

ЛЕКЦИОННОЕ ЗАНЯТИЕ

Графическое представление совокупности

Одномерная совокупность, представленная в виде вариационного ряда, может быть изображена в виде полигона, гистограммы, кумуляты, кривой Лоренца, огивы и т.п.

Полигональная ломаная, или полигон, или многоугольник распределения – строится в прямоугольной системе координат.

Полигон может быть построен для дискретного вариационного ряда и для интервального. Полигон для дискретного вариационного ряда строится следующим образом:

По оси абсцисс откладывают значения вариантов, а по оси ординат – значения частот (или частостей).

Полученные на пересечении этих значений точки соединяют отрезками прямой. Такой график, как очевидно, можно построить только для случая часто повторяющихся вариантов в совокупности.

На практике же чаще всего случается, что наблюдений не слишком много, а потому и повторяющихся значений либо мало, либо вообще повторений нет. Все имеющиеся значения вариантов совокупности встречаются только единожды. Частоты равны единице. График теряет смысл.

Поэтому для небольших по объему совокупностей рациональнее строить полигональную ломаную по интервальному вариационному ряду.

Но, как уже говорилось раньше, и для больших по объему совокупностей проще построить полигон по интервальному вариационному ряду.

Для этого по оси абсцисс откладываются значения середин интервала, – среднеинтервальные значения совокупности, а по оси ординат, как всегда, – частоты (или частости).

Гистограмма распределения строится только для совокупности, представленной в виде интервального вариационного ряда.

Гистограмма также строится в прямоугольной системе координат.

В отличие от полигона, для гистограммы на оси абсцисс откладываются отрезки, соответствующие интервалам значений. На каждом отрезке, как на основании строится прямоугольник, высотой которого служит значение частоты, соответствующей данному интервалу.

Получим как бы ступенчатую гистограмму. При таком построении допускается, что распределение вариантов внутри интервала равномерно.

Можно представить себе, что при последовательном делении интервалов, ступенчатая гистограмма превратится в плавную кривую. Такая кривая носит, название кривой распределения.

Кумулята или кумулятивная ломаная, выполняется в прямоугольной системе координат. По оси абсцисс откладываются значения признака (варианты), а по оси ординат – соответствующие накопленные частоты. Полученные точки пересечений соединяются отрезками прямой.

Кумулятивную ломаную можно построить как дискретного вариационного ряда, так и для интервального.

Для интервального вариационного ряда по оси абсцисс откладываются среднеинтервальные значения. Нижней границе первого интервала соответствует частота равная нулю, а верхней границе последнего интервала – сумма всех частот или общее количество наблюдений.

При выборе соотношений между масштабами по осям абсцисс и ординат целесообразно использовать правило «золотого сечения». График располагается в (прямоугольнике, размеры которого пропорциональны 5:8 или 3:4, и линии графика занимают всю площадь.

Сравнение этих графиков показывает, что переход к интервальным значениям значительно сглаживает график и выявляет сущность

совокупности.

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ

Построим графики для совокупности, представленной в виде вариационного ряда. Воспользуемся таблицами данных: дискретный вариационный ряд – табл. 1, интервальный вариационный ряд – табл. 2.

Таблица 1

x_i	1609	2004	3615	6237	7050	9249	N
m_i	1	1	3	2	2	2	11

Таблица 2

i	$x_{min} \dots x_{max}$	x_i	m_i	n_i	M_i
1	1500 - 3500	2500	2	0,18	2
2	3500 - 5500	4500	3	0,27	5
3	5500 - 7500	6500	4	0,36	9
4	7500 - 9500	8500	2	0,18	11

Построим полигональную ломаную – «полигон». По оси абсцисс отложим значения вариант x_i , а по оси ординат – значения частот этих вариант m_i .

Из табл. 2 следует, что наименьшее $x_i = x_{min} = 1609$, а наибольшее $x_i = x_{max} = 9249$, поэтому на оси абсцисс отложим 1500 и 9500, т. е. значения, включающие min и max. И на полученном отрезке оси отметим точки, соответствующие значениям всех вариант таблицы 2.

Как видим по таблице 1, наибольшее значение частоты $m_i = 3$.

Поэтому ось ординат достаточно разделить на 3 равных части (рис. 1). А масштаб графика выберем так, чтобы выдерживалось «золотое» соотношение: 5:8 или 3:4.

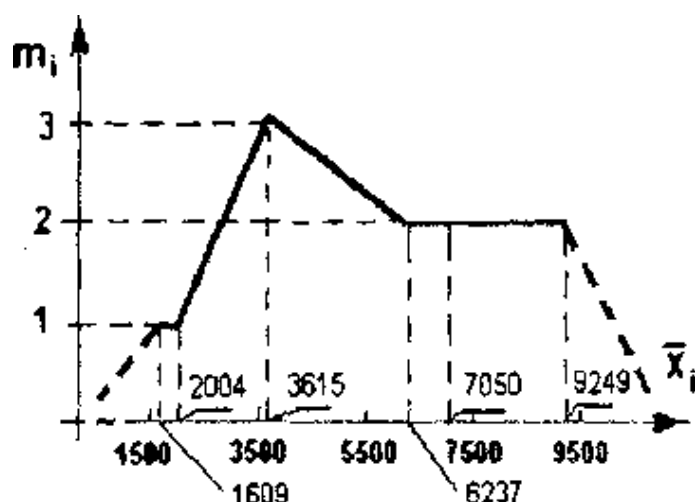


Рис. 1. Полигональная ломаная

Аналогично построим и полигон для совокупности, представленной в виде интервального вариационного ряда.

Используем данные табл. 2. По оси абсцисс графика откладывается значение столбца 3 из табл. 2. Среднеинтервальные значения x_i , варьируют в пределах 2500...8500. Эти значения и должны служить границами графика (см. рис. 1). Наибольшая частота $m_{max} = 4$. Поэтому достаточно ось ординат разделить на 4 равных отрезка.

Нанесем точки на ось абсцисс: 2500, 4500, 6500 и 8500. На оси ординат отложим 2, 3, 4, 2.

На пересечениях этих значений отметим точки полигона и соединим точки отрезками прямой. Можно добавить в таблице 2 две строки – в начале и в конце таблицы:

1	2	3	4	...
0	< 1500	500	0	...
5	> 9500	10500	0	...

Эти добавления дают нам возможность дополнить полигональную ломаную отрезками прямой до пересечения с осью абсцисс. Нанесем эти отрезки пунктиром.

Построим гистограмму.

Для этого используем интервальный вариационный ряд.

На оси абсцисс отложим отрезки, соответствующие интервалам вариационного ряда. На них, как на основании, построим прямоугольники (столбики), высотой, пропорциональной частоте.

Если представить, что интервалы последовательно и многократно делят на два, тогда столбики гистограммы становятся все тоньше и тоньше. И в пределе верхние отрезки столбиков превращаются в точки и получается плавная огибающая линия. Эта линия и носит название кривой распределения. Но этот процесс требует большого количества наблюдений.

В последующих частях анализа используются эти результаты.

По гистограмме (рис. 2) можно графически определить значение M_0 = 6100.

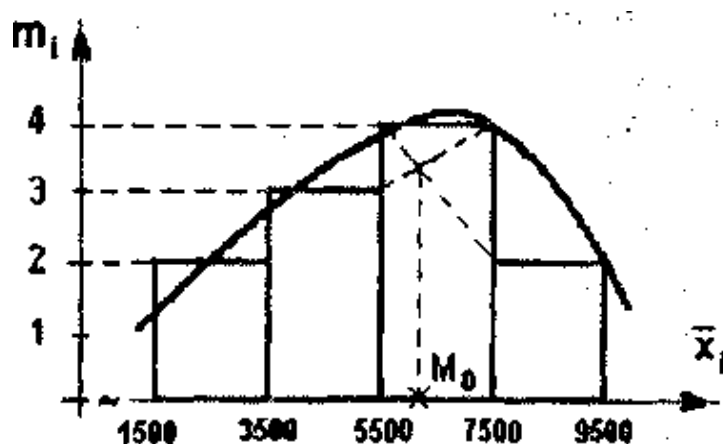


Рис. 2. Гистограмма распределения

Построим кумуляту. Кумулятивная ломаная (рис. 3) строится как для дискретного, так и для интервального ряда.

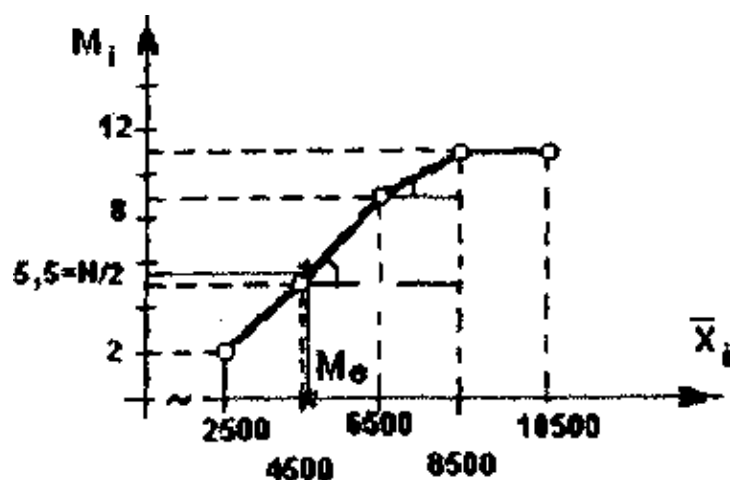


Рис. 3. Кумулятивная ломаная

Не усложняя процесс анализа статистической совокупности, построим кумюляту только для интервального вариационного ряда.

Отложим по оси абсцисс значение середин интервала (среднеинтервальные), а по оси ординат – накопленные частоты и соединим точки пересечения отрезками прямой.

Что можно сказать о совокупности по этому графику? Можно сказать, что основное количество наблюдений находится в средних интервалах, так как угол между осью абсцисс и отрезком кумюляты в этих интервалах больше, чем в последнем интервале. Ведь очевидно, что если бы интервал имел частоту, равную нулю, то соответствующий отрезок кумюлятивной ломаной был бы параллелен оси абсцисс.

Так, если к нашей совокупности, соответственно интервальному ряду, добавить еще один интервал, то частота его будет равна 0.

На кумюляте можно графически определить значение Me .

Как известно, медиана – это значение признака, находящегося посередине совокупности. В нашем примере всего одиннадцать значений, $N = 11$, значит, $N/2 = 5,5$. Откладываем это значение на оси ординат, проводим горизонтальную линию (параллельную оси абсцисс) до пересечения с графиком кумюляты, из точки пересечения опускаем перпендикуляр на ось абсцисс. Это значение и есть значение медианы – $Me_{граф}$. В данном примере $Me_{граф} = 5100$.

Выводы

По графику полигональной ломаной, построенной по дискретным значениям, нельзя сделать серьезного вывода, так как наблюдений немного и они имеют большой разброс, что и сказалось на графике. Несколько сглаженные интервальные значения позволили построить гистограмму достаточно симметричного вида.

Кумулята свидетельствует о том, что в построении вариационного ряда нет пустых интервалов, т. е. интервалов с частотой, равной нулю.

Таким образом, можно считать, что данная совокупность может быть включена в дальнейшее исследование.