

МЕТОДИЧЕСКИЕ РАЗРАБОТКИ И РЕКОМЕНДАЦИИ

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

В разделе публикуются методики и рекомендации, имеющие как общеметодологический, так и узкопредметный характер. Материалы этого раздела призваны помочь в практической организации учебного исследования самому широкому кругу воспитателей — как профессиональным педагогам (и школ, и учреждений дополнительного образования), так и родителям.

Продолжаем занятия в школе начинающего исследователя природы.

На занятиях Вы познакомитесь с основами работы в полевых условиях, а также основными правилами анализа и представления исследовательского материала. Вам будет предложена возможность знакомства с различными

Использование математических методов в биологических исследованиях школьников¹

Хайтов Вадим Михайлович,

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

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

Часть 2. Статистические методы Методы сравнения двух величин

Одним из самых распространенных недостатков самостоятельных исследовательских работ школьников оказывается обсужде-

1

ние различий между двумя совокупностями без доказательств достоверности этих отличий.

Рассмотрим типичный пример². Нередко в работах школьников встречается такого рода утверждения: «В экологически неблагоприятном районе частота заболеваний дыхательных путей составляет 50 %, в то время как в более благоприятных областях эта величина не превышает 25 %». Несмотря на высокую разницу между приведенными величинами, любой грамотный исследователь (или чиновник) не примет данное утверждение за доказательство неблагоприятия в первом районе. Прежде всего, остается непонятным, получены ли эти данные путем подсчета всех без исключения людей или они являются результатом выборочного учета (например, результаты работы с карточками некоторых поликлиник).

Поскольку в подавляющем большинстве случаев работа идет с выборками, то приведенные величины должны сопровождаться всеми атрибутами выборочных данных, то есть точечными и интервальными оценками. Без приведения всех этих величин определение уровня заболеваний оказывается абсолютно неубедительным. Рассмотрим данный пример с точки зрения правильного описания.

Прежде всего, необходимо представить данные в виде исходной таблицы. Например, такой:

Таблица 2. Частота заболеваемости дыхательных путей в двух районах³

	Район	
	Экологически неблагоприятный	Экологически чистый
Общее число обследованных людей (объём выборки, N)	260	800
Число людей, у которых отмечены заболевания дыхательных путей	130	200
Встречаемость заболеваний, %	50	25

Нетрудно заметить, что две выборки отличаются по своему объему. В первом случае людей обследовано почти в три раза меньше, чем во втором.

Поскольку мы имеем дело с выборкой и при этом полученные данные имеют качественный характер, то выборочные оценки будут иметь следующий вид:

$$p_1 = \frac{n}{N} \times 100 \% = \frac{130}{260} \times 100 \% = 50 \% \quad \text{и} \quad p_2 = \frac{200}{800} \times 100 \% = 25 \%$$

$$\sigma_{p_1} = \sqrt{p_1 \times (100 - p_1)} = \sqrt{50 \times (100 - 50)} = 50 \quad \text{и}$$

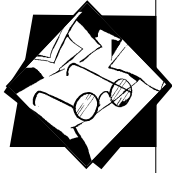
методиками изучения природы и будут даваться задания по их освоению. Материалы занятий будут полезны не только школьникам, но и педагогам, желающим организовать исследовательскую деятельность в своих школах, школьных лесничествах, кружках и клубах. Ведет занятия в школе Николай Харитонов, заведующий отделом Московского городского программно-методического центра дополнительного образования детей (nikol-2@yandex.ru).

2

Почти все примеры, приведенные в данной брошюре, — вымышленные, и автор не уверен, что все полученные при разборе примеров закономерности могут быть обнаружены в реальных условиях.

3

Нумерация таблиц и рисунков сквозная по всей статье с учетом начала в предыдущем номере журнала.



Тема очередного занятия: Математические методы в биологии.
Мы продолжаем тему занятия: математические методы в биологии.

Биология — наука, использующая различные методы представления данных: записи наблюдений, рисунки, численные данные (результаты учетов, измерений, взвешиваний и т.п.). В первой части статьи о математических методах в биологических исследованиях представлены статистические методы.

4

Обратите внимание на то, что при записи результатов после значка «±» стоит одна ошибка. Это не запись доверительного интервала при доверительной вероятности 95%! Это просто краткая форма представления результатов.

$$\sigma_{p_2} = \sqrt{25 \times (100 - 25)} = 43,3.$$

$$m_{p_1} = \frac{\sigma_{p_1}}{\sqrt{N}} = \frac{50}{\sqrt{260}} = 3,1 \text{ и } m_{p_2} = \frac{43,3}{\sqrt{800}} = 1,5.$$

В окончательном виде результаты исследования можно записать в следующем виде: «Встречаемость заболеваний дыхательных путей в экологически неблагоприятном районе составляет $p_1 = 50 \pm 3,1\%$ ($N = 260$ человек), в экологически чистом районе частота заболеваний составляет $p_2 = 25 \pm 1,5\%$ ($N = 800$ человек)⁴». В таком виде результаты исследования будут приняты любым грамотным исследователем. Напомним, что приведенная запись означает, что в генеральной совокупности точное (генеральное) значение параметра $P_{\text{ген}}$ приблизительно равно p , и это точное значение с вероятностью 95% лежит в пределах $\pm 1,96 \cdot m_p$. Так, для случая с экологически неблагоприятным районом с вероятностью 95% $P_{\text{ген}} \in [50 - 1,96 \times 3,1; 50 + 1,96 \times 3,1]$ или $P_{\text{ген}} \in [44,1; 55,9]$. После того, как получены правильные выборочные оценки, можно приступить к решению вопроса о том, достоверны отличия между ними или нет. Иными словами, мы должны решить, отличаются ли генеральные параметры в двух генеральных совокупностях. На языке статистики подобная задача называется тестированием *нулевой гипотезы*.

Нулевая гипотеза, или нуль-гипотеза, — это утверждение, которое звучит так: генеральные параметры двух совокупностей равны. Или в символьном виде:

$$H_0 = P_{\text{ген}1} - P_{\text{ген}2} = 0.$$

Большинство статистических методов направлено на то, чтобы нулевую гипотезу не доказать, а отвергнуть. С чем это связано? Все дело в том, что данные, которыми мы оперируем, — это выборочные данные, которые оценивают генеральные параметры лишь приблизительно. Поэтому всегда останется вероятность того, что мы ошиблись при оценке того или иного параметра. Соответственно строго доказать равенство двух генеральных показателей нельзя, однако можно доказать их различия, то есть отвергнуть H_0 . В связи с этим, вся логика статистического анализа построена не на поиске сходств, а на выявлении различий. Планирование работы надо вести так, чтобы в конечном итоге мы встали **не перед задачей доказательства равенства, а перед задачей доказательства различий выборочных величин**. Если мы сможем отвергнуть нулевую гипотезу, то мы сможем с вероятностью 95% утверждать, что различия между сравниваемыми величинами **статистически значимы, или достоверны**.

Самым простым и наиболее распространенным методом выявления достоверности различий между двумя выборочными оценками можно считать t -критерий Стьюдента. Если не углубляться в теорию, то использование этого критерия сводится к расчету следующей величины:

$$t = \frac{|p_1 - p_2|}{\sqrt{m_{p_1}^2 + m_{p_2}^2}}.$$

В приведенном выше примере вычисления будут иметь следующий вид:

$$t = \frac{|50 - 25|}{\sqrt{3,1^2 + 1,5^2}} = \frac{25}{\sqrt{11,86}} = 7,26.$$

После того как вычислено эмпирическое значение t , его необходимо сравнить с табличным, пороговым значением $t_{\text{табл}}$. Это значение находится по специальным таблицам (см. табл. 1), и для правильного его нахождения необходимо иметь в распоряжении еще две величины. Первая величина — это уже знакомая **доверительная вероятность** ($P_{\text{дов}}$). Вычислять ее не надо, надо лишь выбрать, какая степень достоверности результата нас устроит. Напомним, что в биологии обычно считается, что в качестве такой вероятности следует выбирать 95 %. Это означает, что мы на 95 процентов уверены, что различия между выборочными показателями неслучайны и отражают некоторые реально существующие различия. Лишь 5 % остается на то, что мы ошибаемся в нашем выводе.

Вторая величина требует вычислений. Это так называемое **число степеней свободы**. Оно определяется по следующей формуле⁵:

$$v = N_1 + N_2 - 2.$$

В нашем примере $n = 260 + 800 - 2 = 1058$. Табличное значение при $n = 1058$ (в таблице такая большая величина обозначается как ∞) и доверительной вероятности 95 % равно $t_{\text{табл}} = 1,96$. Сравнение эмпирического значения t и табличного показывает, что $t > t_{\text{табл}}$. Если эмпирическое значение t превышает табличное, то это означает, что нулевую гипотезу можно с вероятностью 95 % (и выше) отвергнуть. То есть, различия между выборочными показателями оказываются **достоверными**, иными словами, мы можем быть практически уверенными (на 95 %), что эти **различия неслучайны и отражают реальную закономерность** — в экологически неблагополучном районе частота заболеваемости дыхательных путей выше.

Если бы при наших расчетах получилось, что $t < t_{\text{табл}}$, то это означало бы, что различия **недостоверны** — нуль-гипотеза не может быть отвергнута. Особо отметим, что если мы не смогли отвергнуть нулевую гипотезу, то это не означает, что величины равны, это означает, что различия не доказаны!

Приведенный пример был основан на качественных данных. Вместе с тем сравнение выборочных величин можно проводить и на основе количественных показателей.

Рассмотрим другой пример. Предположим, что в процессе исследования мы изучили длину соцветий лисохвоста в двух разных популяциях, одна из которых подвержена выпасу скота, а другая —

В этом номере журнала описываются методы сравнения двух величин, методы анализа структуры популяции. В следующем номере будут представлены описание взаимосвязи величин и методы многомерного анализа.

5

Определение числа степеней свободы в разных статистических анализах ведется по-разному.



нет. При этом мы хотим выяснить, влияет ли выпас коров на рост лисохвоста. Для этого мы случайным образом отобрали несколько растений в двух разных популяциях и измерили у них длину соцветий. Как и в предыдущем случае, мы должны первым делом составить исходную таблицу.

Таблица 3. Размеры соцветий в двух популяциях лисохвоста

Длина соцветия, см	
Выпас есть	Выпаса нет
5,7	4,6
4,1	9,6
13,6	14,1
4,4	5,0
3,3	9,1
9,8	16,6
7,2	9,5
9,3	8,6
12,7	10,8
6,3	10,6
$\bar{x}_1 = 7,6$	$\bar{x}_2 = 9,8$
$\sigma_1 = 3,60$	$\sigma_2 = 3,62$
$N_1 = 10$	$N_2 = 10$
$m_{x_1} = 1,13$	$m_{x_2} = 1,13$



Здесь — среднее значение, σ — среднее квадратичное отклонение, N — объем выборки (количество измеренных колосков), m_x — ошибка среднего. Как и в предыдущем случае, бросается в глаза, что на поле, где нет выпаса, колоски в среднем длиннее, поскольку $x_2 > x_1$. Но давайте вспомним, что нас интересует разница не выборочных значений, а разница генеральных параметров! То есть, нуль-гипотеза будет иметь следующий вид: $H_0 = X_1 - X_2 = 0$. Для тестирования этой гипотезы воспользуемся t -критерием Стьюдента. Для сравнения средних он будет иметь следующий вид:

$$t = \frac{d}{m_d},$$

$$d = |\bar{x}_1 - \bar{x}_2|,$$

$$m_d = \sqrt{m_{x_1}^2 + m_{x_2}^2}.$$

Для нашего случая вычисления будут выглядеть так:

$$d = |7,6 - 9,8| = 2,2,$$

$$m_d = \sqrt{1,13^2 + 1,13^2} = 1,60,$$

$$t = \frac{2,2}{1,6} = 1,34.$$

Как и в случае с анализом различий встречаемости, необходимо сравнить эмпирическое значение t и табличное пороговое значение $t_{\text{порог}}$. Для поиска табличного значения надо, во-первых, выбрать доверительную вероятность. Возьмем ее равной 95 %. Во-вторых, надо определить число степеней свободы:

$$v = N_1 + N_2 - 2 = 10 + 10 - 2 = 18.$$

Табличное значение при доверительной вероятности 95 % и $n = 18$ составляет $t_{\text{порог}} = 2,10$. Мы видим, что эмпирическое значение t меньше табличного, порогового. Это означает, что у нас нет оснований отвергнуть нулевую гипотезу. Еще раз напоминаем, что это не означает, что разницы нет, а означает, что разница не доказана.

Таким образом, в нашем примере мы не смогли доказать, что выпас скота влияет на длину соцветий лисохвоста. При этом вполне вероятно, что, увеличив объем выборки (а в нашем случае он чрезвычайно мал), мы выявим достоверные отличия. Вот тогда мы сможем с уверенностью на 95 % утверждать, что выпас коров приводит к уменьшению длины соцветия.

В заключение этой части надо отметить, что при сравнении выборочных средних t -критерий имеет некоторые особенности, о которых мы еще не говорили. Главная особенность заключается в том, что если объемы двух выборок сильно отличаются (т. е. $N_1 > N_2$ или $N_2 > N_1$), то расчеты надо вести немного иначе. Формулы будут такие.

$$d = |x_1 - x_2|,$$

$$m_d = \sqrt{\frac{\sigma_1^2 \cdot (N_1 - 1) + \sigma_2^2 \cdot (N_2 - 1)}{N_1 + N_2 - 2}} \cdot \sqrt{\frac{N_1 + N_2}{N_1 \cdot N_2}},$$

$$t = \frac{d}{m_d}.$$

То есть, различия будут в методе расчета m_d . В остальном же применение этого метода будет абсолютно таким же, как описано выше.

А как сравнить выборки, если в вашем распоряжении есть только балловые данные? Рассмотрим такой пример. Предположим, вы задались целью сравнить две популяции черных кошек, допустим, в центре города и за городом, по степени выраженности у них белых пятен. При этом вы дали следующие балловые характеристики:





Балл	Описание
0	Нет белых пятен, чисто черное животное
1	Одно небольшое белое пятно
2	Несколько небольших белых пятен
3	Одно крупное белое пятно
4	Несколько крупных белых пятен

Таблица 4. Балловые характеристики окраски кошек

Обратите внимание, что здесь мы имеем дело именно с балловыми оценками, а не просто с цифровым обозначением разновидностей. В нашем примере чем больше балл, тем меньше черного в окраске кошки. Мы специально остановились на этом, так как иногда, вводя числовые обозначения, люди попадают под магию цифр и пытаются применить математику там, где ее не должно быть (или применяют не ту математику).

Далее вы собрали материал и оформили его в виде таблицы.

кошек в разных районах

Наличие цифрового материала в данной таблице сразу вызывает у нас соблазн применить простой метод сравнения средних, который был описан выше. Действительно, чего проще рассчитать средний балл и сравнить эти средние баллы. Давайте попробуем. Для городской популяции средний балл составляет 2,19. Можно округлить до 2. Казалось бы, очень просто и логично. Но давайте задумаемся, а что отражает этот средний балл? Честно говоря, он не отражает ничего! Помните, что средняя величина оценивает генеральный параметр распределения χ . А согласно свойству нормального распределения величины, близкие к χ , должны встречаться с максимальной вероятностью

Предположим, такой:

Таблица 5. Результаты учета

В центре города	За городом
0	0
0	0
0	1
0	1
0	1
1	1
1	2
1	2
4	2
4	2
4	3
4	3
4	4
4	4
4	4
4	4
4	4
4	4
4	4
Всего кошек: 16	Всего кошек: 20



6

См. начало статьи в ИРШ.
2008. №1. С. 53.

30

(см. рис. 2)⁶. В нашем же случае, в выборке кошек из центра города, не было отмечено ни одной особи с баллом 2. Стало быть, вычислив среднее значение балла, мы получили откровенную ложь! К сожалению, очень часто балловые оценки подвергаются некорректному статистическому анализу. С баллами надо быть очень аккуратными, далеко не всегда вычисление средних баллов дает объективную информацию о системе.

Как же нам поступить в данном случае? Самый простой способ — это воспользоваться так называемым критерием χ^2 (читается *хи-квадрат*)⁷. Этот критерий служит для сравнения двух распределений. Помните, мы определили распределение как связь между величиной и вероятностью встречи данной величины в генеральной совокупности. Однако, когда мы работаем с выборкой, мы тоже можем говорить о распределении, но уже не о генеральном, а о *выборочном распределении*. Выборочное распределение уже нельзя точно описать какой-то функцией, однако его можно представить в виде так называемого *вариационного ряда*. Обычно вариационный ряд представляют в виде таблицы из двух колонок. В первой колонке выписаны значения признака, который мы анализируем, а во второй — число объектов, обладающих данным признаком, которые попали в выборку. Для нашего примера вариационные ряды будут иметь следующий вид.

Таблица 6. Вариационные ряды для черных кошек из двух популяций

Балл	В центре города	За городом
0	5	2
1	3	4
2	0	4
3	0	2
4	8	8
Всего:	16	20

Когда вариационные ряды составлены, можно приступить к их сравнению. Формула для вычисления критерия χ^2 в данном случае имеет следующий вид:

$$\chi^2 = \frac{1}{N_1 \times N_2} \times \sum \frac{(f_{i1} \times N_2 - f_{i2} \times N_1)^2}{f_{i1} + f_{i2}}$$

В этой формуле N_1 и N_2 — объемы выборок, f_{i1} и f_{i2} — количество объектов, обладающих i -м признаком в первой и во второй совокупностях соответственно. Для вычисления χ^2 проще всего построить такую таблицу:

7

На самом деле, существует целый раздел науки, который называется «непараметрическая статистика». Непараметрические методы нужны тогда, когда тип распределения величины в генеральной совокупности неизвестен или заведомо отличается от нормального. В такой ситуации сравнение выборочных параметров x и s некорректно. Здесь нужны непараметрические критерии достоверности. В данном издании мы коснемся лишь незначительной их части.

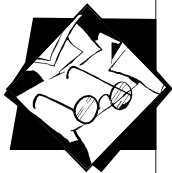


Таблица 7. Ход вычисления χ^2

Балл	В центре города (f_1)	За городом (f_2)	$f_{i1}N_2$	$f_{i2}N_1$	$f_{i1}N_2 - f_{i2}N_1$	$(f_{i1}N_2 - f_{i2}N_1)^2$	$\frac{(f_{i1} \cdot N_2 - f_{i2} \cdot N_1)^2}{f_{i1} + f_{i2}}$
0	5	2	100	32	68	4624	660,6
1	3	4	60	64	-4	16	2,3
2	0	4	0	64	-64	4096	1024,0
3	0	2	0	32	-32	1024	512,0
4	8	8	160	128	32	1024	64,0
	$N_1=16$	$N_2=20$					$\Sigma = 2262,9$

Далее по формуле вычисляем: $\chi^2 = \frac{1}{16 \cdot 20} \times 2262,6 = 7,07$.

Итак, искомое значение критерия получено. Теперь, как и в предыдущих случаях, надо сравнить это значение с табличным (табл. 2). Для поиска табличного, порогового, значения нам опять необходимы две величины — доверительная вероятность и число степеней свободы.

Доверительная вероятность выбирается нами в соответствии с нашими представлениями о степени надежности заключения. Если нас устраивает, что надежность заключения 95 % (то есть с вероятностью 5 % мы ошибаемся), то мы должны выбрать именно эту доверительную вероятность. В принципе, мы можем согласиться и на менее надежные выводы, взяв, например, доверительную вероятность 90 % (но тогда и табличное значение будет другим).

Второй величиной, необходимой нам для поиска порогового значения χ^2 в таблице, как и в предыдущих случаях, оказывается число степеней свободы. В данном случае оно вычисляется по следующей формуле: $n = K - 1^8$. Здесь K — это число классов. В нашем случае число классов определяется количеством баллов. Всего для анализа материала мы использовали 5 баллов, следовательно, число степеней свободы будет: $\nu = 5 - 1 = 4$. Итак, мы выбрали 95 % доверительную вероятность и число степеней свободы равно 4. Этому сочетанию в табл. 2 соответствует $\chi^2_{\text{порог}} = 9,49$. Сравнение табличного и эмпирического значений χ^2 говорит о том, что эмпирическое значение ниже табличного. Это, как и в случае с t -критерием Стьюдента, говорит о том, что различия между двумя вариационными рядами недостоверны. Полученные данные позволяют сделать вывод о том, что у нас нет оснований для утверждения, что черные кошки за городом отличаются от черных кошек в центре города. Весьма вероятно, что при

8

Внимание! В других случаях применения критерия χ^2 формулы для вычисления числа степеней свободы будут другими!

большем объеме выборок, если бы мы проанализировали окраску не десятков, а сотен животных, мы выявили бы достоверные отличия. Но пока эти отличия не доказаны, мы не имеем права делать вывод о том, что эти две популяции отличаются друг от друга. Заметим, что мы не имеем права делать и вывод о том, что эти популяции не отличаются!

Между прочим, пример с черными кошками позволяет поставить и другой вопрос. Например, такой: «Где больше вероятность встретить чисто черную кошку в городе или за городом?» Поскольку вопрос задан по-другому, то и метод анализа будет другим! В данном случае мы должны сравнить встречаемость чисто черных кошек среди черных кошек вообще в городе и за городом. А это уже задача, решаемая с помощью t -критерия Стьюдента. Попробуйте решить эту задачку сами.

Методы анализа структуры популяции

Прежде чем мы приступим к разговору о собственно методах анализа структуры популяции, необходимо договориться о некоторых базовых понятиях популяционной биологии.

Под популяцией мы будем понимать группу особей одного вида, населяющих определенную территорию. Заметьте, что для применения статистических методов нам нет нужды вводить представление о том, что эти особи свободно скрещиваются друг с другом.

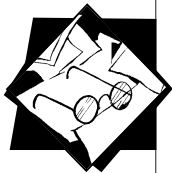
Структуру популяции мы будем понимать как некоторую неоднородность особей, входящих в состав популяции. Например, под половой структурой мы будем понимать соотношение числа особей мужского и женского полов.

Если теперь перейти на язык статистики, то вся популяция для нас будет генеральной совокупностью. Она неоднородна по каким-то признакам. Структуру генеральной совокупности мы будем пытаться анализировать с помощью небольшой выборки особей из этой популяции.

Структуру популяции можно изучать с разных точек зрения. Наиболее часто изучаются следующие аспекты.

1. Половая структура популяции — все особи в популяции различаются по своему полу.
2. Фенетическая структура популяции — особи различаются по некоторым дискретным внешним признакам — фенам, например, по окраске тела.
3. Генетическая структура популяции — особи в популяции рассматриваются как носители того или иного аллеля.
4. Размерно-возрастная структура популяции — особи различаются по своим размерам и возрасту.

Очень часто в первых трех случаях главной задачей оказывается сравнение эмпирического и теоретического распределения. С рассмотрения решения этой задачи мы и начнем.



Предположим, что мы задались целью проанализировать половую структуру популяции мышей в чулане вашего дома. Для этого анализа необходимо сделать выборку особей. Конечно, для дома было бы полезней произвести анализ всей генеральной совокупности, по крайней мере, всех мышей бы извели, но, предположим, что сделать это невозможно, допустим, грызунов слишком много. Итак, вам предстоит сделать выборку. Как это сделать технически, решайте сами, но, так или иначе, зверьков надо отловить и определить у них пол. Допустим, вы поймали 80 мышей. Из них 15 оказались самцами, а 65 самками. Результаты вашей охоты сразу надо представить в виде исходной таблицы (лучше себя сразу приучить грамотно оформлять результаты исследования).

Таблица 8. Соотношение полов в изученной популяции мышей

Число самцов	Число самок
15	65
Всего мышей: 80	



Результаты анализа выборки показывают, что число самцов не равно числу самок. Можем ли мы, исходя из полученных данных, сделать вывод о том, что в нашей популяции самцов меньше, чем самок. Конечно же, нет! Пока нет. Нам необходимо доказать это. Для проверки гипотезы о том, что самцов меньше, чем самок, надо доказать, что наблюдаемое соотношение полов достоверно отличается от соотношения 1:1, при котором количество самцов равно количеству самок. Как это сделать? Для этого используют уже знакомый нам критерий χ^2 , только в несколько иной модификации. В данной ситуации нам необходим критерий для сравнения реального (эмпирического) и теоретического распределения.

Напомним, что все статистические критерии призваны тестировать нулевую гипотезу. В нашей ситуации нулевая гипотеза будет сформулирована так: $H_0 = f_{\text{ген}} - f_{(1:1)} = 0$, или в словесном выражении она будет звучать так: «В генеральной совокупности соотношение численностей разных полов подчиняется соотношению 1:1».

Первый шаг к применению критерия χ^2 — вычисление *теоретически ожидаемых частот*, или, что то же самое, **построение теоретического распределения**. Выражение, выделенное жирным шрифтом, несмотря на непривычное звучание, имеет очень простой смысл. Что такое распределение, мы уже знаем. Под **частотой** подразумевается количество объектов, обладающих данным признаком. Мы с этим уже тоже встречались, когда говорили о вариационном ряде. Что такое «**теоретически ожидаемые частоты**»? Тоже очень просто! Это означает «сколько **было бы** особей с теми или иными признаками, **если бы** популяция обладала структурой такой, какую мы предполагаем».

Для нашего случая мы предположили, что соотношение самцов и самок должно быть 1:1. Стало быть, теоретически частота самцов должна быть равна $\frac{1}{2}$ всех пойманных мышей. Тогда теоретическая частота самцов должна быть 40, и самок — тоже 40. Оформим результаты в виде таблицы.

Таблица 9. Реальные и теоретические частоты самцов и самок мышей

	Самцы	Самки
Реальная частота	15	65
Теоретическая частота	40	40

Теперь пришло время воспользоваться критерием χ^2 . Для данного типа задач применяется другая формула:

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

Здесь E — фактически наблюдаемая (эмпирическая) частота, T — теоретическая частота. Произведем расчеты, занося результаты в таблицу.

Таблица 10. Ход вычисления критерия χ^2

	Самцы	Самки
Реальная частота (E)	15	65
Теоретическая частота (T)	40	40
$E - T$	-25	25
$(E - T)^2$	625	625
$\frac{(E - T)^2}{T}$	15,63	15,63

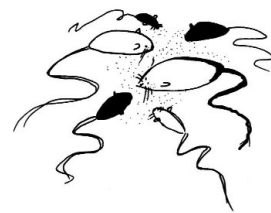
$$\chi^2 = \sum \frac{(E - T)^2}{T} = 15,63 + 15,63 = 31,26.$$

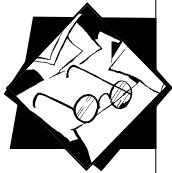
Итак, вычисленное значение $\chi^2 = 31,26$. Теперь, как и в предыдущих случаях, надо сравнить эту величину с табличным значением. Как всегда, надо использовать две величины: доверительную вероятность ($P_{\text{дов}}$) и число степеней свободы (ν).

Как обычно, мы выбираем доверительную вероятность $P_{\text{дов}} = 95\%$. Для вычисления числа степеней свободы используется следующая формула:

$$\nu = K - 1,$$

где K — число классов. Таким образом, в нашем случае $\nu = 2 - 1 = 1$. По таблице значений χ^2 находим пороговое значение для $P_{\text{дов}} = 95\%$ и $\nu = 1$, оно равно $\chi^2_{\text{порог}} = 3,84$. Сравнение табличного и вычисленного значений χ^2 показывает, что $\chi^2 > \chi^2_{\text{порог}}$.





Это говорит о том, что наблюдаемое соотношение полов с вероятностью 95 % отличается от соотношения 1:1. Кстати сказать, если мы возьмем для данного примера доверительную вероятность $P = 99 \%$, то $\chi^2_{\text{порог}} = 6,63$. То есть, даже при столь высокой доверительной вероятности, реально наблюдаемое соотношение полов достоверно отличается от теоретического. Это позволяет говорить о том, что даже с вероятностью 99 % наблюдаемое распределение не соответствует соотношению 1:1.

Итак, мы доказали, что в популяции мышей самцов меньше, чем самок. На этом статистика свое дело закончила. Теперь надо найти причину, по которой самцов меньше. Может быть, самцы менее осторожны и кошки интенсивнее их вылавливают, а может, существует другая причина. Поиск причин доказанных различий — это самое увлекательное в биологии, но тут уж дать какого-то общего совета или алгоритма действий нельзя.

Разберем теперь другой пример, несколько более сложный, когда в распределении присутствуют не два класса, а много. Давайте попробуем повторить опыт Грегора Менделя и посмотреть на его результаты со статистической точки зрения. Вы помните, что Мендель анализировал наследование двух признаков семян гороха: цвета и формы семени. Во втором гибридном поколении у него получилось следующее соотношение семян по фенотипам: 9 частей желтых гладких семян, 3 части желтых морщинистых, 3 части зеленых гладких и 1 часть зеленых морщинистых.

Предположим, что на своем дачном участке вы решили провести аналогичные исследования. Вы проделали те же самые опыты и, собрав урожай от второго гибридного поколения, занялись подсчетами различных типов семян. Полученные данные вы тут же записали в исходную таблицу.

Таблица 11. Число семян с разными фенотипами, полученных от растений второго гибридного поколения

Цвет и форма	Число семян
Желтые гладкие	311
Желтые морщинистые	113
Зеленые гладкие	102
Зеленые морщинистые	40
Всего:	566

Теперь необходимо провести вычисление теоретически ожидаемых частот. Мы будем предполагать, что теоретически ожидаемое расщепление соответствует менделевскому 9:3:3:1. Как найти теоретические частоты для нашего случая? Для этого надо понять, сколько всего частей должно быть представлено в совокупности, если работает соотношение 9:3:3:1. Таких частей должно быть 16



(9 + 3 + 3 + 1 = 16). Если частей 16, то можно вычислить, сколько горошин составит одну часть из 566 штук. Для этого делим общее количество горошин на теоретически ожидаемое число частей: $566/16 = 35,375 \sim 35,4$. Получилось дробное число (многие школьники этого очень боятся)! Однако этого не надо смущаться, нужно лишь округлить конечный результат.

Теперь можно вычислять и теоретические частоты. Если работает соотношение 9:3:3:1, то желтых гладких семян мы должны получить $35,4 \times 9 = 319$, желтых морщинистых $35,4 \times 3 \sim 106$, столько же и зеленых гладких, а зеленых морщинистых должно быть $35,4 \times 1 \sim 35$ (они, в соответствии с менделевским законом, составляют одну часть). Внимание! При округлении необходимо следить за тем, чтобы суммы эмпирических и теоретических частот были равны. В нашем случае данное условие выполнено. Сведем все это в таблицу.

Таблица 12. Эмпирическое и теоретическое распределение по фенотипам семян гороха

	Число семян эмпирическое (E)	Число семян теоретическое (T)
Желтые гладкие	311	319
Желтые морщинистые	113	106
Зеленые гладкие	102	106
Зеленые морщинистые	40	35

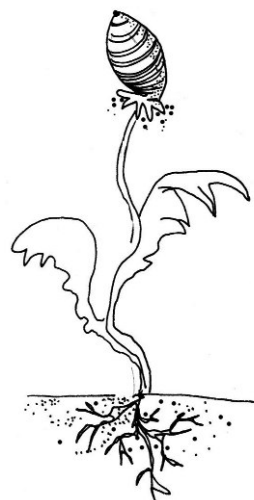
Теперь все готово для вычисления критерия χ^2 , которое ведется по тем же формулам, что и в предыдущем случае.

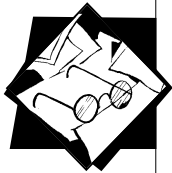
Таблица 13. Ход вычисления критерия χ^2

	E	T	E - T	(E - T) ²	$\frac{(E - T)^2}{T}$
Желтые гладкие	311	319	-8	64	0,2006
Желтые морщинистые	113	106	7	49	0,4623
Зеленые гладкие	102	106	-4	16	0,1509
Зеленые морщинистые	40	35	5	25	0,7143

$$\chi^2 = \sum \frac{(E - T)^2}{T} = 1,528.$$

Как всегда, далее производится сравнение с табличным значением. Число степеней свободы, как и в предыдущем случае, $\nu = K - 1 = 4 - 1 = 3$. При доверительной вероятности $P_{\text{дов}} = 95\%$ табличное значение $\chi^2_{\text{порог}} = 7,81$. Мы видим, что $\chi^2 < \chi^2_{\text{порог}}$. Что это означает? Вы уже можете точно сказать: «Это значит, что **отличия недостоверны**». Стало быть, у нас нет оснований для





утверждения, что в нашем эксперименте наблюдаются существенно другие результаты, нежели предсказывает теория. Но очень важно отметить, что полученные данные ни в коей мере не доказывают, что в нашем случае имеет место расщепление 9:3:3:1. Помните, что доказать равенство мы с помощью статистики не можем! То, что мы не получили достоверных отличий от теоретического соотношения, говорит лишь о том, что мы не отвергли нулевую гипотезу и можем ее использовать как **рабочее предположение**. Всегда надо помнить, что при большем объеме выборки могут быть получены другие результаты, только вероятность этого очень низка. Вот такая упрямая штука — логика статистики!

Теперь перейдем к разговору об анализе размерно-возрастной структуры популяции. С точки зрения размерно-возрастного анализа все особи в популяции отличаются либо по своему возрасту, либо по размеру. Поскольку этот тип анализа структуры популяции имеет свои тонкости чисто биологического плана, то на них и надо остановиться.

Как правило (хотя не всегда), исследователя интересует прежде всего возраст особей. Однако в большинстве случаев возраст измерить крайне трудно. Самой благоприятной оказывается ситуация, когда у организмов имеется какая-то часть тела, на которой этот возраст можно «прочитать». Например, на спиле дерева видны кольца нарастания, на чешуе некоторых рыб и раковинах моллюсков видны кольца зимней остановки роста, у лошадей возраст можно определить по степени изношенности зубов. Однако такого рода признаки встречаются крайне редко и далеко не любой объект можно изучить подобным способом. Более того, если вам посчастливится поработать в тропиках, то там вследствие малозаметности сезонных явлений вы таких признаков не найдете. Поэтому гораздо чаще исследователь исходит из анализа не возраста особей, а их размера. При этом предполагается, что чем старше особь, тем больше ее размер. Это, как не трудно заметить, тоже не всегда справедливо, поскольку есть виды, у которых с возрастом рост замедляется или вовсе останавливается. Однако и это препятствие можно обойти, изучая, например, не рост, а вес организмов или и то, и другое. В любом случае, анализируя популяцию того или иного вида, лучше использовать все доступные для измерения признаки (возрастные признаки, размер тела и отдельных его частей, вес организма и т. п.). Однако при этом не стоит увлекаться, так как анализ большого числа признаков на одной особи может занять очень много времени и вы из-за трудоемкости не сможете обработать достаточно большую выборку.

Итак, вы приступили к анализу размерно-возрастной структуры популяции. Какова ваша цель? Главной целью такого рода исследований оказывается выявление размерно-возрастных групп и определение соотношения их численностей в популяции.



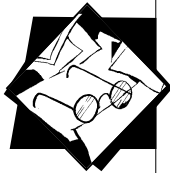
Разберем ход анализа на конкретном примере. В таблице приведены результаты измерения длины тела (L) в мм и веса (P) в мг рачков-бокоплавов *Pontoporeia femorata*, отловленных в мелководном заливе Белого моря.

Таблица 14. Длина тела и вес *Pontoporeia femorata*

L	P		L	P		L	P		L	P
5,7	4		4,6	4		5,2	4		5	6
6,6	10		6,2	7		5,3	6		6,2	11
7,3	11		5,8	7		6,4	7		5,6	6
4,7	5		4,4	3		6,2	8		11,6	64
5,5	6		5,1	5		6,4	6		13,4	64
5,4	5		6,4	6		5,7	6		5,2	10
6,2	10		4,4	5		6,2	10		6,1	7
6,3	11		10,3	42		5,1	4		6,4	7
10,5	32		12,1	63		5,5	5			
13,5	56		12,3	48		6,7	8			
6,3	10		10,8	39		5,8	10			
5,9	7		5,1	6		6,8	9			
6,2	6		5,2	5		9,5	27			
5,8	6		5,7	7		5,4	6			
4,3	4		5,6	7		11,2	41			
4,7	3		4,2	2		5,2	5			
5,1	4		4,9	4		6	6			
5,2	7		4,3	3		5,9	8			
5,8	6		4,2	3		4,9	3			
4,7	4		5	4		5,2	4			
5,6	6		15,5	98		6	6			
4,3	3		13,4	68		6,5	10			
5,8	7		13,2	79		6,1	6			
4,4	2		13,3	60		5,3	6			
4,3	5		6,1	8		6,3	10			
9,2	22		4,6	3		6,9	10			
4,2	3		6,5	10		6	10			
4,4	3		4,8	5		4,9	6			
6,7	8		6,1	7		5,7	10			

ВСЕГО: 95 особей.





Приведенные данные и являются основой для первого шага изучения размерно-возрастной структуры популяции — *частотного анализа*⁹.

Итак, необходимо провести частотный анализ. Суть этого анализа сводится к построению уже известного нам выборочного распределения, т.е. к построению вариационного ряда. Однако, если в предыдущих случаях построение вариационного ряда сводилось к подсчету количества особей относящихся к четко отличимым, как говорят, дискретным классам, то в данной ситуации мы имеем непрерывную величину. Как поступить в этой ситуации? Можно, конечно, поступить просто, например, решить, что в качестве классов мы будем использовать значения конкретных измерений. То есть, в выборке мы будем подсчитывать число объектов, имеющих каждое конкретное значение измеренного признака. В нашем случае это приведет к следующему виду вариационного ряда (для простоты пока остановимся на анализе только длины тела):

Таблица 15. Вариационный ряд при классовых значениях, равных отдельным измерениям

L, мм	Частота (число особей)
4,2	3
4,3	4
4,4	4
4,5	2

и т. д.

Не будем продолжать этот ряд, поскольку в данном случае так его строить неудобно и не совсем правильно. Почему? Во-первых, таких классов получится очень много. Нетрудно подсчитать, что если минимальное значение длины составляет 4,2, а максимальное 15,5, то число классов будет равно 113 (!). При незначительном числе особей большое число классов будет пустым. Во-вторых, и это очень важно, когда производятся измерения с помощью грубых приборов, при таком дробном вариационном ряде начинают играть роль погрешности измерения и округления величин. Наши глаза несовершенны, и при измерениях одного и того же объекта разными людьми очень часто получаются немного разные результаты. Поэтому чрезмерная дробность при частотном анализе вредна, поскольку становятся видны посторонние «шумы», не имеющие никакого отношения к реальной структуре популяции.

Вследствие описанных причин частотный анализ лучше проводить несколько иначе. Для этого разделение на классы ведут в соответствии с выбранным *классовым шагом*. Вся шкалу варьирования признака разбивают на несколько *интервалов*. Как выбрать классовый шаг? Можно сделать это чисто формально. В статистике есть специальная формула:

9

Частотный анализ — метод очень простой и вместе с тем очень мощный, он позволяет понять структуру ваших данных. Поэтому, когда в вашем распоряжении много числовых данных (не важно, для каких целей вы собираетесь их использовать), лучше сначала провести их частотный анализ.

$$i = \frac{L_{\max} - L_{\min}}{1 + 3,32 \lg(N)}$$

Здесь i — величина классового шага, L_{\max} — максимальное значение признака, L_{\min} — минимальное значение признака, \lg — знак десятичного логарифма, N — объем выборки.

Вычислим значение классового шага для данных по размерам *Pontoporeia femorata*.

$$i = \frac{15,5 - 4,2}{1 + 3,32 \lg(95)} = 1,49 \approx 1,5.$$

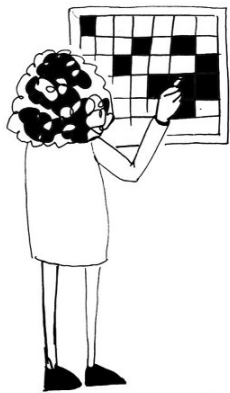
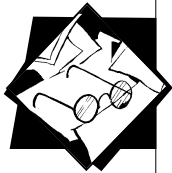
После того как получена величина классового шага, можно приступить к разбивке всей шкалы варьирования признака на интервалы. Первая нижняя граница интервала определяется как L_{\min} , а первая верхняя граница — как $L_{\min} + i$, следующая нижняя граница — это верхняя граница предыдущего интервала. В данном случае мы получим следующие интервалы.

4,2—5,7
5,7—7,2
7,2—8,7
8,7—10,2
10,2—11,7
11,7—13,2
13,2—14,7
14,7—16,2
16,2—17,7

Однако предложенный метод разбивки на интервалы удобен только тогда, когда популяция подвергается разовому анализу и далее она ни с какой другой популяцией сравниваться не будет. Если же вы решили проследить размерную структуру популяции на протяжении нескольких лет, то вполне может сложиться ситуация, при которой объемы выборок в разные годы будут отличаться или изменится размах варьирования размеров. Это приведет к тому, что классовые шаги в разных выборках будут разными, а стало быть, произойдет и разбивка на разные интервалы. Это сделает сравнение невозможным. Поэтому чаще биологи выбирают для каждого вида некоторый свой единый, стандартный шаг. Так, например, опыт показал, что для *Pontoporeia femorata* наиболее удобным оказывается классовый шаг, равный 1 мм. Выбор классового шага, таким образом, очень часто происходит в соответствии с интуицией исследователя и знанием особенностей варьирования признака у изучаемого объекта и точности измерений. Пока у вас не появился опыт, лучше пользоваться приведенной формулой. Однако для дальнейших построений мы все-таки воспользуемся классовым шагом, равным 1 мм. Тогда мы получим следующие интервалы:

4,0—5,0
5,0—6,0
6,0—7,0





7,0–8,0
8,0–9,0
9,0–10,0
10,0–11,0
11,0–12,0
12,0–13,0
12,0–13,0
13,0–14,0
14,0–15,0
15,0–16,0
16,0–17,0

Теперь необходимо подсчитать количество особей, чьи размеры попадают в соответствующий интервал. Это и будет вариационный ряд. Однако здесь необходимо сделать одно очень важное замечание. Нетрудно заметить, что некоторые значения размеров могут быть равны одновременно и верхней границе одного интервала и нижней границе следующего. В такой ситуации для распределения конкретных величин по классам используют очень простое правило: **значение, равное «граничному значению», помещается всегда справа от границы.** Так, например значение равное 10,0 попадет в класс 10,0–11,0, а значение 5,0 — в класс 5,0–6,0.

Обратите внимание на то, что в данном вариационном ряду мы начали «отсчет» от нуля. В принципе, не важно, откуда начинать, но для удобства восприятия иногда полезно начинать от начала координат (хотя иногда это может и затруднить восприятие, например, когда значения признака очень сильно отличаются от нуля, а величина классового шага очень мала).

Итак, вариационный ряд построен. Что делать дальше? Следующим ходом в анализе размерно-возрастной структуры популяции будет построение *частотной гистограммы* или *частотного полигона*.

Таблица 16. Вариационный ряд размеров тела *Pontoporeia femorata*

<i>L</i> , мм	Частота, экз.
0,0–1,0	0
1,0–2,0	0
2,0–3,0	0
3,0–4,0	0
4,0–5,0	22
5,0–6,0	33
6,0–7,0	24
7,0–8,0	1
8,0–9,0	0
9,0–10,0	2
10,0–11,0	3
11,0–12,0	2
12,0–13,0	2
13,0–14,0	5
14,0–15,0	0
15,0–16,0	1
16,0–17,0	0

Построить гистограмму или полигон очень просто. Необходимо изобразить на листе бумаги координатные оси (для данного случая достаточно лишь положительные оси). По оси OX будем изображать величину признака (в нашей ситуации длину), а по оси OY — частоту.

У многих школьников вызывает недоумение, как на оси OX отложить значение признака, когда все измерения у нас уже разбиты на интервалы. Прежде всего, чтобы избежать подобных вопросов и далее, давайте условимся, что при графическом представлении данных **не надо стремиться к абсолютной точности**. Все равно на графиках, гистограммах и т. п. видна лишь общая картина. Некоторые исследователи в такой ситуации поступают очень просто. Они разбивают ось OX на несколько равных промежутков (по числу классов), а снизу подписывают значение классов (рис. 4). Другой способ — это откладывать на оси OX лишь граничные точки интервалов. В нашей ситуации на оси абсцисс наиболее удобно отложить следующие точки: 0; 1; 2; 3; 15; 16; 17 (рис. 4). Можно поступить и еще одним способом: в качестве меток на оси OX можно откладывать среднеклассовые значения. В нашем случае такими значениями будут 0,5; 1,5; 2,5; ... 14,5; 15,5; 16,5 (рис. 4). Обычно при построении гистограммы используют первый или второй способы, а при построении полигона — третий.



Рис. 4. Разные варианты подписи оси OX при построении частотного распределения

При построении графиков, гистограмм и других видов графического изображения материала очень важно правильно выбрать масштаб. Масштаб должен быть таким, чтобы данные воспринимались наилучшим способом. Для достижения этой цели можно воспользоваться несколькими простыми правилами:

1) Если значения (абсцисс или ординат) существенно отличаются от нуля, то совершенно необязательно на графике в качестве начала оси выбирать «0». Можно начать отсчет с любого другого значения.





- 2) На осях OX и OY могут (и должны) быть разные масштабы.
- 3) При выборе масштаба сначала задайте длину осей для будущего графика, а уж потом выбирайте на ней масштаб. В противном случае (если вы сначала выберете масштаб, а затем начнете рисовать ось) длина оси может быть слишком велика, она может даже не поместиться на одном листе. Это будет некрасиво и неинформативно, так как глаз с близкого расстояния воспринимает обычно небольшие рисунки.
- 4) Для выбора длины оси надо выявить минимальное и максимальное значение, которое будет отложено на ней. И уже потом вычислить масштаб.

Теперь можно приступить к построению графического изображения. Оформив ось OX (нанеся границы интервалов или среднечлассовые значения), начинаем откладывать точки, соответствующие тем или иным частотам, которые откладываются по оси OY . У этих точек в качестве ординаты будет частота, а в качестве абсциссы любая точка соответствующего интервала, отложенного по оси OX , как правило, в качестве такой точки берется середина интервала. Если вы по оси OX отложили среднечлассовые значения, то они и будут служить абсциссой для этой точки. Далее, если вы соедините последовательно все точки линией, то вы получите *частотный полигон*. Если построите столбики, упирающиеся в эти точки, то получится *частотная гистограмма* (рис. 5).

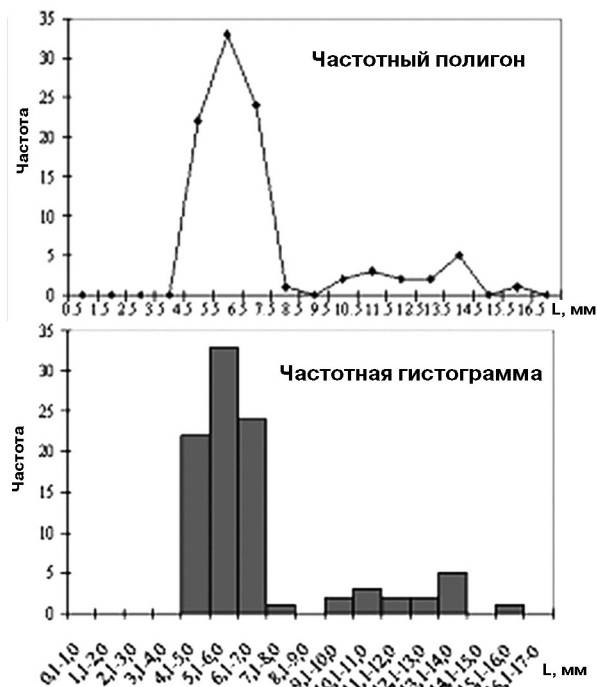
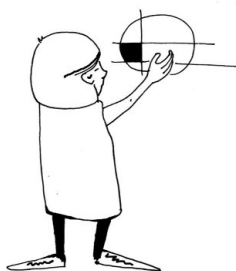


Рис. 5. Частотное распределение значений длины тела бокоплавов *Ponoreia femorata*

Следующий этап — самый сложный. Заключается он в осмыслении той информации, которая приведена на гистограмме или полигоне. Для того чтобы научиться получать эту информацию, необходимо вспомнить, что в генеральной совокупности анализируемая величина распределена в соответствии с некоторым законом. Ранее мы рассматривали в качестве такого закона функцию нормального распределения, но надо постоянно помнить, что величина в генеральной совокупности может иметь и другой закон распределения (в математической статистике описано довольно много таких законов). Однако для простоты давайте остановимся все на том же нормальном распределении. Если в генеральной совокупности действительно оно имеет место, то график, отражающий это распределение, будет иметь уже привычный нам вид колоколообразной кривой (рис. 2).

Однако возможна и другая ситуация, когда вся совокупность объектов включает в себя несколько подсовокупностей, каждая из которых описывается своим собственным нормальным распределением (они отличаются по параметрам распределения: X и S). В такой ситуации в одной генеральной совокупности будут иметь место несколько нормальных распределений (рис. 6).

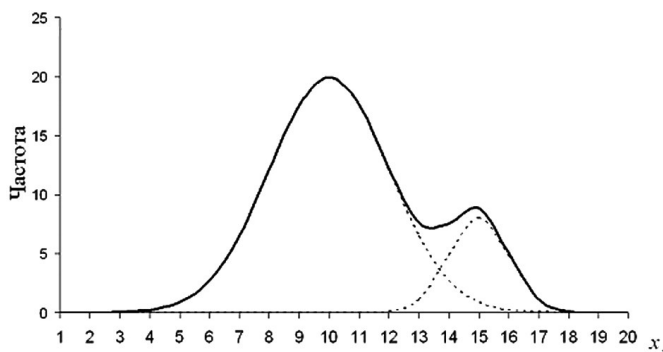


Рис. 6. Двугорбное (бимодальное) частотное распределение, являющееся результатом наложения двух нормальных распределений

Например, в популяции некоторых животных представлены самцы и самки, но самки имеют меньшие размеры. В результате этого, распределение размеров в генеральной совокупности будет состоять из двух распределений, отличающихся параметром X . Но если бы мы не умели отличать самцов и самок, а просто анализировали бы размеры тела животных, то распределение было бы не колоколообразным, а **двугорбым**, или, как говорят, *бимодальным* (рис. 6). Если в генеральной совокупности много подмножеств, то оно называется *полимодальным* (рис. 7).

Поскольку выборочное распределение — это образ генерального распределения, то наличие полимодальности в таком распределении — это верный признак того, что генеральная совокупность



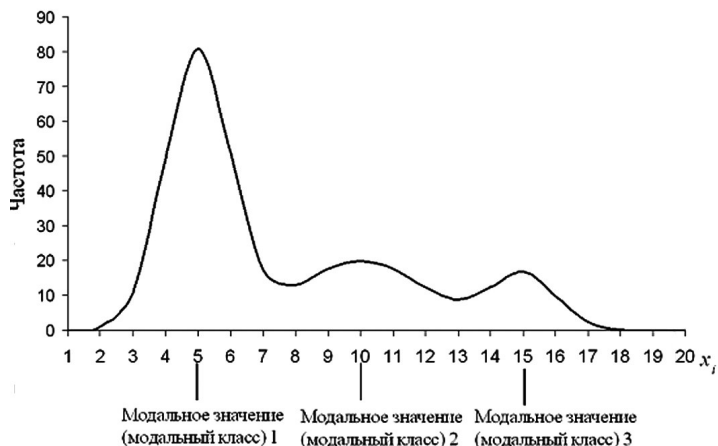
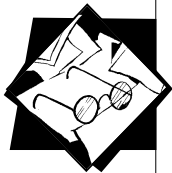


Рис. 7. Полимодальное частотное распределение



состоит из нескольких подмножеств. Что это за подмножества, статистика дать ответа не может. Трактовка этих подмножеств — задача сугубо биологическая. Для ответа на вопрос о том, что это за подмножества, надо очень хорошо знать объект.

Вернемся к нашему примеру с *Pontoporeia femorata*. Распределение по длине тела этих бокоплавов имеет явно полимодальный характер (рис. 5). Известно, что бокоплавов растут линьками, то есть, у них наблюдается дискретный рост. Это позволяет предположить, что наблюдаемые пики в распределении — это отражение подмножеств, соответствующих возрастным когортам. Внимательный читатель может заметить, что в выборочном распределении размеров бокоплавов пиков можно насчитать несколько (до четырех). Это становится видно, если в качестве «маркеров» пиков рассматривать классы, где находится особей больше, чем в соседних классах. Такие классы называются *модальными*. В нашем выборочном распределении модальными оказываются следующие классы: 5,1–6,0; 10,1–11,0; 13,1–14,0; 15,1–16,0. Означает ли это, что надо каждый из этих пиков рассматривать как проявление самостоятельного подмножества в генеральной совокупности? Формально ответить на этот вопрос нельзя! Может быть, да, а может быть, и нет. На чем же остановиться? Если вы анализируете только один признак, то окончательно ответить на вопрос о том, сколько же пиков в вашем распределении, может только очень большая выборка. Чем больше объем выборки, тем ближе выборочное распределение к генеральному, а, стало быть, менее выраженными становятся разные шумы, связанные с неточностями измерения, малым объемом выборки и т. д. Если увеличение объема выборки не убирает пики, выявленные ранее, то это означает, что они отражают реально существующие внутривидовые подгруппы.

Несколько проще решить вопрос о количестве подмножеств в вашей совокупности, если вы анализируете не один, а два или много признаков. Самый простой метод при наличии такого материала — это построение *скеттер-диаграммы*.

Если у вас измерено только два признака, то скеттер-диаграмма, или точечная диаграмма, строится очень просто. Для ее построения надо начертить две координатные оси. На одной из них будут откладываться значения одного признака, а на другой — значения другого. При этом заметьте, что здесь уже нет необходимости выделять классы. Каждый отдельный экземпляр вашего объекта будет нанесен на координатной плоскости в виде точки с координатами, соответствующими значениям измеренных признаков.

Вернемся к примеру с бокоплавами. Мы измеряли два параметра: длину тела и вес организма. Если нанести все 95 пар измерений на координатную плоскость, то получим облако точек (рис. 8). На этом облаке четко различимы две группы точек, два «сгущения». Одно располагается в области малых значений длины и веса. Это сгущение достаточно компактно. Второе сгущение — в области более высоких значений, но это скопление более «размазано», оно обладает большей дисперсией, т.е. большим рассеянием. Если внимательно приглядеться, то можно увидеть и еще одну отдельность, состоящую всего из одной особи в области самых высоких значений длины и веса.

Рассмотрение скеттер-диаграммы и частотного распределения позволяет нам прийти к выводу, что в популяции бокоплава *Pontoporeia femorata* присутствуют три подмножества особей, которые мы, учитывая дискретный характер роста, можем трактовать как возрастные группы. На основании полученных статистических данных, мы можем сделать биологический вывод о том, что в изученной популяции бокоплавы живут до трех лет. Кроме того,

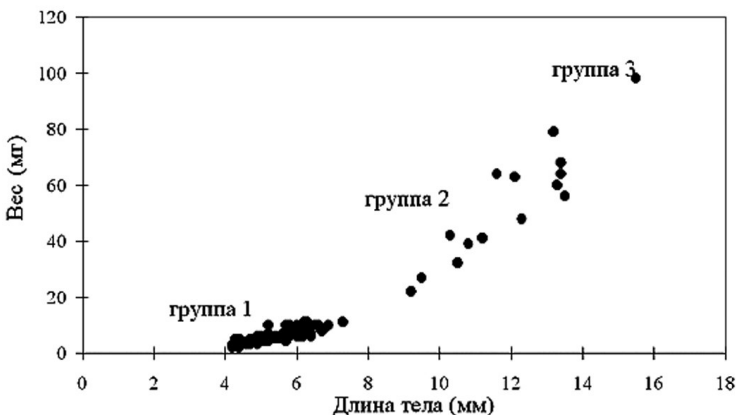


Рис. 8. Скеттер-диаграмма, отражающая соотношение длины и веса тела у бокоплавов *Pontoporeia femorata*



можно сделать еще ряд любопытных выводов. Например, можно высчитать, какова вероятность особи первого года дожить до второго, а особи второго — до третьего. Для этого достаточно подсчитать отношение численности особей второй размерной группы к численности первой и численности третьей ко второй соответственно. Можно сделать и еще много интересного. Главное, что мы смогли обоснованно разделить всю популяцию на подмножества.

Особый случай анализа размерно-возрастной структуры популяции — это случай, когда отдельные распределения, входящие в состав общего полимодального распределения, сильно перекрываются. В этой ситуации четко провести границу между совокупностями нельзя. Существует много разнообразных ухищрений, которыми пользуются биологи для того, чтобы разделить такие подмножества. Все эти методы очень сложны, и в пределах данного пособия мы просто не можем их описать. Поэтому можно предложить вам менее строгий, но достаточно простой способ деления популяции на подмножества — в анализ можно включать только особей, имеющих модальные размеры. Например, пусть вам надо выделить в популяции особей разных возрастных групп и проанализировать их распределение в пространстве. Однако наблюдается существенное перекрывание размерных характеристик у разных возрастов (с полной уверенностью нельзя отнести каждую особь к той или иной совокупности). В такой ситуации вы можете включить в анализ только тех особей, которые имеют размеры, соответствующие модальным классам.

Анализ размерно-возрастной структуры популяции имеет и свои ограничения, суть которых, впрочем, лежит не в статистике, а в биологической природе объектов. Например, не всегда выделенные совокупности можно четко трактовать как возрастные. Они могут иметь и совершенно другую природу. Но в любом случае, если такие совокупности выделяются, их сбрасывать со счетов нельзя. Они что-нибудь да означают!

(Продолжение следует)

