

Раздел I. Эколого-статистические методы анализа.

Тема лекции 3. Вариационно-статистический метод анализа.

Величина признака у отдельной особи называется *вариантой* и обозначается символом "x". Число вариантов, входящих в состав генеральной совокупности определяет её *объём* и обозначается буквой N. Теоретически объём генеральной совокупности ничем не ограничен ($N \rightarrow \infty$).

По объёму выборка может быть:

- малой (численностью менее 30 вариантов) и
- большой (численностью более 30 вариантов).

Число вариантов, входящих в состав выборки также определяет её объем, но обозначается, по сравнению с генеральной, буквой *л*. Метод обработки показателей признаков зависит от объема выборки.

Различают прямой метод обработки показателей признаков, применяемый в малой выборке и, непрямой способ, т.е. метод группировки, приемлемый в больших выборках.

Обработка начинается путем систематизации созданных данных. Процесс систематизации результатов массовых наблюдений, объединения их в относительно однородные группы по некоторому признаку называется *группировкой*. Группировка представляет первичные данные в комплексном виде и, направлено на выявление связей между явлениями. Наиболее распространенной формой группировки являются статистические таблицы.

Выборочный метод в экологометрике. Статистические совокупности могут быть расчленены одновременно по нескольким признакам. В связи с этим различают одномерные, двумерные и многомерные статистические совокупности. Они могут также состоять из дискретных объектов и непрерывных множеств. Признаки объектов или множеств могут также принимать либо дискретные, либо непрерывные значения..

Если генеральная совокупность объектов определена, то выбранные из нее случайным образом n объектов называют случайной выборкой. Математическая статистика имеет дело только со случайными выборками. Числовые или нечисловые характеристики изучаемых объектов, полученных в результате наблюдений или измерений у каждой единицы выборки, называют *статистическими данными*.

Различают следующие виды отбора объектов из генеральной совокупности в выборку:

- 1) *групповой*, когда генеральная совокупность предварительно делится на типические группы (например, участки, делянки, районы);
- 2) *серийный*, когда генеральная совокупность делится на серии, или гнезда;
- 3) *механический*, когда генеральную совокупность предварительно разбивают на несколько частей (или групп) и затем из каждой части отбирают в выборку по несколько объектов, представителей этих частей.

Если учесть, что изучение выборки преследует цель исследования свойств и характеристик генеральной совокупности, то выборка должна быть представительной (репрезентативной). Это означает, что ее объем должен быть не менее некоторого специального определенного объема, что будет рассмотрено далее.

Исследования или измерения каких-либо свойств или характеристик отдельных объектов выборки представляются в виде статистического ряда, они могут быть *дистрибутивные, вариационные, динамики, или временные*.

Дистрибутивные ряды служат для разделения объектов выборки по каким-либо признакам, заранее известным, например, людей по группам крови.

Вариационные ряды показывают закономерность распределения единиц изучаемой выборки по ранжированным значениям варьирующего признака, например, пробы воздуха по содержанию пылевых частиц.

Ряды динамики (временные) показывают закономерность изменения варьирующего признака в зависимости от времени.

При статистической обработке не следует упускать из внимания экологический смысл изучаемого объекта, который всегда устанавливает границы исследования. Таким образом, выборка - группа объектов, отличающихся некоторыми особенностями:

- это часть генеральной совокупности;
- объекты в выборку отбираются в случайном порядке определенным способом;
- объекты выборки исследуются для определения характеристик самой выборки и для оценки характеристик генеральной совокупности.

Основное правило - соблюдение равной вероятности попадания в выборку любого объекта генеральной совокупности, т.е. выбор должен быть совершенно случаен.

Вариационный ряд.

1. Определяют *объем выборки (n)* путем подсчета, который соответствует общему числу вариант.

2. В сводке исходных данных отыскивают минимальную x_{\min} и максимальную x_{\max} варианты (*лимиты*). Лимиты указывают на общий размах разнообразия признаков (R): $R = x_{\min} - x_{\max}$. Лимиты (от лат. *limets* – предел) и размах вариации – это простые и наглядные характеристики варьирования. Чем сильнее варьирует признак, тем больше размах вариации, и наоборот.

3. Устанавливают *число классов (l)* исходя из объема выборки по следующей методике:

Объем выборки – n (от – до)	Число классов
31 – 60	5 – 6
61 – 100	$l = \frac{n}{10}$; полученное число округляют до целого
101 и более	10 – 14

4. Определяют величину *классного промежутка* (k) по формуле:


$$k = \frac{x_{\max} - x_{\min}}{l} = \frac{R}{l} \quad (2)$$

Если полученная величина "к" дробная, то для удобства ее округляют до целого числа по общепринятой методике, т.к. удобнее будет рассчитывать границы классов.

5. После определения n , lim , l , k приступают к построению вариационного ряда.

Двойной ряд чисел, состоящий из классов и соответствующих им частот, называют *вариационным рядом*.

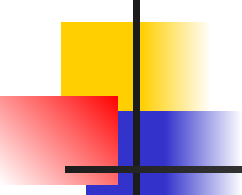
Для построения вариационного ряда строят вспомогательную таблицу, которая включает в себя следующие восемь граф: номер класса – № I; границы классов – W; середина класса – A; разноска; частота – f; отклонение – a; $f \cdot a$; $f \cdot a^2$.



6. *Заполнение вариационного ряда:*

Устанавливают границы классов (W). В графе "границы классов" необходимо вписать две цифры: первая указывает на нижнюю, а вторая – на верхнюю границу соответствующего класса. В пределах указанных границ и располагаются все варианты данного класса.

Следует помнить, что вариационный ряд всегда начинается с минимальной варианты и заканчивается – максимальной.



6.1. Вычисление нижних границ классов вариационного ряда:


– нижняя граница первого класса: x_{\min} ;

– нижняя граница второго класса: для её получения к x_{\min} прибавляют величину классного промежутка k .

Таким же образом находим нижние границы всех последующих классов.

6.2. *Вычисление верхних границ классов вариационного ряда.*

Для получения верхней границы первого класса необходимо от нижней границы второго класса отнять 1, если варианты выражаются целыми числами (при дробных значениях отнимают 0,1; 0,01 и т.д.):



Из всего вышеизложенного следует, что при разбивке вариант на классы необходимо соблюдать следующие *правила*:

- Количество классов должно быть не больше 14-15 и не менее 5.
- Границы классов должны быть такими, чтобы каждая варианта могла быть отнесена только к одному классу.
- Размеры всех классов должны быть равными.
- Первый и последний классы могут быть неполными.

6.3. *Середина класса (A)*. Для заполнения данной графы необходимо для каждого класса нижнюю границу прибавить к верхней и полученную сумму разделить на два.

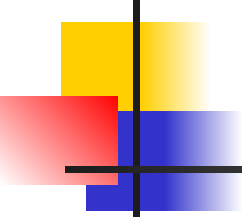
6.4. *Разноска* – это группировка вариантов по классам. Удобно разноску вариант производить по порядку их записи, ставя в классах сначала точки, а когда точек наберется четыре, соединять их отрезками в виде конверта:

-1	3	5	7	9
-2	4	6	8	10

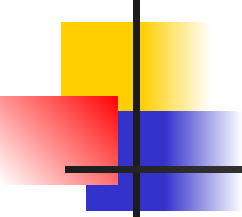
Таковыми фигурами удобно подсчитывать общую сумму вариант.

6.5. Определяют частоты (f) в классах. Для этого подсчитывают отметки разностей каждого класса. Затем, при суммировании частот ($\sum f$) получают общее число вариантов в выборке: $\sum f = n$.

В вариационном ряду при нормальном распределении существует определенная закономерность. Крайние вариации малочисленны, но с приближением к середине ряда частоты вариаций увеличиваются. В середине вариационного ряда или вблизи имеется одна вариация (один класс), на которую приходится наибольшее число частот. Этот класс называют модальным.




6.6. Для заполнения графы «отклонение» сначала находят модальный класс ($I_{\text{мо}}$). Класс, в котором встречается наибольшее число вариантов называется *модальным*. В вариационном ряду в графе «отклонение» модальный класс обозначается нулем. Для классов, предшествующих модальному характерно снижение показателей изучаемого признака. Поэтому в графе «отклонение» классы перечисляются от 0 по порядку вверх со знаком «минус»: -1, -2. Для классов, следующих после модального, характерно увеличение показателей изучаемого признака и они перечисляются со знаком «плюс»: +1, +2.



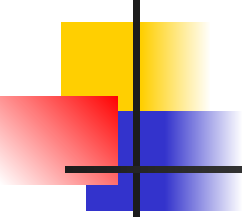
6.7. Для заполнения седьмой графы вычисляют произведения отклонений на частоту: f_a . При этом необходимо помнить, что в графе « f_a » действует правило знаков. В итоге этой графы указывается сумма.

6.8. Для заполнения последней восьмой графы (f_a^2) каждый показатель по классам предшествующей седьмой графы (f_a) умножается на соответствующий показатель шестой графы (a): $f_a \times a = f_a^2$. Получаемые при этом произведения (f_a^2) всегда положительны. Наконец, просуммировав все показатели восьмой графы получают $\sum f_a$.

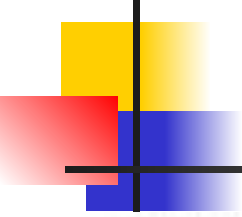


Графическое изображение вариационного ряда.

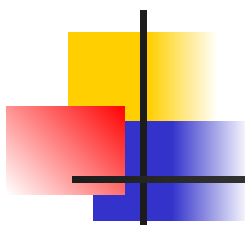
Гистограмма — это вариационный ряд, представленный в виде диаграммы, в которой различная величина частот изображается различной высотой столбиков. Гистограмма распределения представлена на рис. 3.1. На гистограмме более наглядно проявляются особенности распределения. Несколько затруднено при помощи гистограммы сравнение нескольких распределений. Поэтому разработаны другие способы графической иллюстрации особенностей распределения.



Вариационная кривая - это изображение вариационного ряда в виде кривой, ординаты которой пропорциональны частотам вариационного ряда. Вариационная кривая того же распределения представлена на рис.3.2. Вариационная кривая—очень удобный и наглядный способ иллюстрации, особенно в тех случаях, когда на одном графике желательно изобразить несколько распределений.



Графическое изображение вариационного ряда дает наглядное представление о характере распределения признака в изучаемой совокупности. Вариационный ряд можно представить в виде ступенчатой кривой, называемой гистограммой (рис.3.1). Для этого по оси абсцисс наносятся границы классов (W), а по оси ординат – частоты (f). Основанием каждого столбика является соответствующее значение класса, а высота – число особей в нем. Если соединить прямыми линиями середины всех столбиков получится вариационная кривая.



Таким образом, *вариационная кривая* – это линия, указывающая на полигон распределения вариант с учетом середины классов. Вариационная кривая используется для признаков, имеющих прерывистое значение. Двухвершинность вариационного ряда указывает на разнокачественность (разнородность) исследуемых объектов. *Гистограмма* – это ступенчатая кривая, указывающая на полигон распределения вариант с учетом границ классов; используются для признаков, имеющих непрерывный характер.

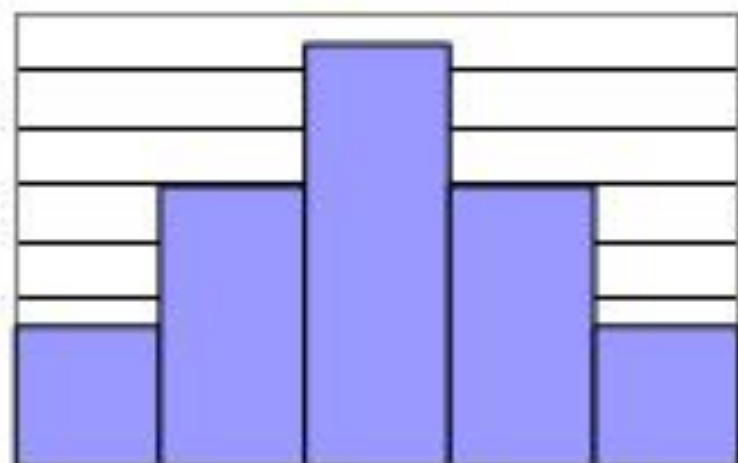


Рисунок 3.1 Гистограмма

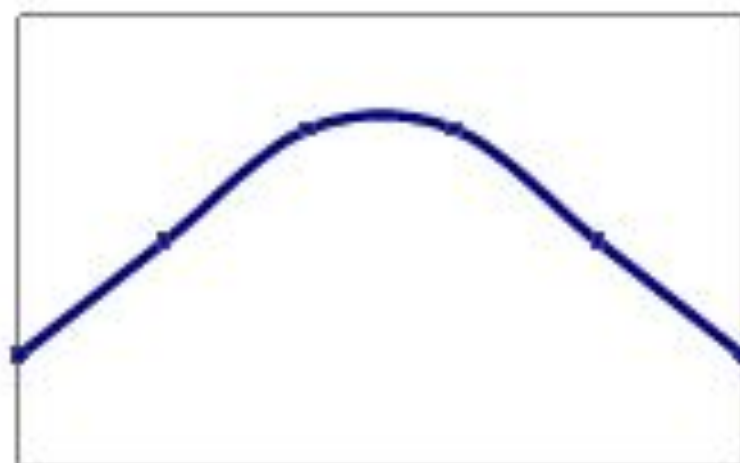


Рисунок 3.2 Вариационная кривая



Рисунок 3.3 Кумулята

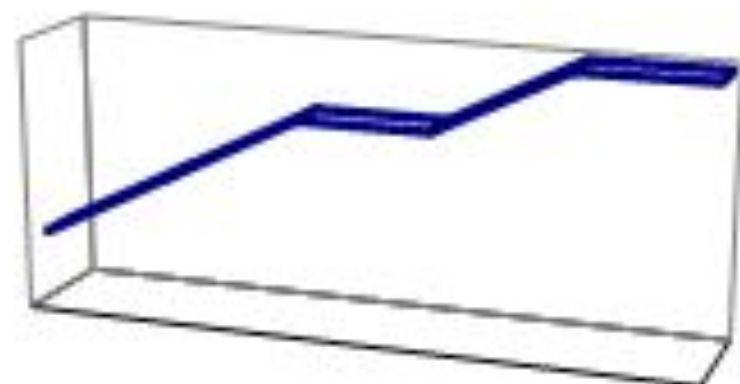



Рисунок 3.4 Огива

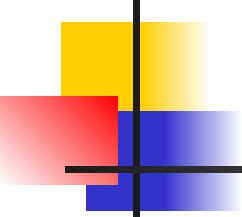
При графическом анализе симметричного вариационного ряда можно видеть следующее:

- а) около середины вариационной кривой располагается наибольшее число вариантов;
- б) распределение вариантов по обе стороны от вершины вариационной кривой примерно симметрично;
- в) варианты убывают к краям вариационного ряда.

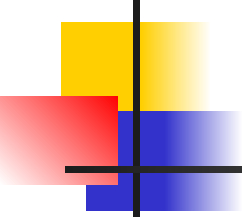
Кумулята — это изображение распределения в виде кривой, ординаты которой пропорциональны накопленным частотам вариационного ряда. Чтобы составить ряд накопленных частот, нужно к частоте первого, наименьшего класса прибавить частоту второго класса (это будет Σf_2 для второго класса), затем прибавить частоту третьего класса (это будет Σf_3 для третьего класса) и т. д. Кумулята для распределений показана на рис. 3.3.



Кумулята иногда имеет преимущество перед вариационной кривой. Некоторые методы экологометрики основаны на использовании кумуляты. К ним относятся критерий лямбда, определяющий достоверность различия двух распределений, пробит - метод, вскрывающий детали действия ядов и других агентов. По кумуляте легко находить значение различных квартилей. На рис. 3.4 показан простой способ нахождения величины первого квартиля, второго квартиля, или медианы, и третьего квартиля. Чтобы найти, например, величину первого квартиля, надо определить число дат, меньших его, т. е. 25% всей группы, затем по оси ординат найти это число, провести на этом уровне горизонтальную прямую до встречи с кумулятой, из точки пересечения опустить перпендикуляр на ось абсцисс и по шкале прочесть искомое значение.



Огива - это изображение распределения, при котором все даты расположены одна за другой в порядке возрастания. При составлении огивы значения признака откладываются по оси ординат, а по оси абсцисс откладывается ранг каждой даты в ранжированном ряду. Огива представлена на рис. 3.4.



Литература:

Основная – 1 [6-88]; 2 [т.1-11-78]; 3 [14-17].

Дополнительная – 3 [28-59]; 5 [26-84].

Контрольные вопросы:

1. Выборочный метод в экологометрике.
2. Дать определение следующим терминам и понятиям: варианта, группировка, вариационный ряд, гистограмма, вариационная кривая.
3. Сущность вариационно-статистического метода анализа.
4. Принципы построения вариационного ряда.
5. Виды графического изображения вариационного ряда.

