

Вариационные ряды распределения и их числовые определения

Преподаватель математики МИПК
им. И. Федорова Епихина Е.В.

Содержание

1. Статистические ряды распределения. Определение. Виды. Примеры.
2. Атрибутивный ряд распределения.
3. Вариационный ряд распределения.
4. Графическое изображение рядов распределения.
5. Полигон.
6. Гистограмма.
7. Кумулята
8. Огива
9. Статистические таблицы.

Статистический ряд распределения – это упорядоченное распределение единиц совокупности на группы по определённому варьирующему признаку.

В зависимости от признака, положенного в основу образования ряда распределения, различают **атрибутивные** и **вариационные** ряды распределения.

Ряд распределения — представляет собой упорядоченное распределение единиц изучаемой совокупности на группы по определенному варьирующему признаку.

- * **Атрибутивными** — называют ряды распределения, построенные по качественными признакам. Ряд распределения принято оформлять в виде таблиц.

N п/п	Виды юридической помощи	Число случаев юридической помощи	
		всего, тыс.	в % к итогу
1	Устные советы	5109	69,43
2	Составление документов	991	13,47
3	Поручения по ведению уголовных дел	1021	13,87
4	Поручения по ведению Гражданских дел	238	3,23
ВСЕГО		7359	100,00

Таблица 1 - Распределение видов юридической помощи, оказанной адвокатами гражданам одного из регионов РФ.

- * **Вариационными** рядами называют ряды распределения, построенные по количественному признаку. Любой вариационный ряд состоит из двух элементов: вариантов и частот.
- * **Вариантами** считаются отдельные значения признака, которые он принимает в вариационном ряду.
- * **Частоты** – это численности отдельных вариантов или каждой группы вариационного ряда, т.е. это числа, показывающие, как часто встречаются те или иные варианты в ряду распределения. Сумма всех частот определяет численность всей совокупности, её объём. Обозначаются (f_i)
- * **Частостями** называются частоты, выраженные в долях единицы или в процентах к итогу. Соответственно сумма частостей равна 1 или 100 %. Обозначаются (W_i)

Виды вариационный рядов

- * В зависимости от характера вариации признака различают **дискретные** и **интервальные** вариационные ряды.
- * Вариационный ряд называется **дискретным**, если любые его варианты отличаются на постоянную величину, и **интервальным**, если варианты могут отличаться один от другого на сколь угодно малую величину. Интервалы в ряду могут быть как равными, так и неравными. Это зависит от характера статистических данных и задач исследования.

* Пример дискретного вариационного ряда:

Таблица 2 - Распределение семей по числу занимаемых комнат в отдельных квартирах в 1989 г. в РФ.

№ П/п	Группы семей, проживающих в квартирах с числом комнат	Число семей	
		всего, тыс.ед.	в % к итогу
1	1	4064	16,3
2	2	12399	49,7
3	3	7659	30,7
4	4 и более	832	3,3
ВСЕГО		24954	100,00

В первой колонке таблицы представлены варианты дискретного вариационного ряда, во второй – помещены частоты вариационного ряда, в третьей – показатели частоты.

- * При расчете средней арифметической в интервальном ряду за значение варианты принимается середина интервала. Середина интервала вычисляется как среднее арифметическое его границ.
- * Медиана (M_e) Срединное значение варьирующего признака в упорядоченном (ранжированном) ряду. Применяется в случаях, когда совокупность статистических данных неоднородна (асимметрична), поскольку M_e менее чувствительна к средним значениям ряда, чем средняя арифметическая.
- * Мода (M_o) Наиболее часто встречающаяся в ряду варианта. В интервальном вариационном ряду определяется модальный интервал. M_o используется для характеристики среднего уровня в неоднородных совокупностях, как и медиана.
- * Размах вариации (R) - разность максимального и минимального значений периода в вариационном ряду.

$$R = x_{\max} - x_{\min}$$

Зависит от случайных колебаний выборки, т.е. для вычисления R используются лишь крайние значения варианты.

- * Дисперсия - представляет собой среднюю арифметическую квадратов отклонений всех вариантов от их средней арифметической.

Графическое изображение рядов распределения

Наглядно ряды распределения представляются при помощи графических изображений.

Ряды распределения изображаются в виде:

- * Полигона
- * Гистограммы
- * Кумуляты
- * Огивы

Полигон

При построении полигона на горизонтальной оси (ось абсцисс) откладывают значения варьирующего признака, а на вертикальной оси (ось ординат) — частоты или частоты.

Домохозяйства, состоящие из:	одного человека	дзух человек	трех человек	5 или более	всего
Число домохозяйств в %	19,2	25,2	22,6	20,5	100,0



Полигон на рис. 6.1 построен по данным микропереписи населения России в 1994 г.

- * **Условие:** Приводятся данные о распределении 25 работников одного из предприятий по тарифным разрядам: 4; 2; 4; 6; 5; 6; 4; 1; 3; 1; 2; 5; 2; 6; 3; 1; 2; 3; 4; 5; 4; 6; 2; 3; 4
- * **Задача:** Построить дискретный вариационный ряд и изобразить его графически в виде полигона распределения.
- * **Решение:** В данном примере вариантами является тарифный разряд работника. Для определения частот необходимо рассчитать число работников, имеющих соответствующий тарифный разряд.
- * Для построения полигона распределения (рис 1) по оси абсцисс (X) откладываем количественные значения варьирующего признака — варианты, а по оси ординат — частоты или частости.

Тарифный разряд X_j	Число работников f_j
1	3
2	5
3	4
4	6
5	3
6	4
Итого:	25

Полигон распределения (рис 1)



Гистограмма

- * Если значения признака выражены в виде интервалов, то такой ряд называется интервальным.
- * Интервальные ряды распределения изображают графически в виде гистограммы.
- * Гистограмма – график, на котором ряд изображен в виде смежных друг с другом столбиков. Для построения гистограммы по оси абсцисс указывают значения границ интервалов и на их основании строят прямоугольники, высота которых пропорциональна частотам (или частостям).



Рис. 6.2. Распределение населения России по возрастным группам

Все население	В том числе в возрасте								
	до 10	10-20	20-30	30-40	40-50	50-60	60-70	70 и старше	Всего
Численность населения	12,1	15,7	13,6	16,1	15,3	10,1	9,8	7,3	100,0

* На рис. 6.2. изображена гистограмма распределения населения России в 1997 г. по возрастным группам.

* Условие: Приводится распределение 30 работников фирмы по размеру месячной заработной платы

* Задача: Изобразить интервальный вариационный ряд графически в виде гистограммы и кумуляты.

* Решение:

1. Неизвестная граница открытого (первого) интервала определяется по величине второго интервала: $7000 - 5000 = 2000$ руб. С той же величиной находим нижнюю границу первого интервала: $5000 - 2000 = 3000$ руб.

2. Для построения гистограммы в прямоугольной системе координат по оси абсцисс откладываем отрезки, величины которых соответствуют интервалам вариационного ряда.

Эти отрезки служат нижним основанием, а соответствующая частота (частость) — высотой образуемых прямоугольников.

3. Построим гистограмму:

Размер заработной платы руб. в месяц	Численность работников чел.
до 5000	4
5000 — 7000	12
7000 — 10000	8
10000 — 15000	6
Итого:	30

Гистограмма



Кумулята

- Распределение признака в вариационном ряду по накопленным частотам (частостям) изображается с помощью кумуляты.
- * Кумулята или кумулятивная кривая в отличие от полигона строится по накопленным частотам или частостям. При этом на оси абсцисс помещают значения признака, а на оси ординат — накопленные частоты или частости .
- * Для построения кумуляты необходимо рассчитать накопленные частоты (частости). Они определяются путем последовательного суммирования частот (частостей) предшествующих интервалов и обозначаются S . Накопленные частоты показывают, сколько единиц совокупности имеют значение признака не больше, чем рассматриваемое.

- * Рассчитаем накопленные частоты:
- * Накопленная частота первого интервала рассчитывается следующим образом: $0 + 4 = 4$, для второго: $4 + 12 = 16$; для третьего: $4 + 12 + 8 = 24$ и т.д.

Размер заработной платы руб в месяц X_i	Численность работников чел. f_i	Накопленные частоты S
до 5000	4	4
5000 — 7000	12	16
7000 — 10000	8	24
10000 — 15000	6	30
Итого:	30	-

- * При построении кумуляты накопленная частота (частость) соответствующего интервала присваивается его верхней границе:

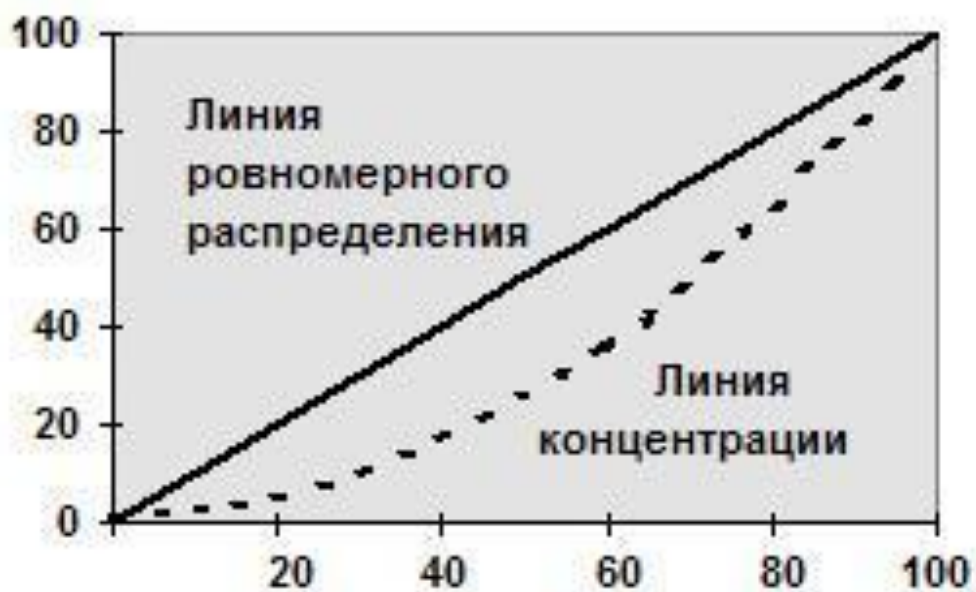
Кумулята



Огива

- * Огива строится аналогично кумуляте с той лишь разницей, что накопленные частоты помещают на оси абсцисс, а значения признака — на оси ординат.
- * Разновидностью кумуляты является кривая концентрации или график Лоренца. Для построения кривой концентрации на обе оси прямоугольной системы координат наносится масштабная шкала в процентах от 0 до 100. При этом на оси абсцисс указывают накопленные частоты, а на оси ординат — накопленные значения доли (в процентах) по объему признака.
- * Равномерному распределению признака соответствует на графике диагональ квадрата (рис. 6.4). При неравномерном распределении график представляет собой вогнутую кривую в зависимости от уровня концентрации признака.

Огиба



6.4. Кривая концентрации

Статистические таблицы

- * Статистическая таблица – это особый способ краткой и наглядной записи сведений об изучаемых общественных явлениях. Статистическая таблица позволяет охватить материалы статистической сводки в целом, она также является системой мыслей об исследуемом объекте, излагаемых цифрами на основе определенного порядка в расположении систематизированной информации.

- * Условие: Приведены данные о размерах вкладов 20 физических лиц в одном банке (тыс.руб) 60; 25; 12; 10; 68; 35; 2; 17; 51; 9; 3; 130; 24; 85; 100; 152; 6; 18; 7; 42.
- * Задача: Построить интервальный вариационный ряд с равными интервалами.
- * Решение:
 1. Исходная совокупность состоит из 20 единиц ($N = 20$).
 2. По формуле Стерджесса определим необходимое количество используемых групп: $n = 1 + 3,322 * \lg 20 = 5$
 3. Вычислим величину равного интервала: $i = (152 - 2) / 5 = 30$ тыс.руб
 4. Расчленим исходную совокупность на 5 групп с величиной интервала в 30 тыс.руб.
 5. Результаты группировки представим в таблице:

Размер вкладов тыс.руб X_j	Число вкладов f_j	Число вкладов в % к итогу W_j
2 — 32	11	55
32 — 62	4	20
62 — 92	2	10
92 — 122	1	5
122 — 152	2	10
Итого:	20	100

При такой записи непрерывного признака, когда одна и та же величина встречается дважды (как верхняя граница одного интервала и нижняя граница другого интервала), то эта величина относится к той группе, где эта величина выступает в роли верхней границы.