

Первичная обработка данных статистического наблюдения

Результатом первичной группировки собранных статистических данных является построение **ряда распределения**, под которым понимают группировку наблюдений за значениями одного и того же показателя в один и тот же момент времени по разным единицам совокупности.

Ряд распределения всегда состоит из двух частей: значений вариант и соответствующих им частот (или частостей).

Вариантой называется значение, которое может принимать признак в ряду распределения, **частотой** – количество единиц наблюдения, обладающих значением данной варианты. Сумма частот всегда равна объему совокупности. Иногда вместо частот рассчитывают **частости** – это частоты, выраженные либо в долях единицы (тогда сумма всех частостей равна 1), либо в процентах к объему совокупности (сумма частостей будет равна 100 %).

Ряды распределения делятся на *атрибутивные* и *вариационные* в зависимости от признака, положенного в основу построения ряда.

Если признак качественный, то ряд распределения называется **атрибутивным**.

Примером атрибутивного ряда является распределение предприятий и организаций по формам собственности.

Распределение предприятий и организаций по формам собственности на 1 января 2003 года

<i>Формы собственности</i>	<i>Число предприятий и организаций, тыс.</i>
Государственная	157
Муниципальная	239
Собственность общественных и религиозных объединений (организаций)	244
Частная	2957
Прочие формы собственности	248
Всего	3845

Если признак, по которому строится ряд распределения, количественный, то ряд называется **вариационным**.

Вариационные ряды бывают *дискретные* и *интервальные*.

У **дискретных** рядов значения признака выражены конкретными числами, чаще всего целыми (варианты представляют собой перечень чисел), у **интервальных** рядов значения показателя задаются в виде интервалов.

Распределение работников предприятия по времени работы в страховой компании

Время работы в компании, полных лет (<i>варианты</i>)	Число работающих	
	Человек (<i>частоты</i>)	В % к итогу (<i>частоты</i>)
до 1 года (0)	15	11,6
1	17	13,2
2	19	14,7
3	26	20,2
4	10	7,8
5	18	13,9
6	24	18,6

Распределение предприятий-заемщиков по величине коэффициентов ликвидности

Коэффициент ликвидности	Количество предприятий, единиц	Доля предприятий, %
до 1,4	20	9,8
1,4-1,6	38	18,6
1,6-1,8	44	21,6
1,8-2,0	67	32,8
2,0 и выше	35	17,2
Итого	204	100,0

Интервалы интервального вариационного ряда имеют две границы: нижнюю и верхнюю.

Интервалы могут быть открытыми и закрытыми. У открытых нет одной из границ.

При построении интервального ряда в зависимости от характера разброса значений признака используют как равные интервальные промежутки, так и неравные.

Если признак принимает ограниченное число значений, обычно не больше 10, строят дискретные ряды распределения.

Если вариант больше, то дискретный ряд теряет свою наглядность и уже целесообразно использовать интервальную форму вариационного ряда.

Кроме того, при непрерывной вариации признака, когда его значения в определенных пределах отличаются друг от друга на сколь угодно малую величину, также строят интервальный ряд распределения.

Пример.

Предположим, что в результате статистического наблюдения получены следующие данные о средней величине процентной ставки 50 коммерческих банков (%):

14,	19,	24,	20,	12,	24,	17,	14,	19,	18,
7	0	5	8	3	6	0	2	7	8
18,	20,	21,	20,	20,	14,	25,	22,	19,	19,
1	5	0	7	4	7	1	7	0	6
19,	18,	17,	20,	13,	25,	13,	19,	18,	21,
0	9	4	0	8	6	0	0	7	1
13,	20,	15,	19,	21,	16,	16,	15,	21,	20,
3	7	2	9	9	0	9	3	4	4
12,	20,	14,	18,	15,	23,	18,	14,	21,	19,
8	0	3	0	1	8	5	4	0	0

Ход построения интервального ряда распределения с равными интервалами

1. Определим число интервалов.

Число интервалов на практике часто задается самим исследователем исходя из задач каждого конкретного наблюдения, но может вычисляться и математически, если распределение значений признака близко к нормальному закону распределения, по **формуле Стерджесса**:

$$n = 1 + 3,322 \lg N$$

где n – число интервалов,
 N – объем совокупности (число единиц наблюдения).

Для нашего примера:

$$n = 1 + 3,322 \lg N = 1 + 3,322 \lg 50 = 6,6 \approx 7$$

2. Определим величину интервалов (i) по формуле:

$$i = \frac{x_{\max} - x_{\min}}{n}$$

Для нашего примера: $i = \frac{25,6 - 12,3}{7} = 1,9$

Интервалы вариационного ряда наглядны, если их границы имеют «круглые» значения, поэтому округлим величину интервала «1,9» до «2», а минимальное значение признака «12,3» до «12,0».

3. Определим границы интервалов.

Интервалы, как правило, записывают таким образом, чтобы верхняя граница одного интервала являлась одновременно нижней границей следующего интервала.

Для нашего примера получим:

12,0 – 14,0; 14,0 – 16,0; 16,0 – 18,0; 18,0 – 20,0; 20,0 – 22,0;
22,0 – 24,0; 24,0 – 26,0.

Подобная запись означает, что признак непрерывный.

Если же варианты признака принимают строго определенные значения, например, только целые, но их количество слишком велико для построения дискретного ряда, то можно создать интервальный ряд, где нижняя граница интервала не будет совпадать с верхней границей следующего интервала.

Кроме того, в нашем примере мы могли бы сделать первый и последний интервалы открытыми, т.е. записать: «до 14,0», «24,0 и выше».

4. Подсчитаем частоты.

При подсчете частот, если значение признака попадает на границу интервала, то данная единица приписывается к тому интервалу, для которого ее значение является верхней границей, например, значение «16,0» в нашем примере будет относиться ко второму интервалу.

Результаты группировки:

***Распределение коммерческих
банков
по величине кредитной ставки***

Процентная ставка банка, %	Количество банков, ед.
12,0 – 14,0	5
14,0 – 16,0	9
16,0 – 18,0	4
18,0 – 20,0	15
20,0 – 22,0	11
22,0 – 24,0	2
24,0 – 26,0	4
—	50