

# Генеральная совокупность и выборка. Дискретные и интервальные вариационные ряды.

Преподаватель: Нургалиева  
А.К.

## КЛЮЧЕВЫЕ ПОНЯТИЯ

Выборка, генеральная совокупность, статистический ряд, частота, гистограмма

При проведении экспериментов часто получают информацию очень большого объема. Например, балловые результаты ЕНТ учащихся в городе или районе, о размерах вкладов населения в банках Казахстана, масса тела и рост призывников на военную службу из конкретной области РК, количество посещения супермаркета в течение дня, и т. д.

Также статистика занимается: сбором и хранением информации, разработкой различных прогнозов, оценкой их достоверности и т. д. Если одна из основных задач математической статистики состоит в надлежащей обработке полученной информации, тогда и остальные задачи достижимы.

Порядок преобразования первоначально полученной информации примерно следующий:

- данные измерений (опыта) упорядочивают и группируют;
- после группировки составляют таблицы измерений данных;
- по таблице распределения строят графики распределения данных;
- составляют своего рода паспорт данного измерения, в котором собрано небольшое количество основных числовых характеристик полученных измерений.

*Генеральной совокупностью* называется совокупность всех подлежащих изучению объектов или возможных результатов всех наблюдений, производимых в одинаковых условиях над одним объектом.

*Выборочной совокупностью* или *выборкой* называется совокупность объектов или результатов наблюдения над объектом, отобранных случайным образом из генеральной совокупности.

*Объемом выборки* называется число объектов или наблюдений в выборке.

Значения выборки называют *наблюдаемыми значениями случайной величины  $X$* .

Для получения достоверных, надежных выводов выборка должна быть достаточно представительной по объему. Большая выборка — это неупорядоченное множество чисел. Для исследования выборку приводят к наглядному упорядоченному виду.

**вы знаете:**

Вариационные ряды состоят из двух элементов: частоты и варианты.

*Вариантой* называют отдельное значение варьируемого признака, которое он принимает в ряду распределения.

*Частота* — это численность отдельных вариантов или каждой группы вариационного ряда.

Пусть дан ряд  $x_1, x_2, x_3, \dots, x_k$ . Здесь варианта  $x_1$  встречается  $n_1$  раз, варианта  $x_2$  встречается  $n_2$  раз, варианта  $x_3$  встречается  $n_3$  раз и т. д.

Тогда равенство  $n_1 + n_2 + n_3 + \dots + n_k = n$  является объемом выборки.  $n_i$  является частотой варианты  $x_i$ , а  $\frac{n_i}{n}$  число — относительной частотой.

**вы знаете:**

Таблица, данная ниже, является статистическим рядом частоты:

Таблица 6

Варианта	$x_1$	$x_2$	$x_3$	...	$x_k$
Частота варианты	$n_1$	$n_2$	$n_3$	...	$\frac{n_k}{n}$

**вы знаете:**

*Полигон частот* (многоугольник) представляет собой ломаную, соединяющую точки, соответствующие срединным значениям интервалов группировки и частотам этих интервалов.

Полигон (*polygon*) в переводе с греческого означает многоугольник.

Составим таблицу, в которой представлены варианты с относительной частотой:

Варианта $x_i$	$x_1$	$x_2$	$x_3$	...	$x_k$
Относительные частоты $\frac{n_i}{n}$	$\frac{n_1}{n}$	$\frac{n_2}{n}$	$\frac{n_3}{n}$	...	$\frac{n_k}{n}$

Данная таблица называется *вариационным рядом относительной частоты*.

Ломаная, соединяющая точки, соответствующие средним значениям интервалов группировки и относительным частотам этих интервалов, называется *полигоном относительной частоты*.

Иначе говоря, для построения полигона относительной частоты на координатной плоскости отмечают точки:  $(x_1; \frac{n_1}{n})$ ,  $(x_2; \frac{n_2}{n})$ , ...,  $(x_k; \frac{n_k}{n})$  и соединяют их отрезками.

**ПРИМЕР**

1. Дан числовой ряд: 2, 3, 4, 5, 2, 2, 5, 3, 3, 3, 4, 5, 5, 5, 5, 4, 2, 2, 2, 2. Найдите объем выборки, варианты выборки, составьте вариационный ряд частоты и вариационный ряд относительной частоты, постройте полигон частоты.

*Решение.* По условию задачи объем равен 20. В данном ряду встречаются числа 2, 3, 4, 5. Они являются вариантами выборки.

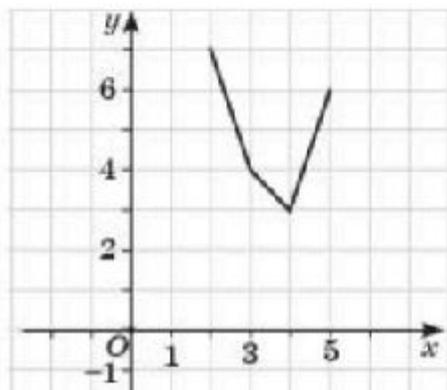
Составим вариационный ряд частоты:

Варианта $x_i$	2	3	4	5
Частота варианты $n_i$	7	4	3	6

Составим вариационный ряд относительной частоты:

Варианта $x_i$	2	3	4	5
$\frac{n_i}{n}$	0,35	0,2	0,15	0,3

Построим полигон частот (рис. 61).



- Дискретные и интервальные вариационные ряды.

**вы знаете:**

*Рядами распределения* называются группировки особого вида, при которых по каждому признаку, группе признаков или классу признаков известны численность единиц в группе либо удельный вес этой численности в общем итоге.

Ряды распределения могут быть построены или по количественному, или по атрибутивному признаку.

Ряды распределения, построенные по количественному признаку, называются *вариационными рядами*.

Вариационные ряды бывают *дискретные* и *интервальные*. Ряд распределения может быть построен по непрерывно варьирующему признаку (когда признак может принимать любые значения в рамках какого-либо интервала) и по дискретно варьирующему признаку (принимает строго определенные целочисленные значения).

# Дискретный вариационный ряд

## ПРИМЕР

1. Имеются данные о тарифных разрядах 20 рабочих. Построить дискретный вариационный ряд распределения рабочих по тарифному разряду: 2, 3, 2, 4, 4, 5, 5, 4, 6, 3, 1, 4, 4, 5, 5, 6, 4, 3, 2, 3.

*Решение.* Сначала составим таблицу. Так как ряды распределения имеют два элемента, то таблица будет состоять из двух строк.

Первая строка — это всегда варианта, в данном примере по условию это — тарифный разряд рабочих; вторая строка — это частота — как часто встречается варианта, а именно, численность рабочих определенного разряда.

Учитывая условие задания, выберем те значения, которые встречаются хотя бы один раз. В данном случае это следующие числа: 1, 2, 3, 4, 5, 6. Затем подсчитаем сколько же раз встречается каждое значение варианты и составим следующую таблицу (табл. 16):

Таблица 16

Тарифный разряд ( $x_j$ )	1	2	3	4	5	6
Численность рабочих ( $n_j$ )	1	3	4	6	4	2

Таким образом, в результате получен дискретный вариационный ряд распределения рабочих по тарифному разряду.

*Дискретным вариационным рядом распределения* называется ранжированная совокупность вариантов с соответствующими им частотами или частностями. Варианты дискретного ряда — это дискретно прерывно изменяющиеся значения признака, обычно это результат подсчета.

Дискретные вариационные ряды строят обычно в том случае, если значения изучаемого признака могут отличаться друг от друга не менее чем на некоторую конечную величину. (ранжирование это сортировка)

В дискретных рядах задаются точечные значения признака.

Разность между  $x_n - x_1$  называется *размахом измерения*, или *разностью между наибольшей вариантой и наименьшей вариантой*.

*Мода ряда данных* — это варианта, которая встречается чаще всего в ряду измерений. Мода равна варианту, кратность которой является наибольшей.

*Медианой ряда из нечетного числа данных*  $x_1 < x_2 < \dots < x_{2k} < x_{2k+1}$  называют число  $m = x_{k+1}$ , а *медианой ряда из четного числа данных*  $x_1 < x_2 < \dots < x_{2k-1} < x_{2k}$  называют число  $m = \frac{x_k + x_{k+1}}{2}$ .

# Интервальный вариационный ряд

## ПРИМЕР

2. Имеются данные о величине вклада в банке 10 вкладчиков — 280, 240, 400, 340, 200, 310, 260, 360, 330, 230 тыс. тг. Построить интервальный вариационный ряд распределения вкладчиков, по размеру вклада, выделив 3 группы с равными интервалами. По каждой группе подсчитать общий размер вкладов.

*Решение.* Сначала составим таблицу. Так как ряды распределения имеют два элемента, то таблица будет состоять из двух строк.

Первая строка — это всегда варианта, в данном примере по условию — размер вклада в банк; вторая строка — это частота — как часто встречаются варианты, а именно число вкладчиков с соответствующим вкладом, в данном случае, попадающие в интервал.

Используя формулу (1) найдем величину интервала. По условию задачи наибольшее значение 400 тыс. тг., наименьшее значение 200 тыс. тг., количество групп — 4.

$$\text{Тогда } i = \frac{400\,000 - 200\,000}{4} = 50\,000.$$

Таблица 17

Размер вклада в банк ( $x$ )	200 000 — 250 000	250 000 — 300 000	300 000 — 350 000	350 000 — 400 000
Число вкладчиков ( $n$ )	3	2	3	2

Таблица 17

Размер вклада в банк ( $x$ )	200 000 — 250 000	250 000 — 300 000	300 000 — 350 000	350 000 — 400 000
Число вкладчиков ( $n$ )	3	2	3	2

Теперь проведем подсчет общего объема размера вкладов по каждому интервалу и в целом. Для этого сложим размеры вкладов по каждому интервалу и получим суммарное значение вкладов.

По первому интервалу:  $230\ 000 + 240\ 000 + 200\ 000 = 670\ 000$ ;

по второму интервалу:  $280\ 000 + 240\ 000 = 520\ 000$ ;

по третьему интервалу:  $310\ 000 + 330\ 000 + 340\ 000 = 980\ 000$ ;

по четвертому интервалу:  $360\ 000 + 400\ 000 = 760\ 000$ .

Таблица 18

Размер вклада в банк ( $x$ )	200 000 — 250 000	250 000 — 300 000	300 000 — 350 000	350 000 — 400 000	Всего
Число вкладчиков ( $n$ )	3	2	3	2	10
Общий объем вкладов	670 000	520 000	980 000	760 000	2 930 000

•

*Интервальным вариационным рядом* называется упорядоченная совокупность интервалов варьирования значений случайной величины с соответствующими частотами или частостями попаданий в каждый из них значений величины.

Интервальные ряды предназначены для анализа распределения непрерывно изменяющегося признака, значение которого чаще всего регистрируется путем измерения или взвешивания. Варианты такого ряда — это группировка.

Если в дискретных вариационных рядах частотная характеристика относится непосредственно к варианту ряда, то в интервальных — к группе вариантов.

Есть несколько способов построения интервала:

1) визуальный способ без дополнительных расчетов на основе логического анализа данных. Расчет производится по формуле, если по условию требуется построить равные интервалы;

2) способ с дополнительными вычислениями. Для расчетов величины интервала используется формула:

$$i = \frac{x_{\max} - x_{\min}}{n}, \quad (1)$$

где  $i$  – величина или длина интервала;

$x_{\max}$  – максимальное значение;

$x_{\min}$  – минимальное значение;

$n$  – требуемое число групп по условию задачи.

Первый интервал начинают строить от минимального значения, к нему добавляется величина интервала и получается верхняя граница первого интервала. Затем верхняя граница первого интервала становится нижней границей второго интервала, к ней добавляется величина интервала и получается второй интервал. И так далее столько раз, сколько требуется построить интервалов по условию.

## Упражнения

- 21.1. Имеются данные о разрядах рабочих: 3, 4, 5, 6, 3, 3, 4, 6, 6, 6, 5, 4, 4, 4, 4, 3, 5, 5, 5, 6. Постройте дискретный вариационный ряд распределения рабочих по разрядам.
- 21.2. У 40 учащихся школы независимо друг от друга попросили назвать любую цифру. Получили следующие данные (табл. 19).

Таблица 19

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

Постройте таблицу распределения кратностей данного измерения и найдите объем выборки и моду.