1.76	1.46 1.71	1.42 1.64	1.39 1.60	1.37 1.56	65
1.84 2.35	1.79 2.28	1.72 2.15	1.67 2.07	1.62 1.98	1.56 1.88	1.53 1.82	1.47 1.74	1.45 1.69	1.40 1.62	1.37 1.55	1.35 1.53	70
1.82 2.32	1.77 2.24	1.70 2.11	1.65 2.03	1.60 1.94	1.54 1.84	1.51 1.78	1.45 1.70	1.42 1.65	1.38 1.57	1.35 1.52	1.32 1.49	80
1.79 2.26	1.75 2.19	1.68 2.06	1.63 1.98	1.57 1.89	1.51 1.79	1.48 1.73	1.42 1.64	1.39 1.59	1.34 1.51	1.30 1.46	1.28 1.43	100
1.77 2.23	1.72 2.15	1.65 2.03	1.60 1.94	1.55 1.85	1.49 1.75	1.45 1.68	1.39 1.59	1.36 1.54	1.31 1.46	1.27 1.40	1.25 1.37	125
1.76 2.20	1.71 2.12	1.64 2.00	1.59 1.91	1.54 1.83	1.47 1.72	1.44 1.66	1.37 1.56	1.34 1.51	1.29 1.43	1.25 1.37	1.22 1.33	150
1.74 2.17	1.69 2.09	1.62 1.97	1.57 1.88	1.52 1.79	1.45 1.69	1.42 1.62	1.35 1.53	1.32 1.48	1.26 1.39	1.22 1.33	1.19 1.28	200
1.72 2.12	1.67 2.04	1.60 1.92	1.54 1.84	1.49 1.74	1.42 1.64	1.38 1.57	1.32 1.47	1.28 1.42	1.22 1.32	1.16 1.24	1.13 1.19	400
1.70 2.09	1.65 2.01	1.58 1.89	1.52 1.81	1.47 1.71	1.41 1.61	1.36 1.54	1.30 1.44	1.26 1.38	1.19 1.28	1.13 1.19	1.08 1.11	1000
1.69 2.07	1.64 1.99	1.57 1.87	1.52 1.79	1.46 1.69	1.40 1.59	1.35 1.52	1.28 1.41	1.24 1.36	1.17 1.25	1.11 1.15	1.00 1.09	∞

превращает критическое при выбранном уровне существенности (F_{05} — верхнее, F_{01} — нижнее)

Приложение V. Критические значения критерия $\chi^2_{\text{ст}}$

ν	0,05	0,01	0,001	ν	0,05	0,01	0,001
1	3,841	6,635	10,827	16	26,296	32,000	39,252
2	5,991	9,210	13,815	17	27,587	33,409	40,790
3	7,815	11,345	16,266	18	28,869	34,805	42,312
4	9,488	13,277	18,467	19	30,144	36,191	43,820
5	11,070	15,086	20,517	20	31,410	37,566	45,315
6	12,592	16,812	22,457	21	32,671	38,932	46,797
7	14,067	18,475	24,322	22	33,924	40,289	48,268
8	15,507	20,090	26,125	23	35,172	41,638	49,728
9	16,919	21,666	27,877	24	36,415	42,980	51,179
10	18,307	23,209	29,588	25	37,652	44,314	52,620
11	19,675	24,725	31,264	26	38,885	45,642	54,052
12	21,026	26,217	32,909	27	40,113	46,963	55,476
13	22,362	27,688	34,528	28	41,337	48,278	56,893
14	23,685	29,141	36,123	29	42,557	49,588	58,302
15	24,996	30,578	37,697	30	43,773	50,892	59,703

Примечание. Различие считается достоверным, если рассчитанное значение превышает критическое при выбранном уровне существенности.

Приложение VI. Арифметический треугольник Паскаля

n	Биномиальные коэффициенты														
					1										
1					1										
2					1	2		1							
3					1	3	3	1							
4					1	4	6	4	1						
5					1	5	10	10	5	1					
6					1	6	15	20	15	6	1				
7					1	7	21	35	35	21	7	1			
8					1	8	28	56	70	56	28	8	1		
9					1	9	36	84	126	126	84	36	9	1	
10					1	10	45	120	210	252	210	120	45	10	1

и т. д.

Приложение VII. Значения функции e^{-M}

<i>M</i>	0	1	2	3	4	5	6	7	8	9
0,00	1,	9990	9980	9970	9960	9950	9940	9930	9920	9910
0,0	1,	9900	9802	9704	9608	9512	9418	9324	9231	9139
0,1	9048	8958	8869	8781	8694	8607	8521	8437	8353	8270
0,2	8187	8106	8025	7945	7866	7788	7711	7634	7558	7483
0,3	7408	7334	7261	7189	7118	7047	6977	6907	6839	6771
0,4	6703	6637	6570	6505	6440	6376	6313	6250	6188	6126
0,5	6065	6005	5945	5886	5827	5769	5712	5655	5599	5543
0,6	5488	5434	5379	5326	5273	5220	5169	5117	5066	5016
0,7	4966	4916	4858	4819	4771	4724	4677	4630	4584	4538
0,8	4493	4449	4404	4360	4317	4274	4232	4190	4148	4107
0,9	4066	4025	3985	3946	3906	3867	3829	3791	3753	3716
1,	3679	3329	3012	2725	2466	2231	2019	1827	1653	1496
2,	1353	1225	1108	1003	0907	0821	0743	0672	0608	0550
3,	0498	0450	0408	0369	0334	0302	0273	0247	0224	0202
4,	0183	0166	0150	0136	0123	0111	0100	0091	0082	0074
5,	0067	0061	0055	0050	0045	0041	0037	0034	0030	0027
6,	0025	0022	0020	0018	0017	0015	0014	0012	0011	0010
7,	0009	0008	0007	0007	0006	0006	0005	0005	0004	0004
8,	0003	0003	0003	0002	0002	0002	0002	0002	0002	0001
9,	0001	0001	0001	0001	0001	0001	0001	0001	0001	0000

Примечание. Нуль целых опущен.

*Приложение VIII. Ординаты нормальной кривой
(значения функции $\theta(t)$)*

<i>t</i>	0	1	2	3	4	5	6	7	8	9
0,0	39894	39892	39885	39876	39852	39844	39822	39797	39767	39733
0,1	39695	39654	39608	39559	39505	39448	39387	39322	39253	39181
0,2	39104	39024	38940	38853	38762	38667	38568	38466	38361	38251
0,3	38139	38023	37903	37779	37654	37524	37391	37255	37115	36973
0,4	36827	36678	36526	36371	36213	36053	35889	35723	35553	35381
0,5	35207	35029	34849	34667	34482	34294	34105	33912	33718	33521
0,6	33322	33121	32918	32713	32506	32297	32086	31874	31659	31443

Приложение VIII (продолжение)

t	0	1	2	3	4	5	6	7	8	9
0,7	31225	31006	30785	30563	30339	30114	29887	29659	29430	29200
0,8	28969	28737	28504	28269	28034	27798	27562	27324	27086	26848
0,9	26609	26369	26129	25888	25647	25406	25164	24923	24681	24439
1,0	24197	23955	23713	23471	23230	22988	22747	22506	22265	22025
1,1	21785	21546	21307	21069	20831	20594	20327	20121	19886	19652
1,2	19419	19186	18954	18724	18494	18265	18037	17810	17585	17360
1,3	17137	16915	16694	16474	16256	16038	15822	15608	15395	15183
1,4	14973	14764	14556	14350	14146	13943	13742	13542	13344	13147
1,5	12952	12758	12566	12376	12188	12001	11816	11632	11450	11270
1,6	11092	10915	10741	10577	10396	10226	10059	9893	9728	9566
1,7	09405	09246	09089	08933	08780	08628	08478	08329	08183	08038
1,8	07895	07754	07614	07477	07341	07206	07074	06943	06814	06687
1,9	06562	06438	06316	06195	06077	05959	05844	05730	05618	05508
2,0	05399	05292	05186	05082	04980	04879	04780	04682	04586	04491
2,1	04398	04307	04217	04128	04041	03955	03871	03788	03706	03626
2,2	03547	03470	03394	03319	03246	03174	03103	03034	02965	02898
2,3	02833	02768	02705	02643	02582	02522	02463	02406	02349	02294
2,4	02239	02186	02134	02083	02033	01984	01936	01888	01842	01797
2,5	01753	01709	01667	01625	01585	01545	01506	01468	01431	01394
2,6	01358	01323	01289	01256	01223	01191	01160	01130	01100	01071
2,7	01042	01014	00987	00961	00935	00909	00885	00861	00837	00814
2,8	00792	00770	00748	00727	00707	00687	00668	00649	00631	00613
2,9	00595	00578	00562	00545	00530	00514	00499	00485	00470	00457
3,0	00443	00430	00417	00405	00393	00381	00370	00358	00348	00337
3,1	00327	00317	00307	00298	00288	00279	00271	00262	00254	00246
3,2	00238	00231	00224	00216	00210	00203	00196	00190	00184	00178
3,3	00172	00167	00161	00156	00151	00146	00141	00136	00132	00127
3,4	00123	00119	00115	00111	00107	00104	00100	00097	00094	00090
3,5	00087	00084	00081	00079	00076	00073	00071	00068	00066	00063
3,6	00061	00059	00057	00055	00053	00051	00049	00047	00046	00044
3,7	00042	00041	00039	00038	00037	00035	00034	00033	00031	00030
3,8	00029	00028	00027	00026	00025	00024	00023	00022	00021	00021
3,9	00020	00019	00018	00018	00017	00016	00016	00015	00014	00014
4,0	00013	00009	00006	00004	00002	00002	00001	00001	00000	00000

Примечание. Ноль целых везде опущен.

Приложение IX. Значения величины z для r

r	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0100	0,0200	0,0300	0,0400	0,0501	0,0601	0,0701	0,0802	0,0902
0,1	0,1003	0,1105	0,1206	0,1308	0,1409	0,1511	0,1614	0,1717	0,1820	0,1923
0,2	0,2027	0,2132	0,2237	0,2342	0,2448	0,2554	0,2661	0,2769	0,2877	0,2986
0,3	0,3095	0,3206	0,3317	0,3428	0,3541	0,3654	0,3769	0,3884	0,4001	0,4118
0,4	0,4236	0,4356	0,4477	0,4599	0,4722	0,4847	0,4973	0,5101	0,5230	0,5361
0,5	0,5493	0,5627	0,5763	0,5901	0,6042	0,6188	0,6328	0,6475	0,6625	0,6777
0,6	0,6931	0,7089	0,7250	0,7414	0,7582	0,7753	0,7928	0,8107	0,8291	0,8480
0,7	0,8673	0,8872	0,9076	0,9287	0,9505	0,9730	0,9962	1,0203	1,0454	1,0714
0,8	1,0986	1,1270	1,1568	1,1881	1,2212	1,2562	1,2933	1,3331	1,3758	1,4219
0,9	1,4722	1,5275	1,5890	1,6584	1,7280	1,8318	1,9459	2,0923	2,2976	2,6467

Приложение X. Значения $-p \log_2 p$ или $-q \log_2 q$

$p; q$	0	1	2	3	4	5	6	7	8	9
0,00	00000	00996	01793	02514	03186	03821	04428	05010	05572	06116
0,01	06643	07156	07656	08144	08621	09088	09545	09993	10432	10863
0,02	11287	11704	12113	12517	12913	13304	13689	14069	14443	14812
0,03	15176	15535	15890	16240	16586	16927	17265	17598	17927	18253
0,04	18575	18893	19208	19519	19827	20132	20434	20732	21027	21320
0,05	21609	21896	22179	22460	22738	23014	23287	23557	23825	24090
0,06	24353	24613	24871	25127	25381	25632	25881	26127	26372	26615
0,07	26855	27093	27330	27564	27796	28027	28255	28482	28706	28929
0,08	29150	29370	29587	29803	30017	30229	30439	30648	30855	31061
0,09	31265	31467	31668	31867	32065	32261	32455	32648	32840	33030
0,10	33219	33406	33592	33776	33959	34141	34321	34500	34677	34853
0,11	35028	35202	35374	35545	35714	35883	36050	36216	36381	36544
0,12	36706	36867	37027	37186	37343	37500	37655	37809	37962	38113
0,13	38264	38413	38562	38709	38855	39001	39145	39289	39430	39571
0,14	39711	39849	39987	40124	40260	40395	40529	40661	40793	40924
0,15	41054	41183	41311	41438	41564	41689	41813	41937	42059	42181
0,16	42301	42421	42540	42658	42775	42891	43006	43120	43234	43346
0,17	43458	43569	43679	43788	43897	44005	44111	44217	44322	44427
0,18	44530	44633	44735	44836	44936	45036	45135	45233	45330	45426
0,19	45522	45617	45711	45805	45897	45989	46081	46171	46261	46350

Приложение X (продолжение)

p: q	0	1	2	3	4	5	6	7	8	9
0,20	46438	46526	46612	46699	46784	46869	46953	47036	47119	47201
0,21	47282	47362	47442	47521	47600	47678	47755	47831	47907	47982
0,22	48057	48131	48204	48276	48348	48420	48490	48560	48629	48698
0,23	48766	48834	48901	48967	49032	49097	49162	49225	49289	49351
0,24	49413	49474	49535	49595	49655	49714	49772	49830	49887	49943
0,25	50000	50055	50110	50164	50218	50271	50324	50376	50427	50478
0,26	50528	50578	50627	50676	50724	50772	50819	50865	50911	50957
0,27	51002	51046	51090	51133	51176	51218	51260	51301	51342	51382
0,28	51422	51461	51499	51537	51575	51612	51649	51685	51720	51755
0,29	51790	51824	51858	51891	51923	51955	51987	52018	52049	52079
0,30	52108	52138	52166	52195	52222	52250	52276	52303	52329	52354
0,31	52379	52403	52427	52451	52474	52497	52519	52541	52562	52583
0,32	52603	52623	52642	52661	52680	52698	52716	52733	52750	52766
0,33	52782	52797	52812	52827	52841	52855	52868	52881	52893	52905
0,34	52917	52928	52939	52949	52959	52968	52977	52986	52994	53002
0,35	53010	53017	53023	53029	53035	53040	53045	53050	53054	53058
0,36	53061	53064	53066	53069	53070	53072	53073	53073	53073	53073
0,37	53072	53071	53070	53068	53066	53063	53060	53057	53053	53049
0,38	53045	53040	53035	53029	53023	53017	53010	53003	52995	52987
0,39	52979	52971	52962	52952	52943	52932	52922	52911	52900	52889
0,40	52877	52864	52852	52839	52825	52812	52798	52783	52769	52753
0,41	52738	52722	52706	52689	52673	52655	52638	52620	52602	52583
0,42	52564	52545	52525	52505	52485	52464	52443	52422	52400	52378
0,43	52356	52333	52310	52287	52263	52239	52215	52190	52165	52140
0,44	52114	52088	52062	52035	52008	51981	51953	51925	51897	51869
0,45	51840	51810	51781	51751	51721	51690	51659	51628	51597	51565
0,46	51533	51501	51468	51435	51402	51368	51334	51300	51265	51230
0,47	51195	51160	51124	51088	51051	51015	50978	50940	50903	50865
0,48	50826	50788	50749	50710	50670	50631	50591	50550	50510	50469
0,49	50428	50385	50344	50302	50260	50217	50174	50131	50087	50044
0,50	50000	49955	49910	49855	49820	49775	49729	49683	49636	49589
0,51	49542	49495	49448	49400	49352	49303	49255	49206	49156	49207
0,52	49057	49007	48957	48906	48855	48804	48753	48701	48649	48597
0,53	48544	48491	48438	48385	48331	48277	48223	48169	48114	48059
0,54	48004	47948	47893	47836	47780	47724	47667	47610	47552	47495
0,55	47437	47379	47320	47262	47203	47143	47084	47024	46964	46904

Приложение X (продолжение)

р. г	0	1	2	3	4	5	6	7	8	9
0,56	46844	46783	46722	46661	46599	46537	46475	46413	46350	46288
0,57	46225	46161	46098	46034	54970	45906	45841	45776	45711	45646
0,58	45580	45514	45448	45382	45316	45249	45182	45114	45047	44979
0,59	44911	44843	44774	44706	44637	44567	44498	44428	44358	44288
0,60	44217	44147	44076	44005	43933	43862	43790	43718	43645	43573
0,61	43500	43427	43353	43280	43206	43132	43058	42983	42909	42834
0,62	42758	42683	42607	42531	42455	42379	42302	42226	42149	42071
0,63	41994	41916	41838	41760	41682	41603	41524	41445	41366	41286
0,64	41206	41126	41046	40966	40885	40804	40723	40642	40560	40478
0,65	40396	40314	40232	40149	40066	39983	39900	39816	39732	39648
0,66	39564	39480	39395	39310	39225	39140	39054	38968	38882	38796
0,67	38710	38623	38536	38449	38362	38275	38187	38099	38011	37923
0,68	37834	37746	37657	37567	37478	37388	37299	37209	37118	37028
0,69	36937	36847	36755	36664	36573	36481	36389	36297	36205	36112
0,70	36020	35927	35834	35740	35647	35553	35459	35365	35271	35176
0,71	35081	34986	34891	34796	34700	34604	34508	34412	34316	34219
0,72	34123	34026	33928	33831	33733	33636	33538	33440	33341	33243
0,73	33144	33045	32946	32846	32747	32647	32547	32447	32347	32246
0,74	32145	32044	31943	31842	31740	31639	31537	31435	31332	31230
0,75	31127	31024	30921	30818	30715	30611	30507	30403	30299	30195
0,76	30090	20985	29880	29775	29670	29564	29459	29353	29247	29140
0,77	29034	28927	28820	28713	28606	28499	28391	28283	28175	28067
0,78	27959	27850	27742	27633	27524	27414	27305	27195	27086	26976
0,79	26865	26755	26645	26534	26423	26312	26201	26089	25978	25866
0,80	25754	25642	25529	25417	25304	25191	25078	24965	24851	24738
0,81	24624	24510	24396	24282	24167	24052	23938	23823	23707	23592
0,82	23476	23361	23245	23129	23012	22896	22779	22663	22546	22429
0,83	22311	22194	22076	21958	21840	21722	21604	21485	21367	21248
0,84	21129	21010	20890	20771	20651	20531	20411	20291	20170	20050
0,85	19929	19808	19687	19566	19444	19323	19201	19079	18957	18835
0,86	18712	18590	18467	18344	18221	18098	17974	17851	17727	17603
0,87	17479	17355	17230	17106	16981	16856	16731	16606	16480	16355
0,88	16229	16103	15977	15851	15724	15598	15471	15344	15217	15090
0,89	14962	14835	14707	14579	14451	14323	14195	14066	13938	13809
0,90	13680	13551	13421	13292	13162	13032	12902	12772	12642	12512
0,91	12381	12250	12119	11988	11857	11726	11594	11463	11331	11199

Приложение X (продолжение)

	0	1	2	3	4	5	6	7	8	9
0,92	11067	10934	10802	10669	10536	10403	10270	10137	10004	09870
0,93	09736	09602	09468	09334	09200	09065	08631	08796	08661	08526
0,94	08391	08255	08120	07984	07848	07712	07576	07439	07303	07166
0,95	07030	06893	06756	06618	06481	06343	06206	06068	05930	05792
0,96	05653	05515	05376	05237	05099	04960	04820	04681	04541	04402
0,97	04262	04122	03982	03842	03701	03561	03420	03279	03138	02997
0,98	02856	02714	02573	02431	02289	02147	02005	01863	01720	01578
0,99	01435	01292	01149	01006	00863	00719	00575	00432	00288	00144

Примечание. Нуль целых опущен.

Приложение XI. Критические значения коэффициента ранговой корреляции ρ_s Спирмена

n	0,05	0,01	n	0,05	0,01	n	0,05	0,01
5	0,94		17	0,48	0,62	29	0,37	0,48
6	0,85		18	0,47	0,60	30	0,36	0,47
7	0,78	0,94	19	0,46	0,58	31	0,36	0,46
8	0,72	0,88	20	0,45	0,57	32	0,36	0,45
9	0,68	0,83	21	0,44	0,56	33	0,34	0,45
10	0,64	0,79	22	0,43	0,54	34	0,34	0,44
11	0,61	0,76	23	0,42	0,53	35	0,33	0,43
12	0,58	0,73	24	0,41	0,52	36	0,33	0,43
13	0,56	0,70	25	0,40	0,51	37	0,33	0,42
14	0,54	0,68	26	0,39	0,50	38	0,32	0,41
15	0,52	0,66	27	0,38	0,49	39	0,32	0,41
16	0,50	0,64	28	0,38	0,48	40	0,31	0,40

Примечание. Различие считается достоверным, если рассчитанное значение ρ_s превышает критическое при выбранном уровне существенности.

УКАЗАТЕЛЬ ЛИТЕРАТУРЫ

Аристархова М. Л., Иванюкович Л. К. Корреляционная структура количественных признаков культурных видов *Sorghum*.— Бот. журн., 1976, т. 61, № 2, с. 219—226.

Арциховский В. М. Рост как видовой признак у растений. — Изв. Имп. Бот. сада, 1906, т. 6, вып. 2 (отд. отд.).

Баранова Е. В. Материалы к анализу конкретных флор Псковской области. — Вестн. Ленингр. ун-та, 1973, № 15, с. 30—37.

Баранова Е. В. Анализ флоры Псковской области: Автореф. канд. дис. Л., 1977. 15 с.

Баранова Е. В., Миняев Н. А., Шмидт В. М. Флористическое районирование Псковской области на фитостатистической основе. — Вестн. Ленингр. ун-та, 1971, № 9, с. 30—40.

Бейли Н. Статистические методы в биологии. М., 1962. 260 с.

Бейли Н. Математика в биологии и медицине. М., 1970. 326 с.

Бернштейн А. Справочник статистических решений. М., 1968. 162 с.

Бежел Р. Флористические соотношения на островах Неоарктики. — Бот. журн., 1969, т. 54, № 6, с. 872—891.

Бобров А. Е. Морфология спор семейства *Osmundaceae*. — Бот. журн., 1966, т. 51, № 10, с. 1452—1460.

Бочанцева В. В. Биометрическая характеристика некоторых представителей сем. *Ranunculaceae* (*Anemone nemorosa* L., *Anemone ranunculoides* L., *Hepatica nobilis* Mill.) в Ленинградской области. — Вестн. Ленингр. ун-та, 1972, № 3, с. 37—45.

Василевич В. И. Статистические методы в геоботанике. Л., 1969. 129 с.

Васильев Б. Р., Звонцова Н. А., Савинов И. П., Шмидт В. М. Математический анализ роста листьев. — Бот. журн., 1973, т. 58, № 9, с. 1294—1301.

Васильев Б. Р., Савинов И. П., Шмидт В. М. К математическому анализу роста листьев. Использование функции Гомперца. — Бот. журн., 1975, т. 60, № 2, с. 209—216.

Вельдре С. Р. О корреляционной структуре внешних морфологических признаков ушастой круглоголовки *Phrynocephalus mystaceus* (Pallas, 1776). — В кн.: Применение математических методов в биологии. Л., 1964, вып. 3, с. 75—85.

Выханду Л. К. Об исследовании многопризнаковых биологических систем. — В кн.: Применение мат. методов в биологии. Л., 1964, вып. 3, с. 19—22.

Гамбарян П. Математический метод таксономии. — Изв. АН АрмССР, 1964, т. 17, № 12, с. 47—53.

Гамбарян П. П. Модификация математического метода таксономии. — Биол. журн. Армении, 1969, т. 22, № 10, с. 94—97.

- Гамбарян П. П. Сравнение типичных методов числовой таксономии. — Биол. журн. Армении, 1971, т. 24, № 4, с. 41—43.
- Глотов Н. В., Животовский Л. А., Хованов Н. В., Хромов-Борисов Н. Н. Биометрия. Л., 1982. 263 с.
- Гордягин А. Я. Наблюдения над изменчивостью *Anemone patens* L. — Тр. о-ва естествоисп. Казанск. ун-та, 1920, т. 59, вып. 5, с. 8—88.
- Грейг-Смит П. Количественная экология растений. М., 1967. 358 с.
- Гроссгейм А. А. Анализ флоры Кавказа. — Тр. Бот. ин-та Азерб. филиала АН СССР, 1936, т. 1. 242 с.
- Гурский А. В. Закономерности роста древесных растений. — Сообщ. Тадж. филиала АН СССР, 1948, вып. 4, с. 19.
- Дорофеюк А. А. Алгоритмы автоматической классификации. — Автоматика и телемеханика, 1971, № 12, с. 78—113.
- Дюльдин А. А. Булевы функции в таксономическом анализе Е. С. Смирнова. — Журн. общ. биол., 1973, т. 34, № 5, с. 745—751.
- Завадский К. М., Колчинский Э. И. Эволюция эволюции. Л., 1977. 236 с.
- Зайцев Г. Н. Методика биометрических расчетов. М., 1973. 256 с.
- Зотин А. И. Термодинамический подход к проблемам развития, роста и старения. М., 1974. 184 с.
- Иванов В. Б. Клеточные основы роста растений. М., 1974. 223 с.
- Кендэл М. Ранговые корреляции. М., 1975. 214 с.
- Кефели В. И., Девяткин В. М., Коренева В. М., Дубовая Л. П. О ритмическом характере ростового процесса. — Физиол. растений, 1964, т. 11, вып. 3, с. 496—505.
- Коновалов И. Н., Лерман Р. И., Михалева Е. Н., Шилова Н. В. Особенности изменения физиологических процессов растений в связи с их приспособлением к новым условиям существования. — В кн.: Экспериментальная ботаника. Л., 1960, с. 7—53 (Тр. Бот. ин-та им. В. Л. Комарова АН СССР, сер. 4, вып. 14).
- Кренке Н. П. Теория циклического старения и омоложения растений и практическое ее применение. М., 1940. 136 с.
- Лакин Г. Ф. Биометрия. М., 1980. 293 с.
- Лоули Д., Максвелл А. Факторный анализ как статистический метод. М., 1967. 274 с.
- Магомедмирзаев М. М. Об изменчивости сосны в условиях Центрального Дагестана. — Лесоведение, 1968, № 1, с. 65—71.
- Магомедмирзаев М. М. О значении корреляций в эволюционных преобразованиях типов побега у березы. — Докл. АН СССР, 1970а, т. 194, № 4, с. 967—969.
- Магомедмирзаев М. М. Анализ морфогенетической организации линейных признаков листового аппарата клена полевого. — В кн.: Сборник научных сообщений, вып. 2. Махачкала, 1970 б, с. 86—102.
- Магомедмирзаев М. М. Об одной закономерности индивидуальной изменчивости количественных признаков березы. — Лесоведение, 1972, № 5, с. 22—32.
- Мальшиев Л. И. Зависимость флористического богатства от внешних условий и исторических факторов. — Бот. журн., 1969, т. 54, № 8, с. 1137—1147.
- Мальшиев Л. И. Площадь выявления флоры в сравнительно-флористических исследованиях. — Бот. журн., 1972а, т. 57, № 2, с. 182—197.
- Мальшиев Л. И. Флористические спектры Советского Союза. — В кн.: История флоры и растительности Евразии. Л., 1972б, с. 17—40.
- Мальшиев Л. И. Количественный анализ флоры: пространственное разнообразие, уровень видового богатства и репрезентативность участков обследования. — Бот. журн., 1975, т. 60, № 11, с. 1537—1550.
- Мальшиев Л. И. Количественная характеристика флоры Путорана. — В кн.: Флора Путорана. Новосибирск, 1976, с. 163—186.

Миняев Н. А. История развития флоры северо-запада европейской части РСФСР с конца плейстоцена. Доклад о работах, представленных к защите на соиск. учен. степени д-ра биол. наук. Л., 1966. 38 с.

Миняев Н. А., Шмидт В. М. Краткая история исследования флоры Псковской области. — В кн.: Конспект флоры Псковской области. Л., 1970, с. 5—10.

Митропольский А. К. Техника статистических вычислений. М., 1971. 576 с.

Окунь Я. Факторный анализ. М., 1974. 200 с.

Плохинский Н. А. Биометрия. М., 1970. 367 с.

Расницын С. П. Применение таксономического анализа для сравнения биотопов по их фауне и населению. — Журн. общ. биол., 1965, т. 26, № 3, с. 335—340.

Ребристая О. В. Флора востока Большеземельской тундры. Л., 1977. 334 с.

Ребристая О. В., Шмидт В. М. Сравнение систематической структуры флор методом ранговой корреляции. — Бот. журн., 1972, т. 57, № 11, с. 1353—1364.

Рокицкий П. Ф. Биологическая статистика. Минск, 1973. 320 с.

Рухадзе Л. П. Корреляционная структура признаков семян некоторых ископаемых и современных европейских видов винограда. — Сообщ. АН ГССР, 1968, т. 51, № 3, с. 711—715.

Рыгова Н. Г. Формирование свободно размещенных растений луговых злаков. — Сельскохозяйственная биология, 1971, т. 6, № 4, с. 532—538.

Сабинин Д. А. Минеральное питание растений. Киев, 1940. 307 с.

Сабинин Д. А. О ритмичности строения и роста растений. — Бот. журн., 1957, т. 42, № 7, с. 991—1010.

Сабинин Д. А. Физиология развития растений. М., 1963. 196 с.

Савинов И. П., Васильев Б. Р., Шмидт В. М. Об одном классе кривых роста растений. — Журн. общ. биол., 1977, т. 38, № 3, с. 432—438.

Серебрякова Т. И. Некоторые закономерности формирования почек у луговых злаков. — Бюл. МОИП, отд. биол., 1961, т. 66, № 4, с. 45—52.

Серебрякова Т. И. О ритмичности в структуре и онтогенезе побегов у злаков. — Бюл. МОИП, отд. биол., 1969, т. 74, № 4, с. 60—71.

Серебрякова Т. И. Морфогенез побегов и эволюция жизненных форм злаков М., 1971. 359 с.

Синнот Э. Морфогенез растений. М., 1963. 603 с.

Синская Е. Н. Проблема популяций у высших растений. Л., 1963. 124 с.

Смирнов Е. С. О строении систематических категорий. — Рус. Зоол. журн., 1923, т. 3, № 3/4, с. 358—391.

Смирнов Е. С. Анализ распределения и соотношения признаков в систематических категориях. — Докл. Рос. АН, 1924, с. 81—83.

Смирнов Е. С. Таксономический анализ рода. — Журн. общ. биол., 1960, т. 21, № 2, с. 89—103.

Смирнов Е. С. Таксономический анализ. М., 1969. 187 с.

Смирнов Е. С. О кодировании признаков для таксономического анализа. — Журн. общ. биол., 1971, т. 32, № 2, с. 224—228.

Снедекор Д. У. Статистические методы в применении к исследованиям в сельском хозяйстве и биологии. М., 1961. 503 с.

Соколова Н. П. Эволюционные взаимоотношения жизненных форм в роде *Rubus* L. — Малина, Ежевика, сем. Rosaceae: Автореф. докт. дис. М., 1974. 30 с.

Сокэл Р. Р. Нумерическая таксономия: методы и современное развитие. — Журн. общ. биол., 1968, т. 29, № 3, с. 297—315.

Старобогатов Я. И. Фауна моллюсков и географическое районирование континентальных водоемов земного шара. Л., 1970. 372 с.

Стрелер Б. Время, клетки и старение. М., 1964. 253 с.

Сукачев В. Н. Биометрические исследования над *Chrysanthemum leucanthemum* L. и *Chr. ircutianum* (DC.) Turcz. — Изв. Рос. АН, 1918, т. 12, ч. 1, № 10, с. 939—970.

Сукачев В. Н. О *Caltha palustris* L. var *Stebutiana* n. в связи с вопросом об изменчивости ее и типичной формы. — Журн. русс. бот. об-ва, 1919, т. 4, с. 111—132.

Табака Л. В., Қлявиня Г. Б., Плотниекс М. Р. Некоторые методические вопросы изучения видового состава флоры Западной Латвии. — В кн.: Флора и растительность Латвийской ССР. Рига, 1977, с. 86—120.

Тамарин П. В. Анализ математических методов систематики. — Журн. общ. биол., 1971а, т. 32, № 3, с. 277—286.

Тамарин П. В. О противоположном характере результатов, получаемых с помощью весового и адансоновского принципов в таксономии. — Журн. общ. биол., 1971б, т. 32, № 4, с. 460—466.

Тамарин П. В. Оптимизация вычислительных работ в таксономии Е. С. Смирнова с помощью весовых таксономических таблиц. — Журн. общ. биол., 1972, т. 33, № 5, с. 592—609.

Тамарин П. В. Доказательство таксономической ценности редких признаков с помощью свойств коэффициента корреляции. — Журн. общ. биол., 1973, т. 34, № 2, с. 264—274.

Тамарин П. В. Унификация и кодирование качественных и количественных признаков в таксономии. — В кн.: Методы современной биометрии. М., 1978, с. 130—156.

Тамарин П. В., Шмидт В. М. Сравнительный анализ некоторых коэффициентов сходства. — В кн.: Успехи биометрии. Л., 1975, с. 45—54. (Тр. Ленингр. о-ва естествоисп., т. 72, вып. 5).

Тахтаджян А. Л. О принципах, методах и символах филогенетических построений в ботанике. — Бюл. МОИП, отд. биол., 1947, т. 52, № 5, с. 95—120.

Тахтаджян А. Л. Система и филогения цветковых растений. М.; Л., 1966. 611 с.

Тахтаджян А. Л. Флористические области Земли. Л., 1978. 247 с.

Терентьев П. В. К вопросу о реальной значимости пирсоновских типов кривых распределения для биологических объектов. — Докл. АН СССР, 1935, т. 1, № 4.

Терентьев П. В. Метод корреляционных плеяд. — Вестн. Ленингр. ун-та, 1959, № 9, с. 137—141.

Терентьев П. В. Дальнейшее развитие метода корреляционных плеяд. — В кн.: Примененные мат. методов в биологии. Л., 1960, с. 27—36.

Терентьев П. В., Ростова Н. С. Практикум по биометрии. Л., 1977. 152 с.

Толмачев А. И. К методике сравнительно-флористических исследований. I. Понятие о флоре в сравнительной флористике. — Журн. рус. бот. о-ва, 1931, т. 16, № 1, с. 111—124.

Толмачев А. И. Флора центральной части Восточного Таймыра. — Тр. Полярн. комисии АН СССР, вып. 8, 1932. 126 с.; вып. 13, 1932. 72 с.; вып. 25, 1935. 80 с.

Толмачев А. И. О количественной характеристике флор и флористических областей. М.; Л., 1941. 40 с. (Тр. Северн. базы АН СССР, вып. 8).

Толмачев А. И. Богатство флор как объект сравнительного изучения. — Вестн. Ленингр. ун-та, 1970, № 9, с. 71—83.

Толмачев А. И. Введение в географию растений. Л., 1974. 244 с.

Урбах В. Ю. Биометрические методы. М., 1964. 415 с.

Фишер Р. А. Статистические методы для исследователей. М., 1958. 268 с.

Харман Г. Современный факторный анализ. М., 1972. 486 с.

Чупов В. С. О географической изменчивости числа лепестков у *Anemone nemorosa* L. и *A. ranunculoides* L. — Бот. журн., 1972, т. 57, № 8, с. 950—954.

Шеляг-Сосонко Ю. Р. О конкретной флоре и методе конкретных флор. — Бот. журн., 1980, т. 65, № 6, с. 761—774.

Шилова Н. В. Темп и ритм формирования элементов вегетативной почки у некоторых *Andromedeidae* (Drude) E. Busch. — Вестн. Ленингр. ун-та, 1962, № 21, с. 23—30.

- Шилова Н. В.* Приспособление вечнозеленых кустарников к условиям Севра. Л., 1967. 63 с.
- Шмальгаузен И. И.* Проблемы дарвинизма. Л., 1969. 493 с.
- Шмидт В. М.* К проблеме биологической значимости полимодальных кривых некоторых признаков растений. — Вестн. Ленингр. ун-та, 1961, № 9, с. 36—45.
- Шмидт В. М.* О методе таксономического анализа Е. С. Смирнова и некоторых возможностях его применения в ботанике. — Бот. журн., 1962, т. 47, № 11, с. 1648—1654.
- Шмидт В. М.* Опыт анализа дивергенции корреляционных структур систематических категорий. — В кн.: Применение мат. методов в биологии. Л., 1964, вып. 3, с. 61—69.
- Шмидт В. М.* О квантированности роста и органогенеза у растений. — Вестн. Ленингр. ун-та, 1968, № 3, с. 99—111.
- Шмидт В. М.* О двух методах таксономического анализа. — Бот. журн., 1970, т. 55, № 3, с. 386—396.
- Шмидт В. М.* О площади конкретной флоры. — Вестн. Ленингр. ун-та, 1972, № 3, с. 57—66.
- Шмидт В. М.* Количественные показатели в сравнительной флористике. — Бот. журн., 1974, т. 59, № 7, с. 929—940.
- Шмидт В. М.* Биометрические исследования в области морфогенеза растений, таксономии и сравнительной флористики: Автореф. докт. дис. Л., 1975. 52 с.
- Шмидт В. М.* О двух направлениях развития метода конкретных флор. — Бот. журн., 1976, т. 61, № 12, с. 1658—1669.
- Шмидт В. М.* О корреляциях. I. Сущность, онтогенетический и филогенетический аспекты явления биологических корреляций. — Вестн. Ленингр. ун-та, 1979, № 3, с. 77—85.
- Шмидт В. М.* Статистические методы в сравнительной флористике. Л., 1980. 176 с.
- Шмидт В. М.* О корреляциях. II. Вопросы классификации биологических корреляций. — Вестн. Ленингр. ун-та, 1981а, № 15, с. 62—68.
- Шмидт В. М.* Приближенное прогнозирование флористических показателей в зависимости от географической широты. — Журн. общ. биол., 1981б, т. 42, № 3 с. 430—439.
- Шмидт В. М.* О коэффициентах корреляции, используемых для сравнения систематической структуры флор. — Вестн. Ленингр. ун-та, 1981в, № 3, с. 57—67.
- Шмидт В. М., Баранова Е. В.* Применение метода таксономического анализа для сравнения флор по их историко-географическим связям. — Журн. общ. биол., 1975, т. 36, № 4, с. 555—562.
- Шмидт В. М., Васильев Б. Р., Колодяжный С. Ф.* К обобщению опыта математического анализа роста и органогенеза у растений. — Журн. общ. биол., 1978, т. 39, № 6, с. 878—885.
- Эсау К.* Анатомия растений. М., 1969. 564 с.
- Юл Д. Э., Кендэл М. Д.* Теория статистики. М., 1960. 779 с.
- Юрцев Б. А.* Дискуссия на тему «Метод конкретных флор в сравнительной флористике». — Бот. журн., 1974, т. 59, № 9, с. 1399—1407.
- Юрцев Б. А.* Некоторые тенденции развития метода конкретных флор. — Бот. журн., 1975, т. 60, № 1, с. 69—83.
- Юрцев Б. А.* Флора как природная система. — Бюл. МОИП, отд. биол., 1982, вып. 4, с. 3—22.
- Bertalanffy L.* Untersuchungen über die Gesetzmäßigkeit des Wachstums. 1. Allgemeine Grundlagen der Theorie mathematische und physiologische Gesetzmäßigkeiten des Wachstums bei Wassertieren. — Roux' Arch. Entwicklungsmech. Organ., 1934, Bd 131, N 4, S. 613—653.
- Bertalanffy L.* A quantitative theory of organic growth (inquires on growth laws II). — Human. Biol., 1938, vol. 10, N 4, p. 184—213.
- Blackman V. H.* The compound interest law and plant growth. — Ann. Bot., 1919, N 33, p. 353—360.

- Cailleux A.* Biogéographie mondiale. Paris, 1953. 128 p.
- De Candolle A.* Géographie botanique raisonnée. Paris; Genève. 1855. 1365 p.
- Huxley J. S.* Problems of relative growth. New York, 1932. 276 p.
- Lebrun J.* Sur la richesse de la flore de divers territoires africaines.— Acad. Roy. Sci. d'outre-mer, Bull. séances, Nouv. sér., 1960, vol. 6, N 4, p. 669—690.
- Losch H.* Über die Variation der Anzahl der Sepalen und der Hüllblätter bei *Anemone nemorosa* L. und über den Verlauf der Variation während einer Blütenperiode nebst einigen teratologischen Beobachtungen.— Ber. Deutsch. Bot. Ges., 1916, N 34, S. 396—411.
- Ludwig F.* Das Gesetz der Variabilität der Zungenblüten von *Chrysanthemum leucanthemum*.— Mitteil. Thüring. Bot. Ver., 1897, Bd 7, N 10, S. 20—23.
- Ludwig F.* Ueber Variationskurven.— Bot. Centralbl., 1898, Bd 75, N 4—5, S. 97—107, 178—183.
- Medawar P. B.* Size, chape and age.— In: Essays on growth and form. Oxford, 1945, p. 157—187.
- Monod T.* Sur un cas exceptionnel de richesse aréale: les sables de Mauritanie orientale.— Rec. Trav. Bot. Géol. Zool. Univ. Montpellier, 1955, Sér. bot., N 7, p. 63—67.
- Plantefol L.* Fondements d'une théorie phyllotaxique nouvelle.— Ann. Sci. Nat. Bot., 1946, vol. 11, N 7, p. 153—229.
- Plantefol L.* La théorie des hélices foliaires multiples. Paris, 1948, 154 p.
- Preston F. W.* The canonical distribution of commonness and rarity.— Ecology, 1962, N 43, p. 185—215 a. 410—431.
- Pütter A.* Studien über physiologische Ähnlichkeit.— Pflüger's Arch. Ges. Physiol., 1920, N 180, S. 298—340.
- Sneath P. H., Sokal R. R.* Numerical taxonomy. San Francisco, 1973. 573 p.
- Sokal R. R., Sneath P. H.* Principles of numerical taxonomy. San Francisco, 1963. 359 p.
- Thompson D'Arcy.* On growth and form. Cambridge; New York, 1942. 1116 p.
- Vogler P.* Variation der Anzahl der Strahlblüten bei einigen Kompositen.— Beihefte Bot. Centralbl., 1910, Bd 25, N 1, S. 387—396.
- Weldon W.* Change in organic correlation of *Ficaria ranunculoides* during the flowering season.— Biometrika, 1901, N 1, p. 125—128.

ОГЛАВЛЕНИЕ

Предисловие	3
ОБЩАЯ ЧАСТЬ	
Глава 1. Изучение изменчивости признаков	6
1.1. Сущность биометрического метода. Генеральная совокупность и выборка	6
1.2. Правила составления выборок	9
1.3. Ошибки выборочного исследования	13
1.4. Вариационный ряд и его обработка	14
1.4.1. Запись наблюдений (с. 14).—1.4.2. Типы варьирования признаков (с. 15).—1.4.3. Группировка данных (с. 16).—1.4.4. Графическое изображение вариационного ряда (с. 19).—1.4.5. Основные выборочные параметры (с. 21).—1.4.6. Косвенные способы расчета основных выборочных параметров (с. 24).—1.4.7. Ошибки репрезентативности выборочных параметров (с. 27).—1.4.8. Проверка принадлежности варианты к данной совокупности (с. 28).	
1.5. Сравнение одноименных признаков в разных выборках	30
1.5.1. Параметрические критерии различия (с. 30).—1.5.2. Непараметрические критерии различия (с. 32).	
Глава 2. Теоретические и эмпирические распределения	37
2.1. Биномиальное распределение	38
2.2. Распределение Пуассона	41
2.3. Нормальное распределение	43
2.4. Эмпирические распределения. Асимметрия и эксцесс	48
2.5. Метод моментов и типы кривых Пирсона	49
Глава 3. Изучение влияния факторов среды на изменчивость признаков (дисперсионный анализ)	53
3.1. Логическая схема дисперсионного анализа	53
3.2. Однофакторный дисперсионный анализ	56
3.3. Двухфакторный дисперсионный анализ	62
Глава 4. Изучение взаимной связи признаков (корреляционный анализ)	66
4.1. Понятие о корреляционной связи	66
4.2. Линейная корреляция	69
4.2.1. Коэффициент линейной корреляции (с. 69).—4.2.2. Частная и множественная корреляция (с. 74).	
4.3. Нелинейная корреляция	76
4.3.1. Корреляционное отношение (с. 76).—4.3.2. Критерии нелинейности связи (с. 80).	
4.4. Непараметрические показатели связи, основанные на методе χ^2	81
Глава 5. Изучение зависимости изменения одного признака от изменения другого (регрессионный анализ)	87
5.1. Понятие о регрессии	87
5.2. Эмпирические линии регрессии	88
5.3. Линейная регрессия	89
5.3.1. Уравнение линейной регрессии и метод наименьших квадратов. Построение теоретической линии регрессии (с. 90).—5.3.2. Коэффициенты регрессии (с. 93).	

5.4. Нелинейная регрессия	94
5.4.1. Параболическое интерполирование (с. 95). — 5.4.2. Другие интерполяционные формулы (с. 99).	
5.5. Множественная регрессия	103

СПЕЦИАЛЬНАЯ ЧАСТЬ

Глава 6. Математические методы изучения роста и органогенеза у растений	106
6.1. Квантированность роста и органогенеза	106
6.1.1. Многовершинные кривые и их биологическая интерпретация (с. 106). — 6.1.2. Многовершинные кривые числа органов и ряд Фибоначчи (с. 108). — 6.1.3. Многовершинные кривые размера органов (с. 114). — 6.1.4. «Принцип кратных отношений» А. В. Гурского и числовые соотношения органов растений (с. 118).	
6.2. Математический анализ роста	129
6.2.1. Описание роста логистической функцией и функцией Гомпертца (с. 132). — 6.2.2. Расчет основных параметров роста (с. 135). — 6.2.3. Обобщенный класс функций роста (с. 141).	
6.3. Соотносительный рост органов и его описание аллометрической функцией	144
Глава 7. Математические методы в систематике	148
7.1. Простейшие приемы сравнения таксонов	148
7.1.1. Полигональные графики (с. 148). — 7.1.2. Коэффициент дивергенции признаков (с. 149). — 7.1.3. Понятие о таксономическом коэффициенте (с. 153).	
7.2. Таксономический анализ	154
7.2.1. О двух направлениях таксономического анализа (с. 155). — 7.2.2. Метод таксономического анализа Е. С. Смирнова (с. 160).	
7.3. Анализ корреляционной структуры таксонов	174
7.3.1. Коэффициент дивергенции корреляций (с. 175). — 7.3.2. Метод корреляционных плеяд П. В. Терентьева (с. 179). — 7.3.3. Основные принципы факторного анализа (с. 188).	
Глава 8. Математические методы в сравнительной флористике	194
8.1. Обеспечение биолого-статистической сопоставимости флор	194
8.2. Метод конкретных флор А. И. Толмачева	197
8.2.1. Полевая методика изучения конкретных флор (с. 201). — 8.2.2. О площади конкретной флоры (с. 203).	
8.3. Количественные показатели (параметры) флор, их свойства и способы сравнения	206
8.3.1. Показатели флористического богатства и систематического разнообразия (с. 206). — 8.3.2. Показатели сходства систематической структуры флор (с. 213). — 8.3.3. Информационные индексы сложности систематической структуры флор (с. 224). — 8.3.4. Коэффициенты сходства систематического состава флор (флористических списков) (с. 231).	
8.4. Географическая изменчивость флористических показателей и возможность их прогнозирования	237
8.5. Сравнение флор по их историко-географическим связям с помощью метода таксономического анализа Е. С. Смирнова	252
8.6. Флористическое районирование на статистической основе	259
Приложения	266
Указатель литературы	281