

П. Ф. Рокицкий

ОСНОВЫ
ВАРИАЦИОННОЙ
СТАТИСТИКИ
ДЛЯ
БИОЛОГОВ

Минск 1961

*Рокицкий
Петр Фомич*

**Основы вариационной статистики
для биологов**

И з д а т е л ь с т в о
Белгосуниверситета имени В. И. Ленина
М и н с к—1961

Редактор *Найдович А. Н.*
Художественный редактор *Лысенко П. П.*
Техредактор *Беленькая И. Е.*
Корректоры *Сушко К. В., Смеян И. В.*

Печатается по постановлению РИС о БГУ им. В. И. Ленина

АТ 04716. Сдано в набор 12.X 1960 г. Подписано к печати 14.III-61 г.
Тираж 8000. экз. Бумага 84×108¹/₃₂. Печ. л. 7. Усл.-печ. л. 11,48.
Уч.-изд. л. 16,5. Заказ 154. Изд. зак. 58. Цена 65 коп.

Типография Издательства Белгосуниверситета имени В. И. Ленина,
г. Минск, ул. Кирова, 24.

Проф. П. Ф. РОКИЦКИЙ

ОСНОВЫ
ВАРИАЦИОННОЙ СТАТИСТИКИ
ДЛЯ БИОЛОГОВ

ИЗДАТЕЛЬСТВО
БЕЛГОСУНИВЕРСИТЕТА имени В. И. ЛЕНИНА
МИНСК — 1961

Книга представляет собой учебное пособие по курсу вариационной статистики для студентов биологических факультетов университетов и других вузов биологического профиля, а также может быть использована научными и практическими работниками—биологами различных специальностей.

В работе подробно и последовательно изложены необходимые для биологических исследований статистические методы: группировка материала, составление вариационных рядов, вычисление важнейших статистических показателей, характеризующих совокупности, измерение корреляции и регрессии. Уделено внимание понятиям вероятности и достоверности и их значению для анализа биологических данных.

Изложенный в книге материал иллюстрируется большим числом конкретных примеров из различных областей биологии. Каждая глава содержит проверочные вопросы и задачи на материале ботаники, зоологии, животноводства, растениеводства, физиологии, генетики. В приложении дается перечень статистических показателей и их формул и 9 статистических таблиц, необходимых для оценки ряда показателей и проверки результатов.

ВВЕДЕНИЕ

Общеизвестно, что в физике и химии, а также в различных областях техники и народного хозяйства широко применяются математические приемы и методы, с помощью которых можно точно характеризовать те или иные явления и выражать с помощью математических формул разнообразные связи и зависимости между ними. В настоящее время биология претерпевает очень резкий переход от чисто описательной науки, которой она была в прошлом, к точной и экспериментальной, пользующейся количественными методами исследования и основывающейся не только на наблюдении, но и на эксперименте. Значение количественных методов особенно подчеркивается тем фактом, что в биологии приходится иметь дело, как правило, не с одним, а со многими объектами. В озерах и реках водится множество рыб, относящихся к одному и тому же и к разным видам, миллионы экземпляров различных видов рачков, моллюсков, коловраток, водорослей, инфузорий и других животных и растений. В совхозах и колхозах разводят сотни и тысячи голов крупного рогатого скота, тысячи и десятки тысяч голов овец, тысячи свиней, кур и других животных, выращивают миллионы различных полевых, огородных, садовых и иных растений.

Все эти организмы характеризуются самыми различными показателями: коровы—удоями за лактацию или за 300 дней лактации, процентом жира молока, живым весом; рыбы—весом, длиной тела; овцы—настригами, длиной и тониной шерсти; куры-несушки—количест-

вом снесенных за год яиц; колосья пшеницы—количеством зерен в колосе, весом отдельных зерен и т. д.

Для того, чтобы суметь разобраться в разнообразных количественных данных о животных и растениях как в условиях производства, так и в научных опытах с ними, нужны определенные приемы и методы, заимствованные из математики.

Сравнение между собою сортов растений или пород животных, изучение влияния различных внешних факторов на физиологические и биологические процессы, на хозяйственные и биологические свойства организмов, изучение видов, подвидов, экотипов, географических рас в природе—все это может быть сделано только путем применения точных количественных методов наблюдения или эксперимента и последующего математического анализа результатов. Только в этом случае можно сделать правильные выводы. Особенно это относится к быстро развивающимся новым отраслям биологии: экологии, генетике, теоретической систематике, селекции, физиологии и фармакологии.

Роль количественного подхода к явлениям природы была замечательно выражена еще в известном завещании Галилея: измерять все, что измеряется, и делать пригодным к измерению то, что пока не поддается измерению.

Дело не только в применении математических приемов для количественного изучения, но и в правильной оценке количественных закономерностей в явлениях природы, ибо на основе количественного анализа можно выявить их качественное своеобразие и отсюда правильно подойти к проведению необходимых наблюдений и к постановке опытов, оценить полученные цифровые данные и предостеречь против возможных исходных ошибок в наблюдениях или опытах.

Применение математических методов к живым существам составляет особую науку—биометрию.

Термин «биометрия» (или «биометрика») был предложен еще в конце XIX века. Биометрия рассматривалась как наука о применении математических методов для изучения разнообразия живых существ. В настоящее время она дает в распоряжение биолога большое количество приемов, подчас очень сложных, с помощью которых можно значительно лучше выяснить особенности

того или иного материала (животных, растений или микроорганизмов). Однако мы ограничимся рассмотрением небольшой группы элементарных математических приемов, которые необходимы для работы биолога, зоолога, ботаника, растениевода, животновода, входящих в область лишь одной математической науки,—вариационной статистики. Современная же биометрия (или биоматематика, как иногда ее называют) использует приемы и некоторых других областей математики. Но приемы вариационной статистики имеют наибольшее значение, ибо они позволяют проводить количественный анализ массовых явлений, то есть таких явлений, которые охватывают значительное количество отдельных изменяющихся величин (слово «вариация» в переводе и означает «изменение», «колебание»). Как раз вариационная статистика и занимается математическим изучением закономерностей, проявляющихся в массовых явлениях. Вот почему наш курс называется не биометрией, а вариационной статистикой.

В излагаемых ниже главах будут даны лишь элементарные основы вариационной статистики в размерах, отведенных на этот курс учебными планами биологических факультетов университетов.

Более полные сведения о вариационной статистике дают специальные руководства, которые указаны в заключительной части книги.

Глава 1

ГРУППИРОВКА ДАННЫХ, СОВОКУПНОСТЬ И ВАРИАЦИОННЫЙ РЯД

Характеристика совокупности. Всякое множество отдельных изменяющихся объектов составляет так называемую совокупность. Совокупностями являются популяции рыжих полевок того или иного района, стадо коров, потомство быка, заготавливаемые в определенном районе беличьи шкурки, растения на опытных делянках, группа цыплят, на которых ставится опыт по применению антибиотиков, мальки окуня в озере и т. д. Понятие совокупности приложимо не только к животным и растениям. Такими же совокупностями являются, например, дети, родившиеся в стране в течение какого-то года или месяца, молекулы газа в том или другом объеме. В состав совокупности входят различные члены или единицы: для популяции животных—каждое отдельное животное, для стада коров единицей является каждая корова, для совокупности шкурок—каждая шкурка, для потомства быка— все телки и бычки, от него полученные, для совокупности зерен гречихи—каждое отдельное зерно.

Обычно число единиц совокупности называют объемом совокупности и обозначают латинской буквой *n*. Единица совокупности может характеризоваться определенными признаками, например, коровы—удоями за лактацию, весом, мастью, молекулы газа—скоростями их движения и т. д. Каждый изучаемый признак принимает разные значения у различных единиц совокупности, он меняется в своем значении от одной единицы совокупности к другой. Это изменение называется вариацией (т. е. изменчивостью) или дисперсией (т. е. рассеяни-

ем). Мы говорим—«признак варьирует». Это означает, что он принимает разные значения у различных членов совокупности, например, коров данной породы, мышей опытной группы, поросят одного помета и т. д. Значение или меру признака для той или иной единицы совокупности называют вариантой и обозначают определенной буквой. Раньше обозначали варианты буквой v , теперь чаще обозначают буквой x . В таком случае ряд вариантов в совокупности следует обозначать как $x_1, x_2, x_3, \dots, x_n$. Самую же варьирующую величину называют случайной переменной. Варианты являются ее числовыми значениями.

Чаще всего в состав совокупностей входят отдельные особи. Так, например, при характеристике стада коров по весу во взрослом состоянии (на 1 января определенного года) за единицу совокупности следует взять каждую корову. Однако возможны случаи, когда единицей совокупности может быть не каждое животное в отдельности, а только какая-то его характеристика. Так, изучая вариацию коров стада по молочной продуктивности, можно взять за единицу каждую лактацию. Тогда при общем количестве коров в стаде, например 100 голов, количество изучаемых за несколько лет лактаций может быть 500 или 600. Отдельными вариантами будут величины удоев за каждую лактацию. Можно изучать вариацию того или иного признака во времени даже на одном животном. Как известно, жирность молока изменяется не только по дням лактации, но и по отдельным дойкам того же дня. Варьирующие данные о проценте жира в молоке определенной коровы, полученные путем измерения жирности за ряд доек и дней лактации, также составляют совокупность, которую можно изучить вариационно-статистическими методами.

Такой же совокупностью, очевидно, является ряд показателей состава крови у одной морской свинки при его изучении в течение какого-то времени.

В общем виде можно сказать, что сумма наблюдений или измерений есть тоже совокупность. Каждое отдельное наблюдение, при котором устанавливается значение случайной переменной, тогда будет единицей совокупности.

Следует иметь в виду, что совокупность может состоять из других, более частных совокупностей. Совокуп-

ность, представляющая собой всех животных данной породы, распадается на частные совокупности—стада отдельных хозяйств, колхозов или совхозов. В пределах стада одного хозяйства можно выделить еще более частные совокупности, например потомство определенных быков.

При постановке опытов по изучению влияния каких-либо антибиотиков на рост крыс внутри совокупности, охватывающей всех опытных и контрольных животных, можно отдельно рассматривать каждую группу, подвергавшуюся воздействию определенных факторов, как самостоятельную, более частную совокупность. Во всех случаях мы будем встречаться с постоянной вариацией как внутри отдельных частных совокупностей, так и между ними.

Задачей изучения всякой совокупности является получение статистических (или, как иногда говорят, биометрических) характеристик или показателей, которые позволяют судить о данной совокупности в целом, о вариации внутри совокупности и об отличии ее от других, сходных с ней или близких к ней совокупностей.

Именно тогда совокупность становится статистической, когда в ее описание вносится количественный метод. Применение количественного метода изучения совокупности и позволяет получать для нее ряд статистических показателей (в специальной литературе статистические показатели нередко называют сокращенно «статистиками»). С их помощью мы получаем основную информацию о совокупности.

Варьирующие признаки и их учет. При изучении единиц совокупности по тем или другим признакам необходимо записать полученные данные.

В настоящее время предпочитают производить такого рода записи на карточках, так как их можно затем группировать любым образом. При большом количестве карточек обработка записей может производиться счетной машиной. В этом случае карточки должны быть перфорированными, т. е. в определенных местах на них должны быть пробиты дырочки или сделаны вырезы в соответствии с записанными цифрами. Машина сама производит необходимые подсчеты по этим дырочкам или вырезам. Наконец, в особо сложных случаях все полученные при опытах или наблюдениях данные переводятся

на условный код, который записывается в соответствующих частях электронно-счетных машин. Такие машины в дальнейшем могут обработать полученные данные, при этом с большой скоростью.

Способы обработки данных сильно зависят от того, каков характер изменчивости (вариации) изучаемых признаков. Различия между вариантами могут выражаться в каких-то качествах. Такую изменчивость называют качественной, или альтернативной. Так, если совокупность животных характеризуют по масти, тогда каждая варианта должна получить качественную характеристику в соответствии с заранее принятыми обозначениями: черная, рыжая, черно-пестрая, черно-рыжая и т. д. В этом простейшем случае подсчет числа особей в каждой из выделенных групп дает представление о совокупности в целом. Подсчет, произведенный в абсолютных числах, можно выразить в процентах и представить в виде диаграммы (столбиками или секторами круга). Однако, как мы увидим в дальнейшем, и при альтернативной изменчивости возможно получение ряда статистических показателей.

В других случаях различия между вариантами являются количественными. Количественная вариация может быть двух типов: прерывистая (дискретная) и непрерывная. В первом случае различия между вариантами, отдельными значениями случайной переменной, выражаются целыми числами, между которыми нет и не может быть переходов. Примером может служить количество детенышей в помете (поросят у свиноматок или щенков у серебристо-черных лисиц), число сосков у свиноматок, число лучей в плавниках рыб, количество лепестков в цветке, число позвонков у птиц и т. д. Для изучения подобного варьирования надо сосчитать у каждой единицы совокупности число изучаемых элементов и записать его на соответствующую карточку. При непрерывной изменчивости значения вариант не обязательно выражаются только целыми числами. Все зависит от того, какая степень точности принимается для характеристики данного количественного признака. Так, например, при изучении веса крупного рогатого скота можно ограничиться значениями вариант, выраженными в килограммах, отбросив граммы, но совершенно недостаточно округлять веса рыб до килограммов, а необ-

ходимо выражать их в граммах, так как грамм здесь имеет большое значение. В опытах же по изучению влияния гормонов на рост гребня у цыплят вес гребня придется измерять в миллиграммах. Молочную продуктивность за лактацию обычно выражают в килограммах, но общая картина удоев не изменится, если округлять количество молока за период лактации до десятков килограммов. Оценка же жирности молока в целых процентах явно недостаточна, ее надо давать в десятых или даже в сотых долях процента. Однако во всех этих и им подобных случаях существует непрерывная вариация, выражающаяся в том, что между вариантами возможны все переходы. При изучении непрерывной изменчивости надо все единицы совокупности характеризовать количественно с той степенью точности, которая заранее намечена и больше всего подходит в данном конкретном случае, и полученные данные (варианты) внести в карточки.

Группировка данных при качественной и дискретной изменчивостях. Чтобы проанализировать ту или иную совокупность, необходимо сгруппировать полученные отдельные варианты и затем представить эту группировку в виде таблицы или ряда. Только при упорядочении полученных данных можно их обработать математически и вывести вариационно-статистические показатели, которые будут исчерпывающе характеризовать изучаемую совокупность. Проблема группировки занимает большое место в статистике вообще (особенно в экономической), так как ошибочная группировка данных может привести к неправильным выводам о существе изучаемого явления.

Наиболее проста группировка при качественной, альтернативной, изменчивости. Так, если при просмотре совокупности из 150 коров рогатых было обнаружено 120, а комолых—30, то полученные данные можно свести в таблицу (табл. 1).

При количественной изменчивости надо предварительно наметить для таблицы классы, охватывающие все изученные количественные данные от минимальных их значений до максимальных. Это легко сделать при прерывистой (дискретной) количественной изменчивости.

Допустим, что была изучена плодовитость 80 самок серебристо-черных лисиц, т. е. число родившихся у каж-

дой самки щенков. Варианты $x_1, x_2, x_3, \dots, x_{80}$ этой совокупности будут выражены цифрами, представленными в табл. 2.

Таблица 1

Группировка коров по рогатости и комолости

Типы коров	Количество	Процент
Рогатые	120	80
Комолые	30	20
Всего	150	100

Таблица 2

Количество щенков у 80 самок серебристо-черных лисиц

4	5	3	4	6	7	8	3	1	4
6	4	4	3	2	5	3	4	5	4
5	3	4	5	4	4	4	6	5	7
6	4	5	4	4	4	4	2	3	4
5	5	4	5	4	4	6	4	4	4
4	8	7	5	4	9	4	3	4	4
5	4	6	4	4	3	4	4	4	2
4	4	5	4	6	4	3	3	4	2

Легко видеть, что минимальное число щенков 1, максимальное—9. Отсюда естественно установить 9 классов: с 1 щенком, с 2, с 3 и т. д. и распределить все варианты по этим 9 классам. Наиболее простым способом разнесения вариантов по классам является следующий. Составляется таблица с намеченными 9 классами и в соответствующие горизонтальные строчки разносятся все варианты, начиная от первой. Обозначаются они так: первые четыре данного класса—точками, а последующие—черточками, соединяющими четыре точки. Число 10 будет в таком случае фигурой $\boxed{\times}$. В табл. 3 произведена разноска первых 20 вариантов, записанных в двух верхних строчках табл. 2.

Предоставляем каждому самому довести разноску данных табл. 2 до конца. В окончательном виде таблица будет иметь вид табл. 4.

Т а б л и ц а 3

Разноска 20 вариант по классам

Классы (число щенков в помете каждой самки)	Частоты (количество вариант в каждом классе)
1	1
2	1
3	4
4	7
5	3
6	2
7	1
8	1

Т а б л и ц а 4

Распределение 80 самок серебристо-черных лисиц
по классам

К л а с с ы	Ч а с т о т ы
1	1
2	4
3	10
4	39
5	13
6	7
7	3
8	2
9	1
$n = 80$	

Вариационный ряд и его графическое изображение.
Таким образом, после распределения всех вариантов по

классам получился ряд, в котором показано, как часто встречается каждый тип или класс вариант и как варьирует число щенков в помете у отдельных самок, начиная от минимальной величины (1 щенок) и кончая максимальной (9 щенков). Поэтому подобные ряды были названы вариационными. По вариационному ряду можно судить не только о границах колеблемости изучаемого количественного признака, но и о характере вариации. В данном примере в вариационном ряду наиболее частым является класс «4 щенка», следующими за ним по частоте являются классы «3 щенка» и «5 щенков». Наиболее же редкими являются крайние классы «1 щенок» и «9 щенков». Класс, обладающий наибольшей частотой, получил название модального, значения же крайних классов называются лимитами, или пределами.

Всякий вариационный ряд можно изобразить графически. Графическое изображение вариационного ряда получило название кривой распределения, или вариационной кривой. При прерывистой изменчивости это будет, конечно, не кривая, а многоугольник, или так называемый полигон, который показан на рис. 1. При построении кривых распределения, или полигонов, нужно всегда доводить их справа и слева до нулевых классов, т. е. тех соседних классов, в которых уже не было ни одной варианты. В нашем примере ими являются классы «0 щенков» и «10 щенков».

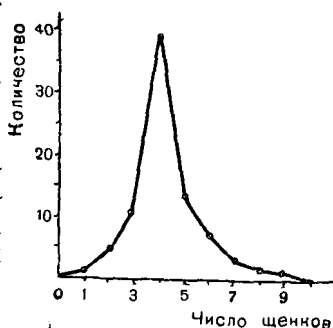


Рис. 1. Кривая распределения 80 самок серебристо-черных лисиц по числу щенков в помете

В приведенном примере модальным был только один класс «4 щенка». Поэтому полигон распределения самок лисиц по числу щенков в помете имел только одну вершину. Однако возможны случаи, когда в вариационном ряду обнаруживается несколько модальных классов, и тогда полигон является многовершинным. Наиболее простой причиной многовершинности, особенно при очень растянутых рядах, является недостаточное количество вариантов в изученной группе. При малом числе особей в

некоторых классах вариационного ряда может вообще не быть ни одной варианты. Вариационный ряд окажется прерывистым, а вариационная кривая—разорванной на части. Однако, если при большом числе особей в изучаемой совокупности сохранится дву- или многовершинность, причину этого надо искать в том, что биологический материал в действительности является смещением двух качественно различных совокупностей, которые или находились в резко отличных условиях внешней среды или принадлежат к разным типам, морфам. Так как многие виды в природе являются полиморфными или диморфными, то соединение в одном ряду особей разных морф может дать внешнюю картину дву- или многовершинности. Например, платиновые лисицы отличаются по числу щенков от серебристо-черных, поэтому было бы неправильно помещать в один вариационный ряд по этому признаку и платиновых, и серебристо-черных лисиц.

Группировка данных при количественной непрерывной изменчивости. В этом случае группировка данных является наиболее трудной. Допустим, что в результате измерения веса 25 кроликов различных пород были получены варианты, представленные в табл. 5, при этом они расположены в так называемом ранжированном виде, т. е. от меньших величин к большим.

Таблица 5

Веса 25 кроликов в кг (для большей наглядности взяты кролики различных пород)

3,2	4,5	5,2	5,6	6,0
3,8	4,7	5,2	5,7	6,3
4,1	4,9	5,3	5,8	6,4
4,3	5,0	5,3	5,8	6,7
4,3	5,1	5,4	5,9	7,3

Здесь нет тех естественных классов, с которыми мы встречались при анализе прерывистой, дискретной вариации. Их необходимо наметить произвольно. Разница между наибольшим и наименьшим значениями вариантов в нашем примере равна 4,1 (7,3—3,2). Но чтобы иметь примерно 8—9 классов, размеры их должны быть 0,5 кг. В таком случае можно наметить следующие классы:

3,0—3,4 кг; 3,5—3,9 кг; 4,0—4,4 кг; 4,5—4,9 кг; 5,0—5,4 кг; 5,5—5,9 кг; 6,0—6,4 кг и т. д.

На правильное построение шкалы для классов надо обращать очень большое внимание. Во-первых, необходимо, чтобы величина классового промежутка была всегда одной и той же. Было бы неправильно, если бы в начале ряда был взят классовый промежуток 0,5 кг, как в нашем примере, а в конце ряда—1,0 кг (величину классового промежутка обычно обозначают буквой *i*). Во-вторых, границы классов должны быть намечены таким образом, чтобы одна и та же цифра не повторялась в двух классах. Если первый класс заканчивался величиной 3,4, то второй класс должен начинаться со следующей по порядку цифры—3,5. Если бы классы были намечены следующим образом: 4,0—4,5; 4,5—5,0; 5,0—5,5 и т. д., то всегда было бы сомнение, к какому классу отнести особь со значением 4,5 или 5,0. Если же один класс будет охватывать значения вариант от 4,0 до 4,4 включительно, а другой—от 4,5 до 4,9 включительно, разноска вариант по намеченным классам не вызовет затруднений. Ее можно проводить тем же методом, который использован при составлении табл. 3 и 4.

В результате данные по весам кроликов будут представлены в табл. 6.

Т а б л и ц а 6

Распределение 25 кроликов по весу

Классы, кг	Частоты	Классы, кг	Частоты
3,0 — 3,4	1	3,0 — 3,9	2
3,5 — 3,9	1	4,0 — 4,9	6
4,0 — 4,4	3	5,0 — 5,9	12
4,5 — 4,9	3	6,0 — 6,9	4
5,0 — 5,4	7	7,0 — 7,9	1
5,5 — 5,9	5		
6,0 — 6,4	3		
6,5 — 6,9	1		
7,0 — 7,4	1		
$i = 0,5 \text{ кг}$	$n = 25$	$i = 1,0 \text{ кг}$	$n = 25$