Лекция 9. Вариационные ряды и их графическое изображение

На лекции рассматриваются вопросы:

1. Предмет математической статистики.
2. Генеральная и выборочная совокупности.
3. Понятие вариационного ряда и его графическое изображение.

1. Предмет математической статистики

Математическая статистика (МС) — наука, изучающая методы сбора, систематизации и обработки результатов наблюдений массовых случайных явлений с целью выявления статистических закономерностей.

2. Генеральная и выборочная совокупности

В основе МС лежат понятия генеральной совокупности и выборки (выборочной совокупности).

Множество всех значений некоторой величины, то есть совокупность *всех* объектов, по которым проводится статистическое исследование, называется ***генеральной совокупностью***.

Количество объектов генеральной совокупности N называется ***объемом генеральной совокупности.***

Пример:

1) Некоторому предприятию необходимо принять 50 деталей. При этом оценивается годность деталей. Все 50 деталей образуют генеральную совокупность, то есть $N = 50$.

2) Проводится исследование средней длины льноволокна в 10 тоннах волокна. Генеральную совокупность образуют все волокна в 10 тоннах. N очень велико.

Генеральная совокупность может быть конечной и бесконечной.

Генеральную совокупность можно изучать следующими *методами*:

1) *сплошное наблюдение* — обследование всех объектов генеральной совокупности (возможно проведение в примере 1);

2) *выборочное исследование* — обследование части генеральной совокупности (целесообразно в примере 2).

Часть объектов генеральной совокупности, используемая для исследования, называется ***выборочной совокупностью*** (***выборкой***).

Число n объектов этой выборочной совокупности называется ***объемом выборки.***

Если выборка позволяет судить об исследуемом признаке генеральной совокупности в целом, то такую выборку называют ***репрезентативной (представительной)***, то есть представляющей генеральную совокупность.

Репрезентативность выборки обеспечивается объемом выборки и случайностью отбора ее элементов (например, используются генераторы случайных чисел).

Способы образования выборки:

1) *повторная* — случайно отобранный элемент после исследования возвращается в генеральную совокупность и может быть отобран повторно;

2) бесповторная — случайно отобранный элемент после исследования не возвращается в генеральную совокупность.

3. Понятие вариационного ряда и его графическое изображение

1. Пусть количественный признак X — **дискретная СВ.**

Пусть для изучения количественного признака X из генеральной совокупности извлечена выборка объемом n.

Наблюдаемые значения x_i признака X называются ***вариантами***.

На первом этапе статистической обработки производят ***ранжирование*** выборки, то есть упорядочивание чисел x_i по возрастанию.

Последовательность вариант, записанных в возрастающем порядке, называется ***дискретным вариационным рядом.***

Пример: Если изучаемым признаком X является успеваемость студентов группы (состоящей из 25 человек), то вариантами показателя успеваемости будут оценки: $x_1 = 2$, $x_2 = 3$, $x_3 = 4$, $x_4 = 5$.

2, 3, 4, 5 — вариационный ряд.

Частотой варианты x_i называется число n_i, показывающее, сколько раз эта варианта встречается в выборке.

Часто выборку записывают в виде *таблицы частот*:

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

В этом случае объем выборки $n = n_1 + n_2 + ... + n_k$, то есть $\sum_{i=1}^{k} n_i = n$.

Пример:

В результате исследования построена таблица частот:

x_i	2	3	4	5
n_i	5	10	7	3

Объем выборки $n = 5 + 10 + 7 + 3 = 25$.

Частостью (относительной частотой) варианты x_i называется число

$$w_i = \frac{n_i}{n}.$$

Оно показывает, какая доля единиц совокупности имеет варианту x_i.

Очевидно, $\sum_{i=1}^{k} w_i = 1$.

Выборку можно записать в виде таблицы относительных частот:

x_i	x_1	x_2	...	x_k
w_i	w_1	w_2	...	w_k

Пример:

В результате исследования построена таблица относительных частот:

x_i	2	3	4	5
w_i	$\frac{5}{25}$	$\frac{10}{25}$	$\frac{7}{25}$	$\frac{3}{25}$

Для наглядности представления используют графические изображения вариационных рядов.

Для изображения дискретного вариационного ряда строят полигон частот или полигон относительных частот.

Полигон частот — ломаная, соединяющая точки плоскости с координатами $(x_i; n_i)$, где $i = 1;2;...k$.

Полигон относительных частот — ломаная, соединяющая точки плоскости с координатами $(x_i; w_i)$, где $i = 1;2;...k$.

(Постройте самостоятельно полигон частот и полигон относительных частот для предыдущего примера.)

Накопленная (кумулятивная) частота m_i — количество вариант, значения которых меньше n_i.

$$m_i = n_1 + n_2 + ... + n_{i-1}.$$

Относительная накопленная частота v_i — отношение накопленной частоты к общему объему выборки, то есть

$$v_i = \frac{m_i}{n}.$$

Пример:

Построим таблицу накопленных частот и накопленных относительных частот:

x_i	2	3	4	5	6
m_i	0	5	15	22	25
$v_i = \frac{m_i}{n}$	0	$\frac{5}{25} = 0{,}2$	$\frac{15}{25} = 0{,}6$	$\frac{22}{25} = 0{,}88$	$\frac{25}{25} = 1$

Кумулята частот — ломаная, соединяющая точки с координатами $(x_i; m_i)$, где $i = 1;2;...k$.

Кумулята относительных частот — ломаная, соединяющая точки с координатами $(x_i; v_i)$, где $i = 1;2;...k$.

(Постройте самостоятельно кумуляту частот и кумуляту относительных частот для предыдущего примера.)

Эмпирической функцией распределения $F^*(x)$ называют функцию, значение которой в точке x равно накопленной относительной частоте, то есть

$$F^*(x) = v_x = \frac{m_x}{n}.$$

Пример:

Найдем эмпирическую функцию распределения:

$$F^*(x) = \begin{cases} 0, \text{ если } x \in (-\infty; 2], \\ 0{,}2, \text{ если } x \in (2; 3], \\ 0{,}6, \text{ если } x \in (3; 4], \\ 0{,}88, \text{ если } x \in (4; 5], \\ 1, \text{ если } x \in (5; +\infty). \end{cases}$$

(Постройте самостоятельно график эмпирической функции распределения).

Свойства эмпирической функции распределения:

1) $E(F^*(x)) = [0;1]$.
2) $F^*(x)$ — неубывающая функция.
3) $F^*(x) = 0$ при $x \le x_{\min}$,
   $F^*(x) = 1$ при $x > x_{\max}$.

Эмпирическая функция распределения — аналог теоретической интегральной функции распределения $F(x)$ в теории вероятностей. Она тем ближе к ней, чем больше *n*.

2. Пусть количественный признак *X* — **непрерывная СВ**, принимающая значения из некоторого промежутка.

Пример: Изучается признак *X* — затраты времени рабочим на обработку одной детали (мин.). Это есть непрерывная случайная величина *X*. Обследовано 100 рабочих, то есть $n = 100$. Получено наименьшее значение варианты $x_{\min} = 22$, наибольшее значение варианты $x_{\max} = 34$.

При исследовании непрерывной СВ диапазон наблюдаемых данных делят на частичные интервалы $[x_{i-1}; x_i)$. Если варианта находится на границе интервала, то ее присоединяют к правому интервалу.

Рекомендуется оптимальную длину частичного интервала (интервальную разность) определять по формуле Стерджерса:

$$h = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \lg n}$$

Тогда количество интервалов:

$$k = \frac{x_{\max} - x_{\min}}{h}.$$

Получим *равновеликий* интервальный ряд. В интервальном вариационном ряде могут встречаться интервалы разной длины. Такие ряды называют *неравновеликими*.

Последовательность частичных интервалов, записанных в возрастающем порядке, называется ***интервальным вариационным рядом.***

Пример:

Определим оптимальную длину частичного интервала по формуле Стерджерса:

$$h = \frac{34 - 22}{1 + 3{,}322 \lg 100} = \frac{12}{7{,}644} \approx 2.$$

Тогда количество интервалов:

$$k = \frac{34 - 22}{2} = 6.$$

Частотой i-го интервала n_i называют число наблюдаемых значений в этом интервале.

Частостью или относительной частотой i-го интервала w_i называют отношение

$$w_i = \frac{n_i}{n}.$$

Частичные интервалы и соответствующие им частоты или относительные частоты записывают в виде таблицы:

Таблица частот:

$[x_{i-1};x_i)$	$[x_0;x_1)$	$[x_1;x_2)$	...	$[x_{k-1};x_k]$
n_i	n_1	n_2	...	n_k

Таблица относительных частот:

$[x_{i-1};x_i)$	$[x_0;x_1)$	$[x_1;x_2)$	...	$[x_{k-1};x_k]$
w_i	w_1	w_2	...	w_k

Пример: При исследовании получили интервальный вариационный ряд. Составили таблицу частот и относительных частот:

$[x_{i-1};x_i)$	$[22;24)$	$[24;26)$	$[26;28)$	$[28;30)$	$[30;32)$	$[32;34]$
n_i	2	12	34	40	10	2
w_i	0,02	0,12	0,34	0,40	0,10	0,20

Первый и последний интервалы могут не иметь соответственно левой и правой границы, то есть быть открытыми. В процессе обработки данных открытые интервалы приходится условно закрывать. При этом величину первого интервала принимают равной величине второго, а величину последнего — величине предпоследнего.

Пример: Изучается признак X — сумма вклада, внесенная клиентом в банк в течение месяца. Это есть непрерывная случайная величина X.

Таблица частот:

X — сумма вклада в у. е.	до 500	500—1000	1000—2000	2000—3000	свыше 3000
n_i	27	11	8	8	2

При этом первый и последний интервалы — открытые. Условно закроем границы открытых интервалов.

Длина второго интервала равна:

$$1000-500=500.$$

Следовательно, нижняя граница первого интервала:

$$500-500=0.$$

Длина предпоследнего интервала равна:

$$3000-2000=1000.$$

Следовательно, верхняя граница последнего интервала:

$$3000+1000=4000.$$

В результате получим следующую таблицу частот:

X — сумма вклада в у. е.	0—500	500—1000	1000—2000	2000—3000	3000—4000
n_i	27	11	8	8	2

Данный интервальный ряд является неравновеликим, так как длины интервалов не одинаковы.

Интервальный ряд изображают в виде гистограммы частот или гистограммы относительных частот.

Гистограмма частот — ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной h_i, а высоты равны ***плотности частот*** $\frac{n_i}{h_i}$.

Площадь гистограммы частот $S_ч$ равна объему выборки:

$$S_ч = n.$$

Гистограмма относительных частот — ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной h_i, а высоты равны ***плотности относительных частот*** $\frac{w_i}{h_i}$.

Площадь гистограммы относительных частот $S_{отн.ч}$ равна единице:

$$S_{отн.ч}=1.$$

В теории вероятностей гистограмме относительных частот соответствует график плотности распределения вероятностей (дифференциальной функции распределения) $f(x)$. Поэтому гистограмму можно использовать для подбора закона распределения генеральной совокупности.

Пример: Постройте самостоятельно гистограмму частот:

$[x_{i-1};x_i)$	$[22;24)$	$[24;26)$	$[26;28)$	$[28;30)$	$[30;32)$	$[32;34]$
$\frac{n_i}{h_i}$	1	6	17	20	5	1

Пример: Постройте самостоятельно гистограмму относительных частот:

$[x_{i-1};x_i)$	$[22;24)$	$[24;26)$	$[26;28)$	$[28;30)$	$[30;32)$	$[32;34]$
$\frac{w_i}{h_i}$	0,01	0,06	0,17	0,20	0,50	0,01

Накопленная (кумулятивная) частота m_i для интервального ряда вычисляется следующим образом:

Накопленная частота нижней границы первого интервала равна нулю.

Накопленная частота верхней границы первого интервала равна частоте этого интервала.

Накопленная частота верхней границы второго интервала равна сумме частот первого и второго интервала.

Далее аналогично.

Относительные накопленные частоты v_i интервального ряда находятся по формуле:

$$v_i = \frac{m_i}{n}$$

При построении ***кумуляты*** частот (относительных частот) интервального ряда соединяют ломаной точки, абсциссами которых служат концы интервалов, а ординатами — накопленные частоты (относительные накопленные частоты).

Пример: Вычислим накопленные частоты m_i и относительные накопленные частоты:

x_i	22	24	26	28	30	32	34
m_i	0	2	2+12=14	14+34=48	48+40=88	88+10=98	98+2=100
v_i	0	0,02	0,14	0,48	0,88	0,98	1

(Постройте самостоятельно кумуляту частот и кумуляту относительных частот).

График кумуляты относительных частот интервального ряды и есть график эмпирической функции распределения.

Пример:

Результаты обследования 20 семей по числу членов оказались такими: 2; 5; 3; 4; 1; 3; 6; 2; 4; 3; 4; 1; 3; 5; 2; 3; 4; 3; 4; 3. Получить по этим данным вариационный ряд и построить полигон распределения относительных частот.

Решение

Проводим ранжирование данного ряда. Для этого запишем результаты наблюдений в порядке возрастания вариант:

1; 1; 2; 2; 2; 3; 3; 3; 3; 3; 3; 3; 4; 4; 4; 4; 4; 5; 5; 6.

По ранжированному ряду определяем частоты различных вариант. Варианта $x_1 = 1$ встречается в заданном ряду два раза. Следовательно, ее частота равна $n_1 = 2$. Варианта $x_2 = 2$ встречается три раза, следовательно, $n_2 = 3$; аналогично для вариант $x_3 = 3$, $x_4 = 4$, $x_5 = 5$, $x_6 = 6$ получаем соответственно: $n_3 = 7$, $n_4 = 5$, $n_5 = 2$, $n_6 = 1$.

Определяем относительные частоты наблюдавшихся в выборке вариант. Они равны отношению соответствующей частоты варианты к общему числу наблюдений:

$$w_i = \frac{n_i}{n}.$$

Так как объем выборки $n = 20$, то относительная частота варианты $x_1 = 1$ будет равна

$$w_1 = \frac{n_1}{n} = \frac{2}{20} = 0{,}10.$$

Аналогично

$$w_2 = \frac{n_2}{n} = \frac{3}{20} = 0{,}15,$$

$$w_3 = \frac{n_3}{n} = \frac{7}{20} = 0{,}35,$$

$$w_4 = \frac{n_4}{n} = \frac{5}{20} = 0{,}25,$$

$$w_5 = \frac{n_5}{n} = \frac{2}{20} = 0{,}10,$$

$$w_6 = \frac{n_6}{n} = \frac{1}{20} = 0{,}05.$$

Проверяем правильность расчетов. Для этого суммируем вычисленные относительные частоты:

$$\sum w_i = 0{,}10 + 0{,}15 + 0{,}35 + 0{,}25 + 0{,}10 + 0{,}05 = 1{,}00.$$

Сумма всех относительных частот равна единице, следовательно, вычисления сделаны верно.

Результаты выполненных расчетов сводим в таблицу, которая называется вариационным рядом:

Таблица. Вариационный ряд

Значения варианты x_i	Частота варианты n_i	Относительная частота варианты w_i
1	2	0,10
2	3	0,15
3	7	0,35
4	5	0,25
5	2	0,10
6	1	0,05
	$\sum n_i = 20$	$\sum w_i = 1,00$

Построим многоугольник или полигон распределения относительных частот. В прямоугольной системе координат по горизонтальной оси откладываем значения вариант x_i, а по вертикальной оси — относительные частоты w_i. Точки с координатами (x_i, w_i) соединяем отрезками прямых. Полученная ломанная является многоугольником или полигоном распределения относительных частот.

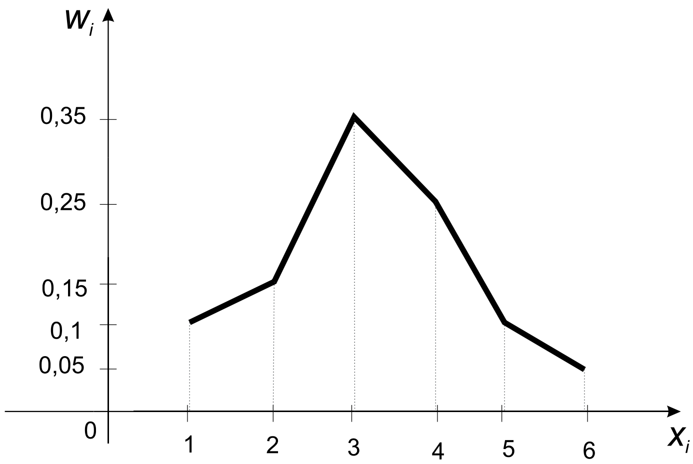

Рис. Полигон относительных частот

Пример:

По некоторому региону известна ежедневная прибыль (тысяч рублей) двадцати однотипных торговых предприятий: 11,3; 10,2; 13,9; 10,7; 11,8; 8,2; 12,4; 9,6; 13,1; 10,6; 6,3; 11,3; 10,2; 15,1; 10,5; 11,0; 15,1; 11,6; 10,4; 11,7. Составить интервальный ряд распределения и построить гистограмму распределения плотности относительных частот.

Решение

Запишем исходные данные в виде ранжированного ряда: 6,3; 8,2; 9,6; 10,2; 10,4; 10,5; 10,6; 10,7; 11,0; 11,3; 11,3; 11,6; 11,7; 11,8; 12,4; 13,1; 13,9; 15,1; 15,1.

Из этого ряда видно, что диапазон изменения вариант в выборке составляет 6-16. Этот диапазон разобьем на несколько интервалов. Размер интервала выбирается произвольно, но следует иметь в виду, что чем меньше интервал, тем точнее результаты. В нашем случае принимаем размер интервала равным 2 единицам, то есть $\Delta x_i = 2$. Получаем пять интервалов: первый 6-8, второй 8-10, третий 10-12, четвертый 12-14, пятый 14-16.

Определяем частоту попадания вариант выборки в каждый интервал. В первый интервал попадает одно значение ряда — 6,3, поэтому $n_1 = 1$. Во второй интервал попадают два значения — 8,2 и 9,6, поэтому $n_2 = 2$. Аналогично находим $n_3 = 12$, $n_4 = 3$, $n_5 = 2$.

Определяем относительные частоты попадания вариант выборки в каждый интервал:

в первый интервал — $w_1 = \dfrac{n_1}{n} = \dfrac{1}{20} = 0{,}05$;

во второй интервал — $w_1 = \dfrac{n_2}{n} = \dfrac{2}{20} = 0{,}10$;

в третий интервал — $w_3 = \dfrac{n_3}{n} = \dfrac{12}{20} = 0{,}60$;

в четвертый интервал — $w_4 = \dfrac{n_4}{n} = \dfrac{3}{20} = 0{,}15$;

в пятый интервал — $w_5 = \dfrac{n_5}{n} = \dfrac{2}{20} = 0{,}10$.

Сумма $\sum w_i = 1$, следовательно, вычисления выполнены верно.

Определим плотность относительных частот вариант как отношение относительной частоты w_i к длине интервала Δx_i, то есть

$$p_i = \frac{w_i}{\Delta x_i}.$$

Для первого интервала — $p_1 = \dfrac{w_1}{\Delta x_1} = \dfrac{0{,}05}{2} = 0{,}025$;

для второго интервала — $p_2 = \dfrac{w_2}{\Delta x_2} = \dfrac{0{,}10}{2} = 0{,}050$;

для третьего интервала — $p_3 = \frac{w_3}{\Delta x_3} = \frac{0{,}60}{2} = 0{,}300$;

для четвертого интервала — $p_4 = \frac{w_4}{\Delta x_4} = \frac{0{,}15}{2} = 0{,}075$;

для пятого интервала — $p_5 = \frac{w_5}{\Delta x_5} = \frac{0{,}10}{2} = 0{,}050$.

Результаты выполнения расчетов сводим в таблицу:

Таблица. Интервальный ряд распределения прибыли

Интервал значений прибыли Δx_i	6-8	8-10	10-12	12-14	14-16	
Частота вариант n_i	1	2	12	3	2	$\sum n_i = 20$
Относительные частоты w_i	0,05	0,10	0,60	0,15	0,10	$\sum w_i = 1{,}00$
Плотность относительных частот p_i	0,025	0,050	0,300	0,075	0,050	

Построим гистограмму, показывающую зависимость плотности относительных частот от значения вариант. По горизонтальной оси наносим шкалу возможных значений вариант, по вертикальной оси — плотность относительных частот; величину относительной плотности считаем постоянной внутри соответствующего интервала. Получаем столбчатую диаграмму, называемую гистограммой распределения плотности относительных частот:

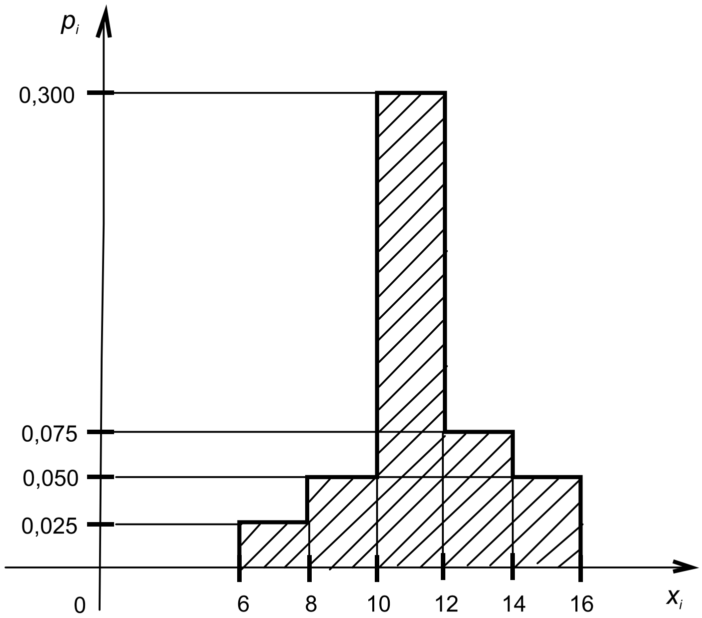

Рис. Гистограмма плотности относительных частот.