Учреждение образования
«Гомельский государственный университет
имени Франциска Скорины»

А. С. СОКОЛОВ

МЕТОДЫ ОБРАБОТКИ ЭКОЛОГИЧЕСКИХ ДАННЫХ

Практическое пособие

для магистрантов специальности 1-33 80 01 «Экология»

Гомель
ГГУ им. Ф. Скорины
2022

УДК 502/.504:311.2(076)
ББК 20.1с9я73
С59

Рецензенты:

кандидат географических наук Е. Н. Карчевская,
кандидат географических наук А. И. Павловский

Рекомендовано к изданию научно-методическим советом учреждения образования «Гомельский государственный университет имени Франциска Скорины»

Соколов, А. С.

С59 Методы обработки экологических данных : практическое пособие / А. С. Соколов ; Гомельский гос. ун-т им. Ф. Скорины. – Гомель : ГГУ им. Ф. Скорины, 2022. – 46 с.
ISBN 978-985-577-856-2

Практическое пособие разработано в соответствии с учебной программой дисциплины «Методы обработки экологических данных» и посвящено рассмотрению методов расчета основных показателей вариационных рядов, доказательства их принадлежности к разным генеральным совокупностям, количественной и качественной корреляции между исследуемыми признаками, а также других методов, применяемых для обработки данных полевых экологических исследований.

Адресовано магистрантам специальности 1-33 80 01 «Экология».

УДК 502/.504:311.2(076)
ББК 20.1с9я73

ISBN 978-985-577-856-2

имени Франциска Скорины», 2022

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Математическая, в том числе статистическая обработка результатов полевых или иных исследований является важнейшим компонентом научно-исследовательской работы. Она позволяет существенно увеличить объём информации, извлекаемой из полевых и экспериментальных данных. Только используя статистический аппарат, мы можем делать выводы о наличии объективных закономерностей, связей между различными характеристиками объектов, степени их влиянии друг на друга; формулировать и доказывать гипотезы.

Важность применения статистических методов исследования неоднократно подчёркивается авторами соответствующих учебных пособий: «Научный работник должен безупречно владеть специальными методиками. Но только этого для научной работы недостаточно. Мало получить научные факты, их необходимо проанализировать. Первый этап – анализ статистический. Именно статистический анализ, а не субъективное ощущение».

Предлагаемое пособие нацелено на изучение основ методов математического анализа результатов экологических исследований, которые станут фундаментом для рассмотрения в дальнейшем более сложных и специализированных методов. В нём освещаются такие вопросы, как понятие о вариационных рядах и их основных характеристиках, включающих среднее значение признака и количественную меру варьирования значения признака; методы сравнения вариационных рядов и определения принадлежности их к одной или различным генеральным совокупностям; методы выявления связей между различными качественными и количественными показателями объектов природной среды; методы фитоиндикации абиотических компонентов экосистем; методы орди-нации и градиентного анализа; методы определения разнообразия экосистем.

Данное практическое пособие адресовано магистрантам специальности 1–33 80 01 «Экология».

ТЕМА 1. ПОКАЗАТЕЛИ ВАРИАЦИОННОГО РЯДА

Всякое множество идентифицируемых объектов, отличающихся друг от друга незначительно по конкретному признаку, но сохраняющих сходство по некоторым существенным характеристикам, называется *совокупностью*. Совокупностью можно называть стадо животных, популяцию растений, пробы воды из водоёма, группу анкетируемых, количество родившихся за определённый период и т. д.

В природе, как правило, невозможно исследовать все члены совокупности (все деревья в лесу, всю воду в водоёме, всех мышей в популяции, все растения на поле пшеницы и т. д.) в силу их огромного количества, поэтому из всей совокупности (которая называется *генеральной совокупностью*) отбирают некоторую часть, называемую *выборочной совокупностью*, или *выборкой*, и изучают её характеристики, чтобы затем по этим характеристикам делать вывод обо всей генеральной совокупности. Количество членов выборки (число наблюдений) называют *объёмом выборки* и обозначают n. Изучаемый показатель (размер, концентрация вещества, вариант ответа, плотность и т. д.) называется *признаком*, величина изучаемого показателя называется *значением* признака и обозначается x.

Важным свойством выборки является её *случайный характер*. Это означает, что каждый член генеральной совокупности равновероятно может попасть в выборку для проведения эксперимента. То есть вероятность оказаться в выборке одинакова для всех членов генеральной совокупности. Случайность выборки обеспечивает исключение субъективизма, неосознанное стремление подогнать данные под заранее выдвинутую гипотезу. Если выборка не является случайной, то статистический аппарат для её анализа применять бессмысленно.

Случайный характер выборки обеспечивает её *репрезентативность* – важнейшее свойство, способность характеризовать всю генеральную совокупность. Если выборка не является репрезентативной, то исследователь может сделать ошибочные выводы обо всех объектах исследования (всей генеральной совокупности).

Любая выборка характеризуется *распределением признака*. К примеру, изучая связь типа почв с высотой мятлика лугового, из каждой популяции, растущей на определённом типе почв, отбирается выборка и измеряется значение признака – высота всех растений, в неё входящих. Расположив все измеренные значения в порядке возрастания или убывания (то есть, создав ранжированный ряд, или *вариационный ряд*), можно увидеть сколько раз то или иное значение высоты появилось в выборке. Это показатель называется *частотой* и обозначается *f*. Построив график распределения, можно графически показать, сколько раз то или иное значение появилось в ходе исследования (рисунок 1).

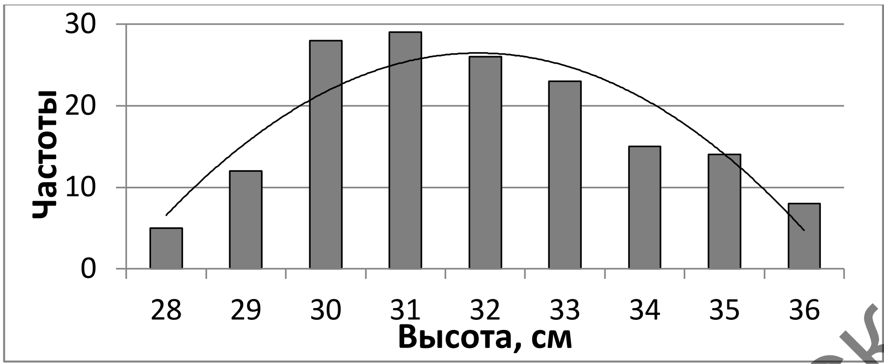


Рисунок 1 – Пример распределения изучаемого признака

Чаще же используют относительные частоты встречаемости, измеряемые в процентах или долях от единицы:

$$f = \frac{k}{n}(\times 100),$$

где k – количество членов выборки с заданным конкретным значением;

n – общий объём выборки.

Сумма относительных частот должна равняться 100 % или 1.

Если кривая, отражающая распределение значений признака по частотам, имеет приблизительно колоколообразный вид, как на рисунке 1, то такое распределение называется *нормальным*.

Нормальное распределение подразумевает, что большая часть значений признака находится в районе так называемого *среднего значения*. При нормальном распределении наиболее часто в выборке встречаются значения, близкие по величине к среднему по выборке и располагающиеся симметрично ему (значений больше среднего и значений меньше среднего приблизительно одинаковое количество). Или если выражать в процентном соотношении (используя относительные частоты встречаемости), то можно говорить, что наибольший процент значений признака находится в районе среднего значения, тогда как всего несколько процентов – по краям кривой.

Именно в зависимости от типа распределения выбираются методы статистического анализа. Если распределение является нормальным, то применяют методы так называемой *параметрической статистики*.

К главным статистическим характеристикам вариационного ряда можно отнести среднее значение признака и количественную меру варьирования значения признака.

Существуют различные виды средних значений вариационного ряда, объединяемые общим понятием *«степенные средние»*. К ним относятся среднее арифметическое, среднее геометрическое, среднее гармоническое, среднее квадратичное.

Наиболее распространённый показатель – *среднее арифметическое*, которое рассчитывается как частное от деления суммы числовых значений признака всех членов вариационного ряда на количество членов вариационного ряда. Если значения признака всех членов заменить на данный показатель, то их сумма не изменится.

Формулой *среднего геометрического* следует пользоваться в случае, если значения осредняемого признака далеко отстают друг от друга или заданы цепными коэффициентами, когда каждый последующий коэффициент зависит от значения предыдущего, например, темп роста, коэффициент прироста. Он вычисляется как корень n-ой степени из произведения числовых значений признака всех членов вариационного ряда, где n – количество членов вариационного ряда. Если значения признака всех членов заменить на данный показатель, то их произведение не изменится.

Среднее гармоническое вычисляют в случаях, когда числа вариационного ряда представлены обратными значениями. Например, это бывает, когда речь идёт о средних значениях различных скоростей за различные промежутки времени, о средней производительности труда у людей или механизмов с различными её значениями и о других случаях вычисления среднего значения относительных динамических показателей у членов ряда с различными их значениями. Среднее гармоническое – это величина, обратная среднему арифметическому обратных величин:

$$\bar{x}_{\text{гарм}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}},$$

где n – количество членов вариационного ряда.

Средняя квадратическая применяется для вычисления среднего значения таких показателей, как диаметры колес, труб, средние стороны квадратов и т. п. Этот показатель является квадратным корнем из частного от деления суммы квадратов отдельных значений признака на их число:

$$\bar{x}_{\text{кв}} = \sqrt{\frac{x_1^2 + x_2^2 + ... + x_n^2}{n}}.$$

Если для чисел x_1, x_2, x_3, ..., x_n существуют все определенные выше средние, то для них всегда верно неравенство:

$$\bar{x}_{\text{гарм}} < \bar{x}_{\text{геом}} < \bar{x}_{\text{ариф}} < \bar{x}_{\text{кв}}.$$

Кроме степенных средних широко используют такие показатели, как мода и медиана.

Мода – наиболее часто встречающееся значение. Преимуществом моды перед средними значениями является то, что она не зависит от крайних значений признака, которые подвержены большим колебанием из-за случайных факторов.

Медиана – это центральное значение в ранжированном ряду данных. То есть, это такое значение, больше которого 50 % значений и не больше которого 50 % значений признака в выборке.

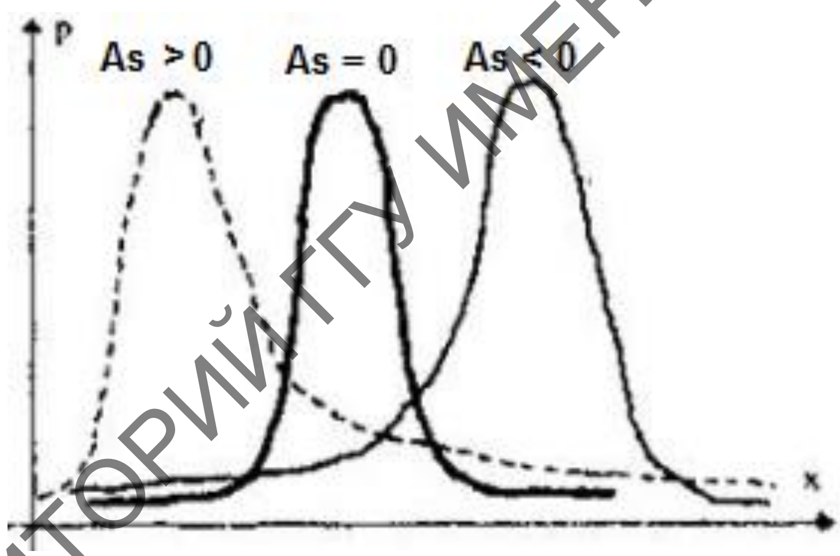
Помимо медианы, делящей вариационный ряд пополам, существуют и другие числа, которые делят его на различные части. Например, *квартили* – числа, делящие вариационный ряд на 4 равные части: первый (верхний) квартиль ($x_{0,75}$) – число, выше которого 25 % и не выше 75 % членов вариационного ряда, второй квартиль ($x_{0,5}$) – это медиана, третий (нижний) квартиль ($x_{0,25}$) – число, выше которого 75 % и не выше 25 % членов вариационного ряда. *Интерквартильный размах* – это показатель разности между первым и третьим квартилями ($x_{0,75} – x_{0,25}$). *Децили* – числа, делящие вариационный ряд на 10 равных отрезков и т. д. Можно вычислить любой *процентиль*, то есть число, выше которого будет определённый процент значений вариационного ряда.

При расчёте средних значений относительных показателей (коэффициентов) для членов ряда с различными значениями характеристики, на которую делится абсолютный показатель (например, среднее значение лесистости для районов с различными площадями или среднюю долю женщин для населённых пунктов с различной численностью населения), рассчитываются показатели, называемые *средние взвешенные*. В этом случае, например, при расчёте *среднего арифметического взвешенного*, в числителе каждое значение показателя, для которого рассчитывается среднее, умножается на соответствующее ему значения «веса». В знаменателе количество членов вариационного ряда заменяется на сумму всех «весов». Аналогичным образом вычисляются взвешенные значения и для других видов средних.

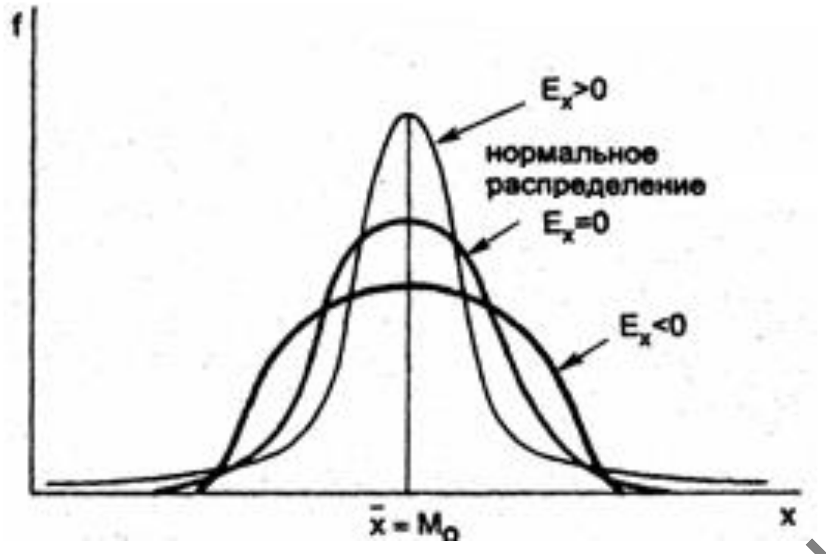
Ещё два показателя, которые характеризуют особенности распределения случайной величины:

– *коэффициент асимметрии* (A_s) – показатель, характеризующий симметричность распределения признака относительно среднего арифметического. Если распределение полностью симметрично, то коэффициент асимметрии равен 0. Если он положителен, то правый хвост на графике распределения длиннее левого, то есть наблюдается смещение значений в сторону больше среднего арифметического. Если он отрицателен, то, наоборот, левый хвост распределения длиннее правого, и наблюдается смещение значений в сторону меньше среднего арифметического (рисунок 2). Величина коэффициента показывает степень смещения. В рассматриваемом на рисунке 1 примере A_s = 0,23;

– *коэффициент эксцесса* (E_x) (коэффициент островершинности) – мера остроты пика распределения случайной величины. Он положителен, если пик распределения около среднего арифметического острый, и отрицателен, если пик очень гладкий (рисунок 2). В примере E_x = –0,72.



возможные значения асимметрии



возможные значения эксцесса

Рисунок 2 – Возможные значения асимметрии и эксцесса

К мерам, показывающим степень варьирования значения признака можно отнести:

– *среднеквадратичное отклонение* (σ). Даёт представление о том, как сильно могут отклоняться от своего центра (среднего арифметического) значения исследуемой случайной величины. Чем больше её величина, тем больше степень её «размазанности» по всему диапазону при одинаковых значениях средней. На рисунке 3 показаны графики распределения значений, характеризующиеся одинаковыми значениями средней и различными значениями среднеквадратичного отклонения. В рассматриваемом нами примере с высотой мятлика σ = 2,04;

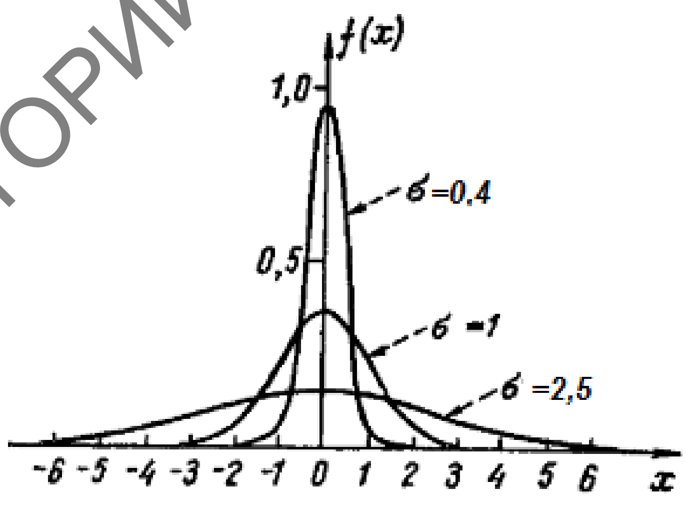


Рисунок 3 – График плотностей (нормальных законов распределения) с различными значениями среднеквадратичного отклонения и одинаковым (нулевым) значением среднего арифметического

– *коэффициент вариации* (V). Среднеквадратичное отклонение дает абсолютную оценку меры разброса. Поэтому чтобы понять, насколько разброс велик относительно самих значений (т. е. независимо от их масштаба), требуется относительный показатель. Такой показатель называется коэффициентом вариации и равен отношению σ к средней арифметической, умноженному на 100 %. В нашем примере V = 6,4 %;

– *ошибку средней* (m), которая рассчитывается как отношение среднеквадратичного отклонения и квадратного корня величины объёма выборки. На графиках ошибка средней может быть показана с помощью планок погрешностей ⊥ ⊤. Математическая статистика даёт доказательство того, что с вероятностью 95 % среднее значение всей генеральной совокупности находится в пределах от $x - 2m$ до $x + 2m$. В нашем примере m = 0,26;

– *ошибку среднеквадратичного отклонения* (m_σ), которую находят как частное от деления среднеквадратичного отклонения на квадратный корень величины объёма выборки, умноженной на 2. Так же, как и в предыдущем случае, с вероятностью 95 % истинное значение среднеквадратичного отклонения лежит в диапазоне $\sigma \pm 2m_\sigma$. Ошибка коэффициента вариации рассчитывается аналогично, путём замены им среднеквадратичного отклонения в числителе.

После того как исследователь вычислил показатели отдельного вариационного ряда, описанные в предыдущем подразделе, он переходит к следующим шагам и сравнивает эти показатели для различных вариационных рядов. Эти сравнения лежат в основе формулировки результатов значительной доли научно-исследовательских работ.

К примеру, исследователь имеет рабочую гипотезу, что высота мятлика лугового не будет одинакова на всех лугах, а будет зависеть от типа почв: на каких-то типах почв он будет более высоким, на каких-то – более низким. Для проверки своей гипотезы он, изучая зависимость высоты мятлика от типа почв, изучил два луга, расположенных на различных почвах. На каждом из этих лугов он отобрал случайным образом по нескольку десятков растений, то есть создал две выборки, составил два вариационного ряда, у каждого из которых определил среднее значение.

Казалось бы, достаточно сравнить два средних арифметических, выявить, в какой выборке этот показатель больше, и можно делать вывод о том, что почвы на данном лугу более благоприятны для произрастания мятлика, чем на втором лугу, где среднее арифметические было меньше. Именно так поступают в тех исследованиях и написанных на их основе научных работах, где статистика не используется.

Почему такой результат невозможно считать достоверным, а само исследование в таком случае не имеет ни научной, на практической ценности? Потому что при использовании такой схемы исследовательского процесса отсутствует механизм отделения случайностей от закономерностей. На высоту мятлика в различных лугах оказывает влияние не только тип почв, но и огромное количество случайных факторов. Даже если взять две случайные выборки с одного и того же луга и определить их средние значения, они не будут абсолютно идентичными. Всегда какое-то значение будет большим, а какое-то меньшим в силу чисто случайных факторов. И такое (влияние случайных факторов) происходит абсолютно во всех исследованиях, связанных с изучением реальной природной среды. Даже если посадить в одинаковые цветочные горшки два генетически идентичных цветка и поместить их в одних и те же условиях влажности, освещённости и других факторов, всё равно их высоты через некоторое время не будут абсолютно идентичны, так как даже в описанном случае влияние случайных факторов будет проявляться, хотя и не в такой степени, как в природных экосистемах.

Поэтому, вернувшись к примеру с мятликом и типами почв, после определения основных характеристик двух выборочных совокупностей, взятых с двух лугов с разными типами почв, необходимо проверить, существует ли между ними статистически доказанное различие, или обе этих выборки принадлежат к одной и той же генеральной совокупности (то есть различия в типах почв этих двух лугов никак не влияют на высоту мятлика), а различия их средних значений вызваны случайными факторами?

Для такой проверки существует показатель, который называется *t-критерий Стьюдента*. Однако прежде – ещё несколько терминов, применяемых при статистических операциях сравнения и оценок взаимосвязей.

Выборки могу быть *независимыми* и *связанными*. Пример *независимых* выборок наблюдается в случаях, когда на каждом луче случайно отбирали растения, при этом одна выборка никак не влияла на другую. *Связанные* выборки рассматриваются в тех случаях, где в качестве второй выступает та же первая выборка, но после проведения определённых манипуляций над ней. Например, имеется случайная выборка пациентов с некоторым заболеванием, которое характеризуется каким-то количественным показателем – признаком. Это первая выборка. Затем этим людям дают определённое лекарство и смотрят, как меняется у них изучаемый признак. По этим данным строят второй вариационный ряд (выступающий в роли второй выборки), определяют его показатели и сравнивают с первым. В данном случае также необходимо определить, имеется ли статистическая достоверность того, что изменение изучаемого признака связано именно с получаемым пациентами лекарством, а не вызван иными, в том числе случайными, факторами.

t-критерий Стьюдента можно применять к независимым выборкам. Ограничение заключается в том, что данный критерий применим только тогда, когда исходные данные имеют *нормальное распределение*. При несоблюдении этого условия вместо t-критерия должны использоваться аналогичные методы *непараметрической статистики*, среди которых наиболее известными являются *U-критерий Манна–Уитни* (для независимых выборок).

Перед выполнением статистических операций, имеющих целью подтвердить различия в нескольких выборках или доказать наличие связи между изменением некоторых показателей, всегда выдвигается так называемая *нулевая гипотеза* (H_0). В качестве нулевой гипотезы выступают гипотезы об отсутствии взаимосвязи, или корреляции между исследуемыми переменными, об отсутствии различий (однородности) в распределениях (параметрах распределений) в двух и/или более выборках. В стандартном научном подходе проверки гипотез исследователь пытается показать несостоятельность нулевой гипотезы, несогласованность её с имеющимися опытными данными, или отвергнуть гипотезу. То есть, исследователь проводит статистическое исследование, рассчитывает t-критерий Стьюдента или другие статистические показатели именно для того, чтобы опровергнуть нулевую гипотезу. Если ему это удаётся, тогда принимается вместо нулевой другая *альтернативная гипотеза* (H_a), которая исключает нулевую.

Итак, в качестве примера, возьмём две выборки с лугов с различными типами почв и определим для каждого растения значение признака, то есть высоту в сантиметрах:

Луг А:	Луг Б:
30, 45, 41, 38, 34, 36, 31, 30, 49, 50,	46, 49, 52, 37, 56, 40, 47, 51, 58,
51, 46, 41, 37, 36, 34, 33, 49, 32, 46,	46, 46, 56, 53, 37, 44, 42, 40, 37,
41, 44, 38, 50, 37, 39, 40, 46, 42.	54, 53, 51, 37, 56, 44, 42, 49.

Для луга А среднее арифметическое $M = 40,2$, среднеквадратичное отклонение $\sigma = 6,4$, ошибка средней $m = 1,2$; для луга Б $M = 47,0$, $\sigma = 6,7$, $m = 1,3$.

Сейчас необходимо проверить, есть ли статистически доказанные различия в двух выборках или их нет, а различие средних значений вызвано случайными факторами.

t-критерий Стьюдента рассчитывается по следующей формуле:

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}},$$

где M_1 – средняя арифметическая первой сравниваемой совокупности (группы);

M_2 – средняя арифметическая второй сравниваемой совокупности (группы);

m_1 – средняя ошибка первой средней арифметической;

m_2 – средняя ошибка второй средней арифметической.

Помимо самого t-критерия необходимо рассчитать показатель, который называется *число степеней свободы f*. В данном случае он определяется как сумма объёмов двух выборок минус два, то есть:

$$f = (n_1 + n_2) - 2.$$

На основе данных о числе степеней свободы и требуемой точности находим критическое значение t-критерия (таблица 1).

Таблица 1 – Критические значения t-критерия

f	$p < 0,05$	$p < 0,01$	f	$p < 0,05$	$p < 0,01$	f	$p < 0,05$	$p < 0,01$
1	12,706	63,65	24	2,064	2,797	56–57	2,003	2,667
2	4,303	9,925	25	2,060	2,787	58–59	2,002	2,663
3	3,182	5,841	26	2,056	2,779	60–61	2,000	2,660
4	2,776	4,604	27	2,052	2,771	62–63	1,999	2,657
5	2,571	4,032	28	2,048	2,763	64–65	1,998	2,655
6	2,447	3,707	29	2,045	2,756	66–67	1,997	2,652
7	2,365	3,499	30	2,042	2,750	68–69	1,995	2,650
8	2,306	3,355	31	2,040	2,744	70–71	1,994	2,648
9	2,262	3,250	32	2,037	2,738	72–73	1,993	2,646
10	2,228	3,169	33	2,035	2,733	74–75	1,993	2,644
11	2,201	3,106	34	2,032	2,728	76–77	1,992	2,642
12	2,179	3,055	35	2,030	2,724	78–79	1,991	2,640
13	2,160	3,012	36	2,028	2,719	80–89	1,990	2,639
14	2,145	2,977	37	2,026	2,715	90–99	1,987	2,632
15	2,131	2,947	38–39	2,024	2,712	100–109	1,984	2,626
16	2,120	2,921	40–41	2,021	2,704	110–119	1,982	2,621
17	2,110	2,898	42–43	2,018	2,698	120–139	1,980	2,617

18	2,101	2,878	44–45	2,015	2,692	140–159	1,977	2,611
19	2,093	2,861	46–47	2,013	2,687	160–179	1,975	2,608
20	2,086	2,845	48–49	2,011	2,682	180–199	1,973	2,605
21	2,080	2,831	50–51	2,009	2,678	200	1,972	2,601
22	2,074	2,819	52–53	2,007	2,674	∞	1,960	2,590
23	2,069	2,807	54–55	2,005	2,670			

Если расчётное значение *t*-критерия выше критического, то нулевая гипотеза отвергается и считается статистически доказанным с вероятностью более 95 %, то есть $p < 0{,}05$. Таким образом, считается статистически доказанной гипотеза о влиянии типов почв на высоту мятлика (разумеется, пока только лишь для изученных типов почв). Если же расчётное значение *t*-критерия окажется меньше критического, то нулевая гипотеза подтверждается, то есть различия сравниваемых величин статистически не значимы, а значит различия в средних значениях случайны, и гипотеза о влиянии типов почв на высоту мятлика не является статистически доказанной. В этом случае необходимо попытаться увеличить объём выборки, постараться исключить влияние наиболее сильных случайных факторов, а именно закладывать площади так, чтобы они различались только по типу почв и были более или менее похожими по другим параметрам – уровню грунтовых вод, экспозиции склонов, подстилающим породам, отсутствию антропогенного воздействия и т. д.

Любые статистические операции используют случайные величины. Поэтому остаётся вероятность ошибки для любого рассчитанного показателя. Эта ошибка называется *уровнем значимости* (*p*). Мы не можем, используя статистический аппарат, её полностью исключить, однако можем снизить вероятность этой ошибки до приемлемого уровня вероятности. В большинстве исследований, в том числе экологических, таким приемлемым уровнем является 5 %. Запись «$p < 0{,}05$» означает, что вероятность ошибки рассчитанного значения менее 5 %, «$p < 0{,}01$» – менее 1 % и т. д.

Следовательно, если критическое значение, определённое по таблице для случая $p < 0{,}05$ (для случаев с меньшей вероятностью ошибки критические значения существенно выше), оказалось меньше рассчитанного, что говорит о статистически достоверных различиях выборок, всё равно существует менее 5 % вероятности, что данный вывод ошибочен. Повторимся, этот уровень вероятности ошибки является приемлемым при естественнонаучных исследованиях.

Итак, рассчитаем t-критерий для двух выборок с лугов с различными типами почв.

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} = \frac{|47,0 - 40,2|}{\sqrt{1,2^2 + 1,3^2}} = 3,84.$$

Вычисляем число степеней свободы, учитывая, что в первой выборке (луг А) её объём $n = 29$, для луга Б $n = 26$:

$$f = (n_1 + n_2) - 2 = (29 + 26) - 2 = 53.$$

По таблице 1 определяем критическое значение t-критерия при числе степеней свободы 53 и $p < 0,05$. Он равен 2,007. Поскольку полученное значение больше, чем критическое, то нулевая гипотеза отвергается, и принимается, что две эти выборки действительно имеют статистически достоверные различия.

Рассчитаем t-критерий для следующего примера. Сравниваются два водоёма по плотности личинок комара. В качестве выборок выступают совокупности проб из этих водоёмов, в качестве признака – количество личинок в каждой пробе.

Водоём А: 8, 4, 6,10, 8, 7,12, 3. Водоём Б: 5, 6, 8, 7, 9, 8, 4, 4.

Здесь для водоёма А $M = 7,25$, $\sigma = 2,96$, $m = 1,05$; для водоёма Б $M = 6,36$, $\sigma = 1,92$, $m = 0,68$. Количество степеней свободы (8+8) – 2 = 14.

Рассчитаем t-критерий:

$$t = \frac{|7,25 - 6,36|}{\sqrt{1,05^2 + 0,68^2}} = 0,71.$$

Поскольку полученное значение меньше критического значения для числа степеней свободы (по таблице оно равно 2,145), то нулевая гипотеза принимается и считается, что статистически достоверных различий по исследуемому показателю между водоёмами нет.

Мы видим, что с увеличением числа степеней свободы снижается величина критических значений t-критерия, поэтому большой объём выборки в большей степени позволяет делать статистически доказанные выводы.

Помним, что расчёт t-критерия Стьюдента относится к *методам параметрической статистики*: исходные данные должны иметь *нормальное распределение*.

Однако данные могут и не иметь нормального распределения. В этом случае для сравнения двух независимых выборок по уровню какого-либо признака используют *методы непараметрической статистики*, например, *U-критерий Манна-Уитни*.

U-критерий не требует наличия нормального распределения сравниваемых совокупностей. Он подходит для сравнения малых выборок: в каждой из выборок должно быть не менее 3 значений признака. Допускается, чтобы в одной выборке было 2 значения, но во второй тогда должно быть не менее пяти.

Методика расчёта следующая: сначала из обеих сравниваемых выборок составляется *единый ранжированный ряд* путем расставления единиц наблюдения по степени возрастания признака и присвоения меньшему значению меньшего ранга. В таблице 2 в первой строчке все значения обеих выборок расположены в единый ряд в порядке возрастания. Значения, относящиеся к первой выборке, показаны полужирным шрифтом, ко второй – обычным.

Таблица 2 – Пример данных для расчета *U*-критерия

Значение признака	**3**	4	4	**4**	5	6	**6**	**7**	7	8	8	**8**	**8**	9	**10**	**12**
Ранг	**1**	2	3	**4**	5	6	**7**	**8**	9	10	11	**12**	**13**	14	**15**	**16**
Скорректированный ранг	**1**	3	3	**3**	5	6,5	**6,5**	**8,5**	8,5	11,5	11,5	**11,5**	**11,5**	11,5	**15**	**16**

Во второй строчке таблицы расставлены ранги значений по возрастанию – от самого меньшего значения, имеющего самый низкий ранг, к большему. Видно, что некоторые значения признака равны (три значения равны 4, два значения равны 6, четыре значения равны 8 и т. д.). Поэтому ряд, в котором показаны ранги, нужно скорректировать. В случае равных значений признака у нескольких единиц каждой из них присваивается среднее арифметическое последовательных значений рангов. То есть последовательно идут три значения, равные 4 с рангами 2, 3, 4. Из этих значений рангов мы вычисляем среднее (2 + 3 + 4) / 3 = 3, и это получившееся число присваиваем рангу всех трёх одинаковых значений признака. Если в ряду значений признака идут два значения равных 7 с рангами 8 и 9 (как в таблице 2), то находим среднее их рангов – 8,5 – и присваиваем его рангам обоих этих значений. В результате получается скорректированный ряд.

Подсчитаем отдельно сумму рангов для первой и второй выборок. Для водоёма А она равна 73, для водоёма Б – 63.

Находим значение *U*-критерия Манна-Уитни по формуле:

$$U = n_1 \cdot n_2 + \frac{n_x \cdot (n_x + 1)}{2} - T_x\ ,$$

где n_1 – объём первой выборки;

n_2 – объём второй выборки;

n_x – объём выборки, которая соответствует бóльшему значению суммы рангов;

T_x – бóльшая из двух сумм рангов.

Полученное значение U-критерия сравниваем по таблице с критическим значением U при заданной численности сопоставляемых выборок (таблица 3):

– если полученное значение U меньше табличного или равно ему, то признается статистическая значимость различий между уровнями признака в рассматриваемых выборках (принимается альтернативная гипотеза). Чем достоверность различий выше, тем меньше значение U;

– если же полученное значение U больше табличного, принимается нулевая гипотеза, то есть отсутствие статистически достоверных отличий.

Таблица 3 – Критические значения критерия Манна-Уитни

n_1	n_2													
	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
3	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4	3	4	4	5	6	7	8	9	10	11	11	12	13	13

Окончание таблицы 3

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	12	15	17	20	23	26	28	30	34	37	39	42	45	48
10	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	16	19	23	26	30	33	37	40	44	48	51	55	58	62
12	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	22	26	30	36	40	45	50	55	59	64	67	74	78	83
15	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	26	31	37	42	48	53	59	64	70	75	81	86	92	98
17	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	32	38	45	52	58	65	72	78	85	92	99	106	113	119

20	34	41	48	55	62	69	76	83	90	98	105	112	119	127

В нашем случае U = 27. Обе выборки имели одинаковый объём – 8. Ищем в таблице соответствующее этим объёмам значение. Оно равно 13. Рассчитанное значение намного больше табличного, что говорит об отсутствии статистически доказанных различий между выборками.

t-критерий Стьюдента и U-критерий Манна-Уитни применяются для сравнения *независимых* выборок. Если выборки *связанные,* или полученные при повторных измерениях одного параметра у одинаковых объектов (например, плотность вредителей на определённых растениях до обрызгивания инсектицидом и плотность вредителей на тех же растениях через три дня после обрызгивания), то применяется *парный t-критерий Стьюдента* (параметрический) или *T-критерий Уилкоксона* (непараметрический).

Парный t-критерий рассчитывается по формуле:

$$t = \frac{M_d}{\sigma_d / \sqrt{n}},$$

где M_d – средняя арифметическая разностей показателей одних и тех же объектов, измеренных до и после;

σ_d – среднеквадратичное отклонение разностей показателей;

n – число наблюдений (объём выборки).

Число степеней свободы оценивается по формуле $f = n - 1$. Далее, аналогично t-критерию для независимых выборок, полученное значение парного t-критерия сравнивается с критическим значением по той же таблице 1 для соответствующего числа степеней свободы. Если вычисленное значение больше или равно табличному, то нулевая гипотеза отвергается и принимается альтернативная.

T-критерий Уилкоксона не требует наличия нормального распределения. Он показывает, достоверно ли меняются показатели в сторону увеличения или уменьшения в связанной выборке после воздействия какого-либо фактора. Объём выборки должен быть не менее 5 и не более 50. Методика расчёта довольно проста. Сначала необходимо вычислить разность значений для каждой пары измерений – «до» и «после», определить, в какую сторону направлено большая часть изменений показателя: в сторону возрастания или в сторону уменьшения. На основании этого выделить типичные изменения (в ту сторону, куда направлено большинство изменений) и нетипичные (в противоположную). Затем для каждой вычисленной разности взять её модуль и ранжировать по модулю от меньших значений к большим и отдельно посчитать сумму рангов, соответствующих нетипичным изменениям. Полученное значение сравнить с критическим по таблице 4 для соответствующего числа наблюдений. Если полученное значение меньше критического или равное ему, то нулевая гипотеза отвергается; если больше, то признается статистическая значимость изменений показателя в типичную сторону (принимается альтернативная гипотеза).

Таблица 4 – Критические значения Т-критерия Уилкоксона

Т	$p <$		Т	$p <$		Т	$p <$		Т	$p <$		Т	$p <$	
	0,05	0,01		0,05	0,01		0,05	0,01		0,05	0,01		0,05	0,01
5	0		14	25	15	23	83	62	32	175	140	41	302	252
6	2		15	30	19	24	91	69	33	187	151	42	319	266
7	3	0	16	35	23	25	100	76	34	200	162	43	336	281
8	5	1	17	41	27	26	110	84	35	213	173	44	353	296
9	8	3	18	47	32	27	119	92	36	227	185	45	371	312
10	10	5	19	53	37	28	130	101	37	241	198	46	389	328
11	13	7	20	60	43	29	150	110	38	256	211	47	407	345
12	17	9	21	67	49	30	151	120	39	271	224	48	426	362
13	21	12	22	75	55	31	163	130	40	286	238	49	446	379

Например, на 8 метеостанциях региона измерили количество осадков за 2015 и 2016 гг. (таблица 5). Нужно определить, имеется ли достоверное увеличение количества осадков в 2016 г. по сравнению с 2015.

Таблица 5 – Пример для применения Т-критерия Уилкоксона

Метеостанция	Осадки в 2015 г., мм	Осадки в 2016 г., мм	Разность	Модуль разности	Ранг разности
1	634	642	8	8	6,5
2	602	612	10	10	8
3	**643**	**635**	**–8**	**8**	**6,5**
4	**652**	**648**	**–4**	**4**	**2**
5	597	603	6	6	3
6	625	632	7	7	4,5
7	**636**	**629**	**–7**	**7**	**4,5**
8	612	613	1	1	1

Для каждой метеостанции вычислили разность осадков в 2015 и 2016 гг. и определили типичные (в сторону увеличения, т. к. их 5) и нетипичные (в сторону уменьшения, их только 3) изменения. Нетипичные изменения выделили полужирным шрифтом. Для каждой разности определили и записали её модуль, затем в последней колонке проранжировали значения разности рангов. Сумма рангов, соответствующих нетипичным изменениям, равна:

$$T = 6,5 + 2 + 4,5 = 13.$$

По таблице смотрим критическое значение для числа наблюдений, равного восьми. Оно равно 5. Поскольку 13 > 5, то принимается нулевая гипотеза об отсутствии статистически достоверного изменения количества осадков.

Практическая работа 1
Расчёт показателей распределения случайной величины

1. При ежегодной оценке численности популяции вида животных в заказнике получены следующие результаты (таблица 6).

Таблица 6 – Варианты к заданию практической работы 1

Вариант	1 год	2 год	3 год	4 год	5 год	6 год
I	2 565	3 625	5 852	8 562	12 586	19 542
II	28	29	34	40	45	52
III	86	89	95	102	105	108
IV	586	625	701	798	863	985

Рассчитайте коэффициенты годового роста популяции для каждого года, начиная со второго, определите вид степенной средней, который корректно использовать для расчёта среднегодового значения коэффициента, и вычислите его.

2. Для вариационного ряда, согласно вариантам, рассчитайте показатели: среднее арифметическое и его ошибку, медиану, верхний и нижний квартили, децили, коэффициенты асимметрии и эксцесса, среднеквадратичное отклонение и его ошибку, коэффициент вариации и его ошибку.

Варианты:

– I: 29; 30; 28; 30; 27; 31; 28; 25; 24; 22; 37; 35; 33; 35; 32;

– II: 9,8; 8,6; 9,5; 5,8; 10,4; 14,5; 8,2; 10,9; 10,2; 9,4; 8,5; 12,8; 9,0;

– III: 86; 88; 91; 81; 83; 85; 86; 92; 86; 90; 82; 85; 82; 92; 84;

– IV: 4,3; 4,8; 5,0; 4,5; 3,8; 4,7; 3,7; 4,7; 4,6; 5,3; 3,8; 4,2; 5,5; 5,3; 2,6.

3. По данным о площадях и лесистости районов Беларуси вычислите лесистость областей методом расчёта среднего арифметического

взвешенного для лесистости районов соответствующей области (варианты: I – Витебская; II – Минская; III – Гродненская; IV – Брестская).

4. Для ряда чисел вычислите среднее арифметическое, среднее геометрическое, среднее гармоническое и среднее квадратичное. Сравните полученные результаты.

Варианты:

– I: 5,8; 6,3; 9,5; 6,8; 9,6; 4,3; 3,8; 5,9; 6,9; 5,4; 4,6; 8,8; 7,2;
– II: 25; 16; 38; 26; 32; 27; 22; 15; 18; 29; 33; 29; 18;
– III: 44; 63; 52; 22; 36; 48; 63; 74; 54; 41; 43; 29; 63;
– IV: 16; 36; 11; 18; 34; 25; 21; 19; 36; 25; 16; 18; 19.

5. Определите с помощью параметрического и непараметрического методов, имеются ли достоверные различия между выборками, либо они принадлежат к одной и той же генеральной совокупности?

Варианты:

I		II		III		IV	
Связанные выборки		Связанные выборки		Независимые выборки		Независимые выборки	
54	60	38	39	25	33	85	73
63	63	47	47	33	29	56	79
55	58	39	38	42	39	69	58
59	61	43	45	39	36	88	73
68	67	52	51	25	42	76	63
52	60	36	44	26	43	78	58
59	61	43	45	23	36	81	46
61	64	45	48	38	31	59	68
48	56	32	40	26	29	61	71
49	51	33	35	35	33	74	54
60	62	44	46	36	39	80	
59	57	43	41	19	26	78	
50	50	34	30	29	41	85	
58	63	42	44	34	35	89	

ТЕМА 2. МЕТОДЫ ОЦЕНКИ СВЯЗИ МЕЖДУ ПОКАЗАТЕЛЯМИ

Изучение природных объектов, обладающим большим количеством различных числовых характеристик позволяет поставить перед исследователем вопрос о наличии связей (*корреляций*) между различными характеристиками. Например, между средней температурой за сутки и уровнем активности каких-либо насекомых; между уровнем грунтовых вод и проективным покрытием мха кукушкин лён; между плотностью населения на определённых территориях и значением лесистости этих территорий. Корреляция может быть положительной (когда возрастание одного показателя сопровождается возрастанием второго) или отрицательной.

Во многих работах эта связь декларируется умозрительно, без привлечения статистического аппарата, что не даёт оснований утверждать о её объективном существовании. Применение статистических методов позволит:

1) определить само наличие связи между изменениями двух сравниваемых параметров;

2) вычислить количественное значение тесноты этой связи.

Как и в других разделах, здесь могут использоваться *параметрические* (условием применения которых являются нормальные распределения сопоставляемых переменных) и *непараметрические* (не имеющие этого ограничения) методы.

Наиболее популярным показателем является *коэффициент корреляции Пирсона* – метод параметрической статистики, позволяющий определить наличие или отсутствие линейной связи между двумя количественными показателями, а также оценить ее тесноту и статистическую значимость. В статистических расчетах и выводах коэффициент корреляции обычно обозначается как r_{xy} или R_{xy}.

При корреляционном анализе необходимо помнить следующее:

– понятия «связь» и «зависимость» не тождественны. Используя коэффициент Пирсона, мы можем лишь утверждать, что связь есть, но не можем утверждать, что эта связь имеет причинно-следственный характер и что один из параметров зависит от второго. В этом заключается одна из основных ошибок исследователей, которые, получив результат о наличии корреляционной статистический связи между двумя параметрами, интерпретируют его так, что один параметр зависит от другого. Они могут оба зависеть от какого-то третьего фактора;

– коэффициент Пирсона может определить наличие только *линейной* связи. Соответственно, низкое значение коэффициента Пирсона или статистическая незначимость рассчитанного коэффициента говорят не о том, что связи нет вообще, а о том, что нет линейной связи, хотя какая-нибудь другая связь вполне может быть;

– с помощью коэффициента Пирсона можно определить только наличие и тесноту связи. Прочие характеристики связи, в том числе направление (прямое или обратное), характер изменений (прямолинейный или криволинейный), а также наличие зависимости одной переменной от другой определяются при помощи *регрессионного анализа*;

– коэффициент Пирсона может определить связь только между *двумя* показателями. Если необходимо определить наличие взаимосвязей между тремя и более показателями, пользуются методами *факторного анализа*;

– значения коэффициента Пирсона могут быть от –1 до +1. Эти два значения свидетельствуют об абсолютной связи, и в реальности в естественнонаучных исследованиях практически не встречаются. Значение 0 означает полное отсутствие связи; значение коэффициента менее 0,3 свидетельствуют о слабой связи, 0,3 до 0,7 – о связи средней тесноты, более 0,7 – о сильной связи.

Коэффициент корреляции Пирсона вычисляется по формуле:

$$r_{xy} = \frac{\sum (d_x \cdot d_y)}{\sqrt{\left(\sum d_x^2 \cdot d_y^2\right)}},$$

где d_x – отклонение значения каждого показателя в вариационном ряду x от среднего арифметического для этого ряда;

d_y – отклонение значения каждого показателя в вариационном ряду y от среднего арифметического для этого ряда;

x и y – это ряды данных, между которыми мы проверяем наличие связи.

При расчёте коэффициента Пирсона удобно сначала заполнить таблицу. В качестве примера проверим связь между проективным покрытием мха *кукушкин лён*, % (x) и уровнем грунтовых вод, м (y), показанную в таблице 7.

Таким образом, подставляя полученные значения в формулу, вычисляют коэффициент корреляции Пирсона:

$$r_{xy} = \frac{-68,98}{\sqrt{3485,50 \cdot 1,59}} = -0,93.$$

Таким образом, связь между уровнем грунтовых вод и проективным покрытием мха *кукушкин лён* оказалась отрицательной и очень тесной.

Таблица 7 – Пример для расчёта коэффициента корреляции Пирсона

x	y	d_x	d_y	d_x^2	d_y^2	$d_x \cdot d_y$

65	0,2	–9,75	0,29	95,06	0,08	–2,80
23	1,3	32,25	–0,81	1 040,06	0,66	–26,20
88	0,1	–32,75	0,39	1 072,56	0,15	–12,69
45	0,6	10,25	–0,11	105,06	0,01	–1,15
27	1,1	28,25	–0,61	798,06	0,38	–17,30
59	0,4	–3,75	0,09	14,06	0,01	–0,33
73	0,1	–17,75	0,39	315,06	0,15	–6,88
62	0,1	–6,75	0,39	45,56	0,15	–2,62
Ср. = 55,3	**Ср. = 0,49**			**Σ = 3 485,50**	**Σ = 1,59**	**Σ = – 69,98**

Осталось проверить статистическую значимость полученного значения. Она осуществляется при помощи *t*-критерия, рассчитываемого по следующей формуле:

$$t_r = \frac{|r_{xy}| \cdot \sqrt{n-2}}{\sqrt{1-r_{xy}^2}}.$$

В нашем примере:

$$t_r = \frac{0,93 \cdot \sqrt{8-2}}{\sqrt{1-0,86}} = 6,09.$$

Число степеней свободы $f = n - 2$ равно 6. По таблице 1 критических значений *t*-критерия находим критическое значение для этого числа степеней свободы. Видим, что полученное нами значение t превышает критическое как для $p < 0,05$, так и для $p < 0,01$, а значит, наличие связи между рассматриваемыми характеристиками считается статистически доказанным, следовательно связь является статистически значимой при $p < 0,01$.

Коэффициент Спирмена применяется, когда зависимость между двумя показателями не является линейной или распределение параметров, между которыми проверяется наличие связи, не является нормальным. Он применяется также в тех случаях, когда известны не числовые значения характеристик показателей, а лишь их ранги. Поэтому его ещё одно название – *коэффициент ранговой корреляции*.

Свойства данного коэффициента такие же, как у коэффициента Пирсона: изменение от –1 до +1. Чем больше абсолютное значение, тем больше теснота связи.

Для расчёта необходимо для каждого абсолютного значения рассматриваемых параметров вычислить значение его ранга. Затем найти разности рангов каждой пары, возвести в квадрат каждую разность, суммировать полученные результаты и рассчитать коэффициент Спирмена по формуле:

$$\rho_{xy} = 1 - \frac{6 \cdot \sum d^2}{n \cdot (n^2 - 1)},$$

где d – разность рангов каждой пары параметров.

Для примера рассчитаем коэффициент Спирмена по тому же примеру, по которому рассчитывали коэффициент Пирсона (таблица 8).

$$\rho_{xy} = 1 - \frac{6 \cdot 158}{8 \cdot (8^2 - 1)} = -0,88.$$

Таблица 8 – Пример для расчёта коэффициента корреляции Симпсона

x	y	ранг x	ранг y	d	d^2
65	0,2	6	4	2	4
23	1,3	1	8	–7	49
88	0,1	8	2	6	36
45	0,6	3	6	–3	9
27	1,1	2	7	–5	25
59	0,4	4	5	–1	1
73	0,1	7	2	5	25
62	0,1	5	2	3	9
					Σ = 158

Теперь необходимо проверить статистическую значимость полученного результата. Это осуществляется точно так же, как для прошлой формулы – по значению t-критерия, только вместо r_{xy} в формулу подставляем вычисленное значение коэффициента Спирмена ρ_{xy}. Число степеней свободы тоже определяется аналогично. В данном случае t-критерий будет равен 6,22, а значит, проверив по таблице критических значений, убеждаемся, что корреляция между этими показателями статистически значима. При расчёте коэффициента желательно, чтобы значения показателей каждого параметра не повторялись, это обеспечит более точный результат.

Кроме количественных, в природных системах существуют те параметры, которые характеризуются качественно: типы почв, экологические группы организмов, положение в пределах речных бассейнов, наличие или отсутствие определённых организмов или признаков в сообществе и т. д. Между такими параметрами тоже может существовать статистическая связь. Рассмотрим статистические показатели, которые описывают связь между двумя параметрами, характеризующимися такими признаками, как «наличие» или «отсутствие». К примеру, необходимо определить, есть ли связь какого-либо вида растения с определённым типом почв, то есть растёт ли этот вид преимущественно на этой почве или никакой связи нет. Для этого потребуется совокупность наблюдений, в каждом из которых фиксировалось бы наличие или отсутствие вида растений, а также наличие или отсутствие типа почв. В результате, все наблюдения (N) можно будет разделить на 4 группы:

A – наблюдения, где фиксировался и вид растений, и тип почвы;

B – наблюдения, в которых фиксировался только вид растений, а тип почв отсутствовал;

C – наблюдения, в которых фиксировался только тип почв, а вид растений отсутствовал;

D – наблюдения, где не были отмечены ни вид, ни почва.

Одним из наиболее популярных статистических методов, с помощью которых проверяют наличие связи между двумя качественными показателями, является *критерий хи-квадрат Пирсона*.

При применении критерия хи-квадрат полученные данные наблюдений заносятся в таблицу, которая называется четырёхпольная таблица сопряжённости (таблица 9). Таблица может состоять и более, чем из четырёх полей, а каждый параметр может оцениваться более, чем двумя качественными состояниями, однако такие методы не затрагиваются в настоящем пособии.

Таблица 9 – Четырёхпольная таблица сопряжённости

Параметр II / Параметр I	есть (1)	нет (0)	Всего
есть (1)	A	B	$A + B$
нет (0)	C	D	$C + D$
Всего	$A + C$	$B + D$	$A + B + C + D$

При расчёте критерия хи-квадрат необходимо сначала *рассчитать ожидаемое количество наблюдений* для каждой из ячеек таблицы сопряженности (то есть, как распределились бы наблюдения при условии справедливости нулевой гипотезы об отсутствии взаимосвязи) путем перемножения сумм рядов и столбцов с последующим делением полученного произведения на общее число наблюдений (таблица 10).

Таблица 10 – Расчёт ожидаемого количества наблюдений

Параметр II / Параметр I	есть (1)	нет (0)	Всего

есть (1)	$(A+B) \times (A+C) / (A+B+C+D)$	$(A+B) \times (B+D) / (A+B+C+D)$	$A + B$
нет (0)	$(C+D) \times (A+C) / (A+B+C+D)$	$(C+D) \times (B+D) / (A+B+C+D)$	$C + D$
Всего	$A + C$	$B + D$	$A + B + C + D$

Затем полученные для каждого из 4 полей таблицы значения подставляем в формулу для расчёта критерия хи-квадрат:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

где i – номер строки (от 1 до r);

j – номер столбца (от 1 до c);

O_{ij} – фактическое количество наблюдений в ячейке ij;

E_{ij} – рассчитанное ожидаемое число наблюдений в ячейке ij.

Число степеней свободы $f = (r - 1) \times (c - 1)$. То есть для четырехпольной таблицы, в которой 2 ряда ($r = 2$) и 2 столбца ($c = 2$), число степеней свободы составляет всегда 1 ($f_{2x2} = (2 - 1) \times (2 - 1) = 1$).

Осталось только сравнить полученное значение χ^2 с критическим значением по таблице 11 при числе степеней свободы, равном 1.

Таблица 11 – Критические значения критерия χ^2

f	$p < 0{,}05$	$p < 0{,}01$	f	$p < 0{,}05$	$p < 0{,}01$	f	$p < 0{,}05$	$p < 0{,}01$
1	3,841	6,635	4	9,488	13,277	7	14,067	18,475
2	5,991	9,21	5	11,07	15,086	8	15,507	20,09
3	7,815	11,345	6	12,592	16,812	9	16,919	21,666

Если расчётное значение выше критического, то нулевая гипотеза отвергается, и наличие связи между двумя параметрами считается статистически доказанным.

Для определения количественного значения тесноты этой связи предлагается ряд коэффициентов, из которых можно выделить трансформированный коэффициент Дайса:

$$K_D = \frac{A - \min(B, C)}{A + \min(B, C)},$$

где A – число наблюдений в ячейке А таблицы 9 (то есть число наблюдений, где присутствовали оба параметра), min (B, C) – минимальное значение из ячеек B и C. Коэффициент может принимать значения от –1 (когда нет ни одного случая совместной встречи двух параметров) до +1 (когда по крайней мере один из двух параметров никогда не встречался без второго).

В том случае, если число ожидаемого явления меньше 10 хотя бы в одной ячейке, при анализе четырехпольных таблиц должен рассчитываться *критерий хи-квадрат с поправкой Йейтса*. Данная поправка позволяет уменьшить вероятность ошибки первого типа, т. е. обнаружения различий там, где их нет. Формула для расчёта критерия хи-квадрат с поправкой Йейтса следующая:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(\left|O_{ij} - E_{ij}\right| - 0,5\right)^2}{E_{ij}}.$$

К примеру, необходимо выяснить взаимосвязь в распространении двух видов растений – крапивы двудомной и недотроги обыкновенной. Для этого были использованы описания 132 пробных площадей. Из них на 58 встречались оба вида, на 22 – только крапива, на 38 – только недотрога. Таким образом, N = 132, A = 88, B = 22, C = 12, D = 10. Рассчитав ожидаемое количество наблюдений, получаем результаты, представленные в таблице 12.

Таблица 12 – Результаты расчёта ожидаемого количества наблюдений при вычислении критерия хи-квадрат

Параметр II / Параметр I	Крапива есть (1)	Крапивы нет (0)	Всего
Недотрога есть (1)	(88+22) × (88+12) / (88+22+12+10) = **83,3**	(88+22) × (22+10) / (88+22+12+10) = **26,7**	$A + B$
Недотроги нет (0)	(12+10) × (88+12) / (88+22+12+10) = **16,7**	(12+10) × (22+10) / (88+22+12+10) = **5,3**	$C + D$
Всего	$A + C$	$B + D$	$A + B + C + D$

Поскольку в ячейке D значение получилось меньше 10, то вычислим *хи-квадрат с поправкой Йейтса*:

$$\chi^2 = \frac{\left(|88 - 83{,}3| - 0{,}5\right)^2}{83{,}3} + \frac{\left(|22 - 26{,}7| - 0{,}5\right)^2}{26{,}7} +$$

$$+ \frac{\left(|12 - 16{,}7| - 0{,}5\right)^2}{16{,}7} + \frac{\left(|10 - 5{,}3| - 0{,}5\right)^2}{5{,}3} = 5{,}26.$$

Поскольку 5,26 больше, чем критическое значение 3,84, то считается доказанным, что крапива и недотрога распространены не независимо. Между ними существует взаимосвязь, а значит, они обладают сходной реакцией на те факторы и условия природной среды, которые были представлены на исследованных пробных площадях. Значение K_D = (88 – 12) / (88 + 12) = +0,76.

Если хотя бы в одной ячейке ожидаемое явление меньше 5, то для анализа должен использоваться *точный критерий Фишера*. Он вычисляется по формуле:

$$P = \frac{(A+B)! \cdot (C+D)! \cdot (A+C)! \cdot (B+D)!}{A! \cdot B! \cdot C! \cdot D! \cdot N!},$$

где N – общее число наблюдений;

! – знак факториала, представляющего собой произведение числа на последовательность чисел, каждое из которых меньше предыдущего на 1 (например, 6! = 6·5·4·3·2·1).

Отличительной особенностью точного критерия Фишера является то, что полученное значение P уже является уровнем значимости при котором отвергается нулевая гипотеза и принимается альтернативная. Если это значение меньше 0,05, то признаётся доказанное наличие статистической связи между двумя показателями.

К примеру, изучалась взаимосвязь распространения ослинника двулетнего и песчаных почв. Всего было сделано 85 описаний, из них 15 описаний, где отмечены и песчаная почва, и ослинник, 2 описания, где отмечен ослинник, но почва не песчаная, 40 описаний, где почва песчаная, но нет ослинника, и 28 описаний, где не отмечено ни песчаной почвы, ни ослинника. Таким образом, N = 85, A = 15, B = 2, C = 40, D = 28. Вычислим *точный критерий Фишера*:

$$P = \frac{(15+2)! \cdot (40+28)! \cdot (15+40)! \cdot (2+28)!}{15! \cdot 2! \cdot 40! \cdot 28! \cdot 86!} = 0{,}016.$$

Полученное значение 0,016 меньше 0,05, то есть связь между песчаными почвами и ослинником двулетним считается доказанной.

Практическая работа 2
Расчёт коэффициентов корреляции

1. Используя коэффициенты корреляции Пирсона и Спирмена и данные таблицы 13, определите, есть ли связь между:

а) плотностью сельского населения административных районов Гомельской области (*p*, чел./км^2) и их лесистостью (*f*, %);

б) коэффициентом урбанизации (*D*, %) и смертностью от новообразований (*M*, умерших на 100 000 человек населения);

в) уровнем преступности (*C*, число зарегистрированных преступлений в год на 1000 жителей) и долей занятых от общего количества трудоспособных (*e*, %);

г) возрастанием численности населения в районе с 2009 по 2019 гг. (*K*, %) и долей населения района, указавшего русский язык родным в ходе переписи 2019 г. (*L*, %).

2. При исследовании связи между распространением клёна ясенелистного и мелколепестника канадского было описано 89 площадок, где они встречались совместно, 56 площадок, где встречался только клён, 43 площадки только мелколепестник и 80 площадок, где не было зафиксировано ни одного из этих видов. Определите, есть ли связь между распространением двух названных видов.

3. Из 47 обследованных регионов в 14 зафиксировано наличие африканской чумы свиней и бешенства, в 3 только чума, в 6 только бешенство. Есть ли связь между этими двумя заболеваниями?

4. При исследовании влияния курения на артериальную гипертензию было исследовано 25 человек, из которых курящих и одновременно страдающих гипертензией 8, курящих и не страдающих гипертензией – 2, не курящих и страдающих – 5. Определите, есть ли связь между курением и наличием артериальной гипертензии?

Таблица 13 – Исходные данные к заданию 1

Район	Брагинский	Буда-Кошелевский	Ветковский	Гомельский	Добрушский	Ельский	Житковичский	Жлобинский	Калинковичский	Кормянский	Лельчицкий	Лоевский	Мозырский	Петриковский	Наровлянский	Октябрьский	Речицкий	Рогачевский	Светлогорский	Хойникский	Чечерский
p	3,8	13,9	6,4	32,4	11,5	5,2	6,8	12,4	7,8	7,6	4,6	6,2	11,6	6,1	2,0	5,6	11,2	12,0	7,5	3,7	6,2
f	50,6	25,8	49,7	36,3	28,5	57,3	55,7	33,9	51,5	34,9	69,7	38,4	55,2	56,6	71,4	58,6	45,5	34,9	53,1	66,6	49,7
D	62,5	42,9	57,7	49,1	53,8	43,2	53,1	44,1	56,8	59,8	62,5	54,7	58,3	52,1	51,8	51,7	38,9	34,2	51,8	53,6	69,1
M	102,8	161,9	162,4	158,4	113,0	112,5	139,8	153,3	136,6	125,7	102,8	188,3	179,5	142,9	97,0	100,2	177,3	171,8	181,1	157,1	147,6
C	10,4	8,8	12,1	10,9	11,1	8,0	12,0	12,7	11,4	10,2	9,0	11,9	11,3	11,9	6,6	9,9	10,7	14,6	8,7	14,0	9,9
e	83,2	83,2	67,1	67,8	76,6	88,8	73,6	83,4	71,4	75,6	77,4	74,8	70,9	73,9	84,0	89,9	79,1	73,8	79,0	75,3	72,9
K	–12,1	–12,3	–6,1	3,4	–9,7	–14	–14,4	–3,5	–9,7	–8,4	–9,5	–16,8	–0,8	–5,8	–11,2	–18,4	–5,7	–11,1	–10,2	–12,5	–6,5
L	23,2	39,2	47,4	55,2	75,3	45,9	42,5	58,7	34,8	48,7	26,3	27,1	45,7	37,6	32,3	22,7	57,0	42,1	48,8	33,8	32,9

ТЕМА 3. ФИТОИНДИКАЦИЯ, ОРДИНАЦИЯ, ОЦЕНКА РАЗНООБРАЗИЯ

Помимо статистических существует большое количество других методов получения количественных характеристик экосистем на основе камеральной обработки данных полевого материала. Среди таких методов можно выделить методы фитоиндикации характеристик абиотических компонентов экосистем (экологических режимов влажности, кислотности, освещённости и др.), методы оценки разнообразия сообществ, методы количественной оценки сходства и различия между сообществами, методы ординации (то есть упорядочения видов и сообществ по какому-либо фактору или по осям координат двух или трёх факторов).

Фитоиндикация экологических режимов. Абиотические компоненты находятся в неразрывной связи с живыми организмами, формируя вместе с ними целостные природные системы. Любые их изменения по системам прямых и обратных связей отражаются на характеристиках живых организмов, которые, в свою очередь, воздействуют на неживые компоненты и изменяют их. Поэтому изолированное изучение только живых организмов, не обращая внимания на характеристики их природного окружения, не позволит в полной мере познать особенности экосистем и выявить объективные закономерности из развития.

Для биоиндикации состояния неживых компонентов – почв, горных пород, микроклимата и т. д. – применяются главным образом растительные организмы, так как они более тесно связаны с абиотическими компонентами, чем животные, более объективно в силу своей неподвижности отображают пространственные особенности изменения природной среды, более физиономичны, что облегчает их изучение.

С помощью методов биоиндикации обычно определяют такие характеристики, как влажность почвы, условия освещения, кислотность, гранулометрический состав, трофность, содержание солей и т. д.

Для оценки количественных значений этих характеристик и их динамики (то есть экологических режимов) были разработаны специальные экологические шкалы. Одним из первых их стал использовать Л. Г. Раменский для экологической оценки лугов в сельскохозяйственных целях (середина 20 века). В дальнейшем экологические шкалы разрабатывались и другими исследованиями, наиболее известными из которых были Д. Н. Цыганов, Г. Элленберг и Э. Ландольт.

Суть экологических шкал состоит в том, что весь возможный диапазон значений факторов среды (например, влажности почв) разбивается на некоторое число градаций, обозначаемых баллами. Каждому виду растения по каждому экологическому режиму присваивается определённый балл (шкалы Элленберга и Ландольта) или диапазон баллов (шкалы Раменского и Цыганова) в соответствии с экологическими свойствами этого вида по различным факторам среды.

Например, система экологических шкал Д. Н. Цыганова состоит из 10 шкал:

– увлажнения почв (23 градации),
– богатства почв азотом (11),
– трофности почв (19),
– кислотности почв (13),
– переменности увлажнения почв (11),
– режима освещённости/затенения (9),
– термоклиматическая шкала (17),
– шкала континентальности климата (15),
– шкала аридности/гумидности климата (15),
– криоклиматическая шкала (15).

Каждому растению присваивается диапазон баллов по каждому фактору. Например, по шкале влажности почв цмину песчаному присвоено значение от 3 до 11 баллов, а осоке дернистой – от 11 до 19 баллов; по шкале режима освещённости/затенения иван-чаю обыкновенному присвоено значение от 1 до 6 баллов, а кислице обыкновенной – от 3 до 9.

Система экологических шкал Г. Элленберга состоит из 6 шкал:

– увлажнения почв (12 градаций),
– богатства почв азотом (9),
– кислотности почв (9),
– режима освещённости/затенения (9),
– термоклиматическая шкала (9),
– шкала континентальности климата (9).

Каждому растению присваивается балльное значение. Например, по шкале влажности почв паслёну сладко-горькому присвоено 8 баллов, а винограду культурному – 4 балла; по шкале кислотности почв зверобою изящному 8 баллов, а мальве приземистой – 5 баллов. Чем больше градаций содержит шкала, тем детальнее дифференцируются местообитания. Все 4 шкалы (Элленберга, Цыганова, Раменского, Ландольта) можно скачать по адресу: http://cepl.rssi.ru/bio/flora/ecoscale.htm.

Естественно, что для каждой шкалы имеются ограничения в применении. Бессмысленно применять шкалу континентальности климата, если все исследования проводятся на относительно небольшой территории,   в пределах одного-двух административных районов, так как очевидно, что степень континентальности там будет одинакова. Может отсутствовать также смысл применения шкалы влажности почв для характеристики экосистем различных природных зон, так как изменение этого показателя определяется в первую очередь локальными факторами.

На основе балльных оценок отдельных видов переходят к балльным оценкам экологических режимов всей экосистемы в целом. Итоговая балльная оценка по некоторому фактору вычисляется как среднее значение из балльных оценок всех видов по этому фактору, взвешенное на относительное обилие видов:

$$I = \sum_{i=1}^{n} \frac{p_i \cdot K}{P}.$$

где I – искомое значение какого-либо экологического фактора для всей экосистемы в целом;

p_i – значение показателя обилия вида (например, значение проективного покрытия, балла по шкале обилия и т. п.);

K – значение балла, присвоенного данному виду по шкале экологического режима, если шкала точечная (или по значению середины интервала баллов, если шкала интервальная);

P – сумма показателей обилия всех видов в сообществе.

Оценки, полученные для различных местообитаний, сравниваются, и делаются выводы об особенностях экологических режимов в различных условиях. Например, в результате полевых работ описаны 3 местообитания (таблица 14).

Для каждого местообитания определён видовой состав растительности и проективное покрытие каждого вида. Эти данные занесены в таблицу. Там же показаны значения каждого вида по интервальным экологическим шкалам Цыганова (min – нижняя граница интервала, max – верхняя граница интервала, ср. – значение середины интервала), по шкалам влажности и освещённости, вычисленной как среднее арифметическое минимального и максимального.

Рассчитаем показатель влажности почв для соснового леса:

$$I = \frac{40 \cdot 13 + 5 \cdot 12 + 25 \cdot 14 + 5 \cdot 10 + 20 \cdot 11 + 5 \cdot 13 + 15 \cdot 13}{40 + 5 + 25 + 5 + 20 + 5 + 15} = 12,7.$$

Таким образом, по шкале влажности почв Цыганова, где 23 градации (1 баллу соответствуют самые сухие, а 23 баллам – самые влажные условия), описанный сосновый лес получил 12,7 балла.

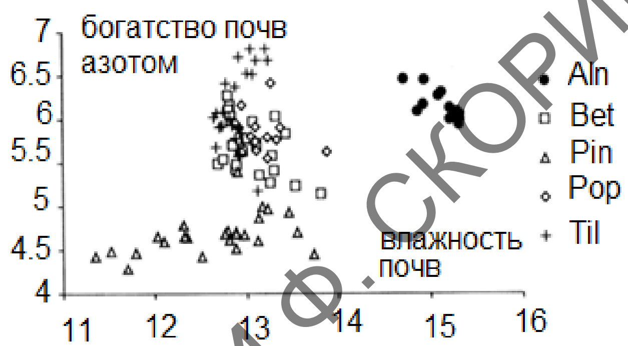
Таблица 14 – Пример трёх описанных местообитаний

Вид	Проективное покрытие, %	Влажность почв			Освещённость		
		min	max	ср.	min	max	ср.
Сосновый лес							
Мох Шребера	40	9	17	13	1	7	4
Марьянник лесной	5	9	15	12	1	8	4,5
Вереск обыкновенный	25	9	19	14	1	6	3,5
Ястребинка волосистая	5	5	15	10	1	5	3
Овсяница овечья	20	7	15	11	1	6	3,5
Ландыш майский	5	8	18	13	3	7	5
Орляк обыкновенный	15	9	17	13	2	8	5
Дубовый лес							
Сныть обыкновенная	20	9	15	12	1	8	4,5
Гравилат городской	5	9	19	14	1	7	4
Звездчатка лесная	10	11	15	13	3	9	6
Кислица обыкновенная	10	10	16	13	3	9	6
Мятлик дубравный	5	7	19	13	3	9	6
Зеленчук жёлтый	15	9	15	12	3	9	6
Будра плющевидная	5	7	17	12	3	7	5
Копытень европейский	5	11	15	13	4	9	6,5
Черноольховый лес							
Вербейник обыкновенный	35	11	21	16	1	7	4
Шлемник обыкновенный	10	11	21	16	1	6	3,5
Щитовник картузианский	5	11	19	15	3	9	6
Зюзник европейский	10	11	21	16	1	5	3
Осока ложносытевая	15	13	21	17	1	5	3
Подмаренник болотный	5	11	21	16	1	8	4,5
Белокрыльник болотный	5	11	21	16	1	9	5
Паслен сладко-горький	10	11	21	16	2	9	5,5
Сабельник болотный	15	12	21	16,5	1	5	3

Аналогично вычисляется балл увлажнения почв для дубового (12,5 балла) и черноольхового леса (16,2). По показателю освещённости, шкала которой делится на 9 градаций (где 1 баллу соответствуют самые освещённые, а 9 баллам – самые тёмные местообитания), сосновый лес получает 4,0 балла, дубовый – 5,4, черноольховый – 3,9 балла. Полученные результаты можно отобразить графически путём построения прямоугольной системы координат, где осями являются данные экологические шкалы, и отметить на неё точки, имеющие соответствующие координаты по влажности и освещённости (рисунок 4а).

При большом количестве описаний, выполненных в различных экологических условиях, их расположение в подобной системе координат позволит выделить группы сходных по экологическим условиям описаний, упорядочить и классифицировать их, сделать выводы о факторах их организации (пример на рисунке 4б).

а



б

Aln – ольшаники; Bet – березняки; Pin – сосняки; Til – липняки; Pop – осинники

Рисунок 4 – Графическое представление положения описаний в пространстве экологических факторов

Оценка сходства и разнообразия сообществ. При обработке материалов полевых исследований всегда встаёт вопрос о сравнении различных описаний (фитоценозов, экосистем, зооценозов и т. д.) между собой по каким-либо численным критериям. Одним из важнейших критериев является видовой состав сообщества. Каким образом численно определить степень сходства двух и более сообществ между собой и сравнить, какие из них более схожие? Для этого предложен ряд показателей, которые называются коэффициентами сходства.

Первая попытка количественного выражения степени сходства между сообществами принадлежала в 1901 г. швейцарскому исследователю П. Жаккару. *Коэффициент флористического сходства Жаккара* до сих пор широко используется в геоботанике. Для сравнения сходства видового состава двух сообществ необходимо посчитать число видов в первом сообществе (*a*), число видов во втором сообществе (*b*) и число видов, общих для первого и второго сообществ (*c*). Далее с помощью этих данных вычислить коэффициент Жаккара по формуле:

$$K_G = \frac{c}{a+b-c}.$$

Не менее широко применяется *коэффициент сходства Серенсена*:

$$K_S = \frac{2c}{a+b}.$$

Оба этих коэффициента изменяются от 0 до 1. Они равны 1 при полном сходстве, то есть, когда оба сообщества имеют идентичный видовой состав. Если в двух сообществах нет ни одного общего вида, оба они равны нулю.

Несмотря на почти повсеместную традицию использовать для оценки сходства биоценозов указанные коэффициенты, ряд авторов считает, что их наиболее существенный недостаток – это то, что они учитывают лишь факт наличия или отсутствия вида в сообществе и совершенно не учитывают его обилие. Между тем, обилие видов является важным показателем, отражающим экологические условия обитания сообщества. Если некоторые виды в одном сообществе встречаются единично, а во втором преобладают, другие же виды наоборот, то коэффициент сходства Жаккара или Серенсена всё равно покажет высокое значение сходства между сообществами, хотя очевидно, что они существуют в экологически неравноценных условиях и по-разному будут влиять на животный компонент, почву и т. д. Кроме того, если исследователь потратил время и усилия на определение количественных характеристик обилия видов, а не просто зафиксировал их наличие, логично было бы, если бы его труд не был напрасен и был бы использован для более точного подсчёта определяемого показателя, в данном случае сходства. При подсчете мер сходства показателей обилия, выраженных в абсолютных или относительных значениях видовой численности или биомассы, возможно использование коэффициента К. Чекановского:

$$M_T = \frac{\sum_{i=1}^{S} \min(X_i, Y_i)}{\sum_{i=1}^{S} X_i + \sum_{i=1}^{S} Y_i},$$

где $\sum \min(X_i, Y_i)$ – сумма минимальных значений из показателей обилия видов, общих для двух сравниваемых описаний;

$\sum X_i$ – сумма значений обилия видов первого описания;

$\sum Y_i$ – сумма значений обилия видов второго описания;

S – общее количество видов.

Разнообразие также является важной характеристикой экосистемы. Проблема разнообразия является одной из центральных проблем экологии, и изучение этого вопроса не может осуществляться без конкретных количественных показателей, характеризующих те аспекты живой природы, которые объединяются под данным термином: количество составляющих биологической системы, её сложность, разнокачественность компонентов, неравномерность их представленности в системе и др. на каждом иерархическом уровне организации живой материи, изучаемом в рамках экологии.

Единого подхода к количественной оценке разнообразия экосистем и их компонентов до настоящего времени нет. Предложено огромное количество различных показателей разнообразия, которые отображают различные аспекты этого сложного явления, по-разному интерпретируются, и ни один из них не может претендовать на универсальность. Большинство из этих показателей так или иначе в самых различных формах учитывают два аспекта: 1) число видов; 2) относительное обилие видов.

Простейшими показателями являются *видовая напряжённость* – число видов на определённую площадь, объём, пробу и т. д.; *нумерическое видовое богатство* – число видов на определенное число обследованных особей (например, число видов на 1 000 рыб).

Соотношение числа видов (s) и числа особей (или другого показателя суммы обилия видов) (N) положено в основу *коэффициента разнообразия Маргалефа*:

$$d = \frac{s-1}{\ln(N)}.$$

К индексам, отображающим одной величиной оба аспекта разнообразия – число видов и их относительное обилие, относятся:

– *индекс разнообразия Симпсона*:

$$D = 1 - \sum_{i=1}^{k} \frac{n_i}{N};$$

– *индекс разнообразия Шеннона*:

$$H = 1 - \sum_{n=1}^{k} \frac{n_i}{N} \ln \frac{n_i}{N},$$

где n_i – обилие i-го вида;

N – сумма обилий всех видов.

Из этих двух обобщенных индексов индекс Симпсона придает обычным видам больший вес (поскольку при возведении в квадрат малых отношений n_i / N получаются очень малые величины). Индекс Шеннона придает больший вес редким видам. В этом его достоинство: как указывает Ю. Г. Пузаченко, «в теории информации наиболее информативны редкие типы событий. С этим информационным эффектом, в частности, можно связать и ощущаемую человеком необходимость сохранения редкостей. Редкости информативны уже потому, что они существуют на границе области устойчивости, и это неопределенное положение создает условия для получения информации о пределах возможного в динамике систем». Обычно значения индекса Шеннона лежат в пределах от 1,5 до 3,5, редко превышая 4,5.

Если же необходимо рассчитать только степень равномерности распределения видов в сообществе, чтобы этот показатель не зависел от числа видов, применяют *индекс выравненности Пиелу*.

$$E = \frac{H}{\ln S},$$

где S – количество видов. Индекс меняется от 0 до 1, причём 1 соответствует ситуация, когда относительная численность всех видов в сообществе одинакова.

Индекс разнообразия Бриллуэна вычисляется по формуле:

$$HB = \frac{\ln N! - \sum \ln n_i!}{N}.$$

Когда оценивают разнообразие единичной выборки, индекс Шеннона всегда будет больше индекса Бриллуэна. Индекс Шеннона в отличие от индекса Бриллуэна не меняется, если число видов и их относительные доли постоянны.

Отразить оба компонента разнообразия – богатство видов и выравненность – можно графически, построив график, где по оси *y* отложено число особей, биомасса или иной рассматриваемый показатель, а по оси *x* – ранжированная последовательность видов от наиболее до наименее обильного. На рисунке 5 показаны некоторые варианты кривых, которые могут получиться при построении графика.

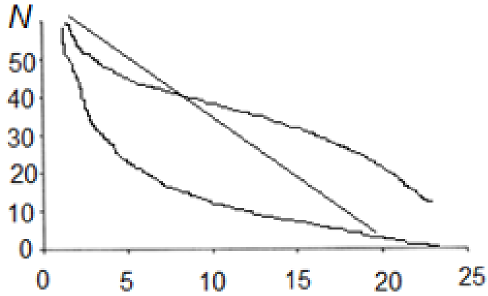


Рисунок 5 – Возможные формы кривых доминирования–разнообразия

Кроме рассмотренных выделяются индексы доминирования, уделяющие основное внимание обилию самых обычных видов, а не видовому богатству. Один из лучших среди таких индексов – *индекс доминирования Симпсона*:

$$S = \sum_{i=1}^{k} \frac{n_i}{N}.$$

Также этот индекс можно вычислить, отняв значение индекса разнообразия Симпсона от единицы: $S = 1 - D$.

Индекс доминирования Макинтоша:

$$DM = \frac{N - \sqrt{\sum n_i^2}}{N - \sqrt{N}}.$$

Индекс разнообразия Бергера-Паркера выражает относительную значимость наиболее обильного вида:

$$d = \frac{N_{\max}}{N},$$

где $N_{\max}$ – показатель обилия (число особей, биомасса, проективное покрытие и т. п.) самого обильного вида.

Ординация сообществ. Поскольку растительный покров носит в основном непрерывный характер, то сообщества (экосистемы) переходят друг в друга постепенно, переходные зоны между сообществами различных типов могут превышать размер самих сообществ. Изменения сообществ под влиянием антропогенных факторов также проявляются в пространстве постепенно, усиливаясь по мере приближения к источнику воздействия. Всё это делает почти невозможным установление объективно существующих чётких границ между различными экосистемами. Выделение сообществ из непрерывного континуума носит в большинстве случаев субъективный характер и определяется целями исследования и рабочей гипотезой.

Результатом полевого исследования, заключающегося, например, в определении влияния какого-либо антропогенного фактора на структуру и характеристики экосистемы, является множество описаний, выполненных в экосистемах с различным уровнем трансформации – от слабонарушенных до сильнонарушенных. Эти нарушения, как указано выше, проявляются в пространстве постепенно без чётко выраженных границ, по которым можно отделить различные стадии или уровни антропогенной трансформации. Тем не менее, в подавляющем большинстве случаев исследователю всё-таки необходимо множество описаний свести в несколько групп, причём этот процесс должен быть максимально избавлен от субъективизма и отражать их реальные качественные отличия. Именно для этой цели была разработана группа разнообразных методов, получивших название методов ординации.

Ординация – это собирательное понятие для обозначения многомерных методов обработки данных о связи растительности и условий среды. Она позволяет расположить описания растительности вдоль некоторых осей, опираясь на данные их видового состава, что дает возможность проследить существующие взаимосвязи между экологическими факторами и составом растительности. Кроме того, с помощью методов ординации можно представлять результаты классификации растительности и оценивать взаиморасположения выделенных групп по отношению к факторам среды.

Существует две группы методов ординации: *прямая* и *непрямая*.

Прямая ординация отображает изменение видового состава вдоль некоторого, выбранного исследователем, экологического фактора (влажности, высоты над уровнем моря и т. д.). Эти характеристики могут быть измерены напрямую или определены косвенным путем через использование экологических шкал.

Непрямая ординация показывает изменение видового состава вдоль некоторой абстрактной оси, которая отражает максимальную изменчивость в структуре данных.

К достоинствам прямой ординации можно отнести легкость ее построения и интерпретации осей. Однако, поскольку выбор осей осуществляется вручную, всегда существует вероятность пропустить какой-либо фактор, играющий доминирующую роль. С другой стороны, непрямая ординация позволяет найти оси, максимально влияющие на изменчивость видового состава, но в дальнейшем требуется их интерпретация, т. е. нахождение реальных экологических факторов максимально приближенных к построенным гипотетическим осям, что, к сожалению, не всегда возможно.

Рассмотрим два метода прямой ординации – прямой однофакторный градиентный анализ и прямой многофакторный градиентный анализ.

Суть метода *прямого однофакторного градиентного анализа* заключается в том, что одновременно с геоботаническим описанием площадок измеряется интересующий исследователя фактор среды. По этому фактору ранжируются сделанные описания. Далее все описания группируются по классам выбранного градиента (обычно достаточно 5–7 групп). После группировки по классам и вычисления среднего обилия чертится график, на одной оси которого расположен измеренный фактор среды, по другой обилие вида (рисунок 6).

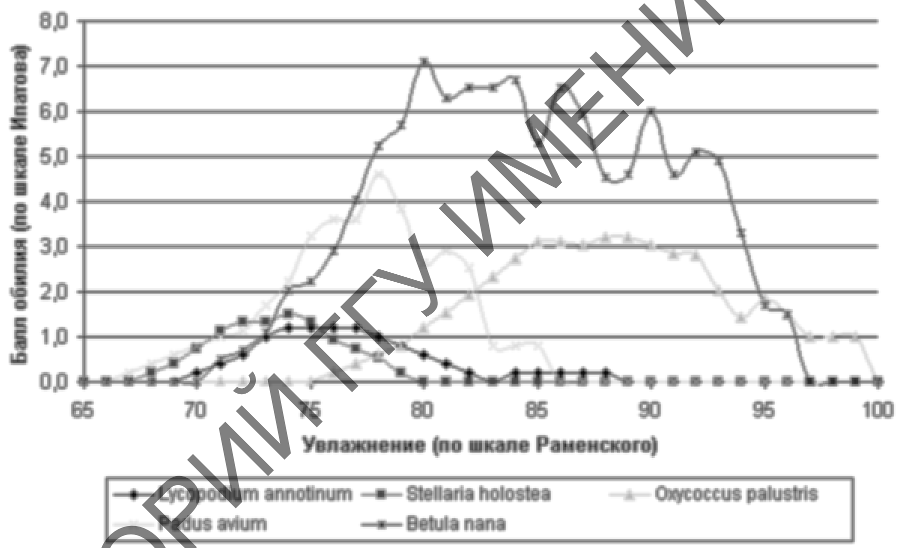


Рисунок 6 – Прямой однофакторный градиентный анализ: изменение обилий растений в связи с изменением фактора увлажнения

Для более эффективного изучения распределения популяций по выбранному градиенту среды необходимо, чтобы описания были схожи по всем другим характеристикам. Например, для изучения влияния высотной поясности на состав растительности надо рассматривать только открытые склоны и только южной экспозиции. Анализ графика позволяет определить, как выбранный фактор влияет на видовой состав растительности для рассматриваемой территории.

На практике часто возникает необходимость рассматривать одновременно несколько экологических факторов. Для изучения подобных сообществ, определяемых более чем одним фактором, применяют *прямой многофакторный градиентный анализ* (обычно двухфакторный), когда на осях откладывают значения различных экологических факторов, которые и являются координатами точек, соответствующих описаниям.

Существенным недостатком методов прямой ординации является то, что нужно заранее знать, какой именно фактор или группа факторов определяет изменчивость видового состава описанных сообществ, что не всегда возможно. Кроме того, эти методы не позволяют проверить, насколько выбранные оси объясняют всю изменчивость растительного покрова. Бывают ситуации, когда значимым фактором может оказаться что-то неучтённое, например, какой-то фактор, который не был замечен исследователем. Изучая зависимость характеристик соснового леса от уровня грунтовых вод, исследователь закладывает пробные площади на территориях с различным его значением. При этом часть площадок может случайно попасть на территорию, характеризующуюся химическим загрязнением почв или воздуха. Этот фактор может ускользнуть от внимания исследователя и оказать существенное искажающее влияние на результаты исследования. Поэтому в ходе дальнейшего развития методов ординации в экологии наряду с прямой ординацией появилась непрямая ординация.

Методы непрямой ординации опираются на видовой состав рассматриваемых описаний, а не на измеренные напрямую или определенные другим образом факторы среды. К ним относятся полярная ординация, метод Чекановского, реципрокное взвешивание, неметрическое многомерное шкалирование и многие другие.

Висконсинская полярная ординация, или *анализ Брея – Кёртиса* является наиболее простым методом. Вначале строится таблица, в которой записываются коэффициенты сходства (КС) и несходства (КН) между описаниями. В качестве индексов сходства можно применять любые, например, Жаккара или Серенсена, индекс несходства вычисляется, как КН = 1 – КС (таблица 15).

Таблица 15 – Пример матрицы коэффициентов сходства и несходства между описаниями (фрагмент)

Номера описаний		1	2	3	4	5	6	7
		Коэффициенты сходства						
1	Коэффициенты	XXXX	0,50	0,89	1,00	0,23	0,52	0,45
2		0,50	XXXX	0,23	0,14	0,36	0,00	1,00
3		0,11	0,77	XXXX	0,00	0,48	1,00	0,88
4		0,00	0,86	1,00	XXXX	0,12	0,58	0,47

5		0,77	0,64	0,52	0,88	XXXX	0,69	0,00
6		0,48	1,00	0,00	0,42	0,31	XXXX	0,00
7		0,55	0,00	0,12	0,53	1,00	1,00	XXXX

Затем выбирают два самых различных сообщества (т. е. сообщества, между которыми имеется максимальный коэффициент несходства). Если максимальное значение этого коэффициента достигается у нескольких пар сообществ, выбирается та пара, у которой максимальна сумма всех значений коэффициентов несходства с остальными описаниями. Эти сообщества определяют противоположные концевые точки первой оси. Координаты каждого описания на этой оси можно вычислить по формуле:

$$C_x = \frac{L_{AB}^2 + L_{AC}^2 - L_{BC}^2}{2L_{AB}^2}.$$

Подобным образом строится вторая (а при необходимости и последующие) ось ординации. Пара концевых точек для второй оси должна отвечать следующим условиям: оба описания должны находиться в средней части первой оси; сходство между этими описаниями должно быть минимальным. После того как выбраны концевые точки второй оси, относительно нее все вычисления повторяются так же, как и для первой оси и т. д. Теперь геоботанические описания могут быть представлены как точки в пределах некоторого пространства осей. Сходные по видовому составу (а следовательно, сформировавшиеся под воздействием сходных значений экологических факторов) местообитания будут образовывать естественные скопления точек, которые можно отделить визуально от других подобных скоплений.

Практическая работа 3
Расчёт значений экологических режимов, сходства и разнообразия сообществ

1. На основе описаний напочвенной растительности различных сообществ сосновых лесов в таблице 16, где указаны значения проективного покрытия видов в процентах, найдите значения экологических режимов данных местообитаний по шкалам Цыганова. Расположите точки, соответствующие описаниям, в прямоугольной системе координат с осями значений по шкалам влаголюбивости и трофности.

Таблица 16 – Исходные данные к практической работе 3

Вид	1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10	11
Pleurozium schreberi (Brid.) Mitt.			80				15		75	
Dicranum undulatum Schrad. ex Brid			1						8	
Cladonia pyxidata (L.) Hoffm.	40							3		50
Polytrichum commune Hedw.				20	13	2	25			
Sphagnum palustre L.					68	85				
Melampyrum sylvaticum L.			1				1		1	

Окончание таблицы 16

1	2	3	4	5	6	7	8	9	10	11
Calluna vulgaris (L.) Hull	1		2							1
Hieracium pilosella L.	1		1							1
Convallaria majalis L.		13		1				13	1	
Hypericum perforatum L.	1							1		
Vaccinium vitis-idaea L.		15	3	30			9	5	5	
Vaccinium myrtillus L.		10	6				45	15	5	
Pteridium aquilinum (L.) Kuhn		30					2	1		
Festuca ovina L.								1		2
Thymus serpyllum L.	1		1					1		
Helichrysum arenarium (L.) Moench	1		1							
Oxalis acetosella L.		20								
Maianthemum bifolium (L.) F.W.Schmidt		1								
Ledum palustre L.				1	3	1				
Vaccinium oxycoccos L.					3	1				

2. Рассчитайте индексы разнообразия сообществ.

3. Проведите анализ Брея – Кёртиса с двумя осями ординации. Нанесите точки, соответствующие сообществам, на диаграмму, согласно полученным координатам. Сравните полученный результат с диаграммой, созданной при выполнении задания 1.

ЛИТЕРАТУРА

1. Атраментова, Л. А. Статистические методы в биологии : учебник для студ. высш. учеб. зав. / Л. А. Атраментова, О. М. Утевская. – Горловка : Лихтар, 2008. – 248 с.

2. Филандышева, Л. Б. Статистические методы в географии : учеб.-метод. пособие / Л. Б. Филандышева, Е. С. Сапьян ; отв. ред. А. В. Пучкин. – Томск : Изд. Дом Томск. гос. ун-та, 2015. – 164 с.

3. Любимов, В. Б. Математические методы в экологии : учеб. пособие / В. Б. Любимов, И. В. Мельников, А. В. Силенок. – Брянск : РИО БГУ, 2017. – 201 с.

4. Крючков, А. В. Биометрия : учеб. пособие / А. В. Крючков, И. В. Маракулин. – Киров : Изд-во ВятГУ, 2011. – 87 с.

5. Соколов, А. С. Методы статистической обработки данных / А. С. Соколов // Геаграфія. – № 2. – 2018. – С. 3–12.

6. Булохов, А. Д. Экологическая оценка среды методами фитоиндикации / А. Д. Булохов. – Брянск : БГПУ, 1996. – 104 с.

7. Булохов, А. Д. Фитоиндикация и ее практическое применение / А. Д. Булохов. – Брянск : Изд-во БГУ, 2004. – 245 с.

8. Цыганов, Д. Н. Фитоиндикация экологических режимов в подзоне хвойно-широколиственных лесов / Д. Н. Цыганов. – М. : Наука, 1983. – 197 с.

9. Пузаченко, Ю. Г. Разнообразие ландшафта и методы его измерения / Ю. Г. Пузаченко, К. Н. Дьяконов, Г. М. Алещенко // География и мониторинг биоразнообразия. – М. : Изд-во НУМЦ, 2002. – С. 76–178.

10. Новаковский, А. Б. Методы ординации в современной геоботанике / А. Б. Новаковский // Вестник Института биологии Коми НЦ УрО РАН. – № 10. – 2008. – С. 3–8.

11. Миркин, Б. М. Словарь понятий и терминов современной фитоценологии / Б. М. Миркин, Г. С. Розенберг, Л. Г. Наумова. – М. : Наука, 1989. – 223 с.

Производственно-практическое издание

Соколов Александр Сергеевич

МЕТОДЫ ОБРАБОТКИ ЭКОЛОГИЧЕСКИХ ДАННЫХ

Практическое пособие

Редактор А. А. Банчук
Корректор В. В. Калугина

Подписано в печать 04.07.2022. Формат 60x84 1/16.
Бумага офсетная. Ризография.
Усл. печ. л. 2,8. Уч.-изд. л. 3,1.
Тираж 10 экз. Заказ 328.

Издатель и полиграфическое исполнение:
учреждение образования
«Гомельский государственный университет имени Франциска Скорины».
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий № 3/1452 от 17.04.2017.
Специальное разрешение (лицензия) № 02330 / 450 от 18.12.2013.
Ул. Советская, 104, 246028, Гомель.

А. С. СОКОЛОВ

МЕТОДЫ ОБРАБОТКИ ЭКОЛОГИЧЕСКИХ ДАННЫХ

Гомель
2022