

Сводка и группировка

Группировка – это разбивка изучаемых объектов на группы в зависимости от какого-либо критерия (признака).

Примеры: предприятия можно группировать по прибыли и рентабельности, формам собственности; суды – по числу рассмотренных дел; страны – по размеру ВВП, преступления – по раскрываемости и т.п.

Важная характеристика любой группировки – это группировочный признак.

Группировочный признак или **основание группировки** – это тот параметр, по которому осуществляется группировка.

Выделяют две группы:

- 1. Количественные признаки** (возраст, зарплата, цена и т.п.);
- 2. Атрибутивные или качественные признаки** (пол, вид деятельности, форма собственности и т.п.).

Ряды распределения

Ряды распределения

Атрибутивные ряды (группы строятся по качественному признаку)

Вариационные ряды (группы строятся по количественному признаку)

Дискретные ряды (группы строятся по признаку, изменяющемуся дискретно)

Интервальные ряды (группы строятся по признаку, принимающему в определенном интервале любые значения)

Ряды распределения

Ряд распределения характеризуется **двумя параметрами** (элементами):

1. **вариантами (x)** – отдельные значения (для дискретного ряда) или диапазон значений (для интервального ряда), которые принимает группировочный признак в данном ряду;

2. **частотами (f)** – числа, показывающие, сколько раз (насколько часто) встречаются те или иные варианты в ряду распределения. Их сумма определяет численность всей совокупности.

Построение ряда распределения

При построении ряда распределения важная задача – это определение числа и величины интервала.

Два варианта:

А. Если выбран **качественный** (атрибутивный) **признак**, то групп можно получить столько, сколько имеется градаций признака.

Пример: необходимо получить ряд распределения людей по полу – групп может быть только две (мужчины и женщины).

Построение ряда распределения

Б. Если выбран **количественный признак**, то количество групп определяется исследователем самостоятельно, но с учетом числа единиц совокупности и размаха варьирования значений группировочного признака. Другими словами, в данном случае также **нет жестких требований** к числу групп.

Однако существует формальный подход, отражающий связь между численностью единиц совокупности N и числом групп n . Эта связь представлена в **формуле Стерджесса** :

$$n=1+3,3221\lg N$$

Номограмма Стерджесса

Число единиц совокупности N	15-24	25-44	45-89	90-179	180-359	360-719	720-1439
Число групп n	5	6	7	8	9	10	11

Построение ряда распределения

При построении интервального ряда распределения необходимо также определить **величину интервалов**. Т.е. это делается только для интервальных (не дискретных) группировок, т.е. таких, в которых значения группировочного признака лежат в определенных границах.

Каждый интервал имеет:

А. величину (размер),

Б. нижнюю и верхнюю границы (наименьшее и наибольшее значение признака в интервале) или хотя бы одну из них.

Построение ряда распределения

Если в группировке представлены обе границы интервала (например, 20-40), то такие интервалы называются **закрытыми**. Если же указана только одна граница (например, >20 или <40), то такие интервалы называются **открытыми**.

Если требуется построить ряд распределения таким образом, чтобы величина интервалов была **равной** для всех групп, то такая величина определяется по формуле:

$$h = \frac{x_{\max} - x_{\min}}{n},$$

где x_{\max} и x_{\min} – максимальное и минимальное значения группировочного признака.

Построение ряда распределения

Пример: пусть по 30 гостиницам имеются следующие первичные статистические данные о прибыли за год (тыс. руб.): 200, 350, 600, 800, 750, 680, 960, 150, 110, 120, 200, 600, 800, 450, 560, 130, 450, 600, 800, 450, 500, 800, 960, 750, 1000, 450, 500, 600, 120, 100.

Необходимо построить интервальный ряд распределения этих гостиниц по прибыли.

Решение.

1. Варианта или признак (x) – это прибыль, частота (f) – это гостиницы.
2. Число групп или интервалов определим по формуле Стерджесса (или возьмем из номограммы):

$$n = 1 + 3,322 \cdot \lg 30 = 5,907 \approx 6.$$

Построение ряда распределения

3. Определим величину интервала:

$$h = \frac{1000 - 100}{6} = 150.$$

4. Построим интервалы прибыли гостиниц по возрастанию x . Для этого примем нижнюю границу первого интервала равной 100 тыс. руб., а его верхнюю границу найдем путем прибавления только что найденной величины интервала, т.е. $100 + 150 = 250$ тыс. руб. Таким образом, первый интервал имеет границы от 100 до 250 тыс. руб.

Далее примем нижнюю границу второго интервала равной верхней границе предыдущего интервала, т.е. 250 тыс. руб., а верхнюю границу найдем аналогично $250 + 150 = 400$ и т.д. Таким образом, второй интервал имеет границы от 250 до 400 тыс. руб. и т.д.

Обычно первый и последний интервалы оставляют открытыми: до 250 тыс. руб. и 850 тыс. руб. и более.

Построение ряда распределения

5. Определим частоту для каждого интервала f .

Например, прибыль от 100 до 250 тыс. руб. встречается у 8 гостиниц, прибыль от 250 до 400 тыс. руб. – только у 1 гостиницы и так далее.

При этом нижняя граница интервала включается в соответствующую группу, а верхняя – не включается. Например, от 100 млн. руб. (включительно) до 250 млн. руб. (не включительно), от 250 (включительно) до 400 млн. руб. (не включительно) и т.д.

Построение ряда распределения

6. Построим для интервального ряда распределения таблицу.

Прибыль гостиниц, тыс. руб.	Число гостиниц
до 250	8
250-400	1
400-550	6
550-700	6
700-850	6
850 и более	3
Итого:	30

Ряды распределения

Ряды распределения часто изображаются графически. Возможно построение:

1. Полигонов распределения
2. Гистограмм распределения
3. Кумулянт распределения
4. Огив распределения

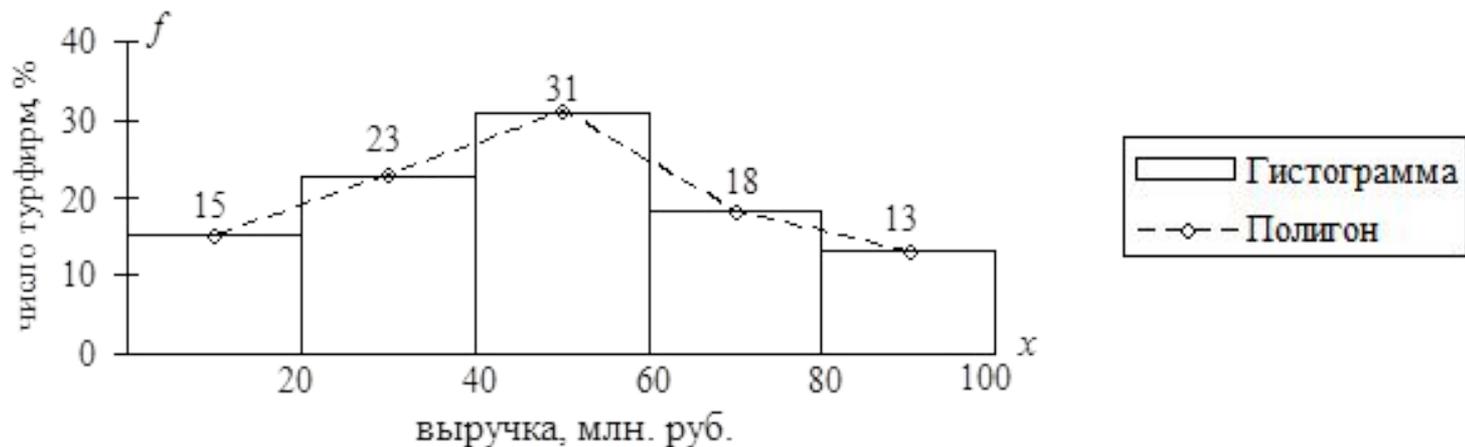
Ряды распределения

Полигон – график, на горизонтальной оси которого откладываются значения вариант x , а на вертикальной – число (доля) единиц совокупности f .

Гистограмма – это диаграмма, иллюстрирующая интервальный вариационный ряд. Она представляет собой столбцы (прямоугольники), число которых соответствует числу групп (интервалов). Ширина каждого столбца равна размеру интервала, а высота – соответствующей частоте f .

Ряды распределения

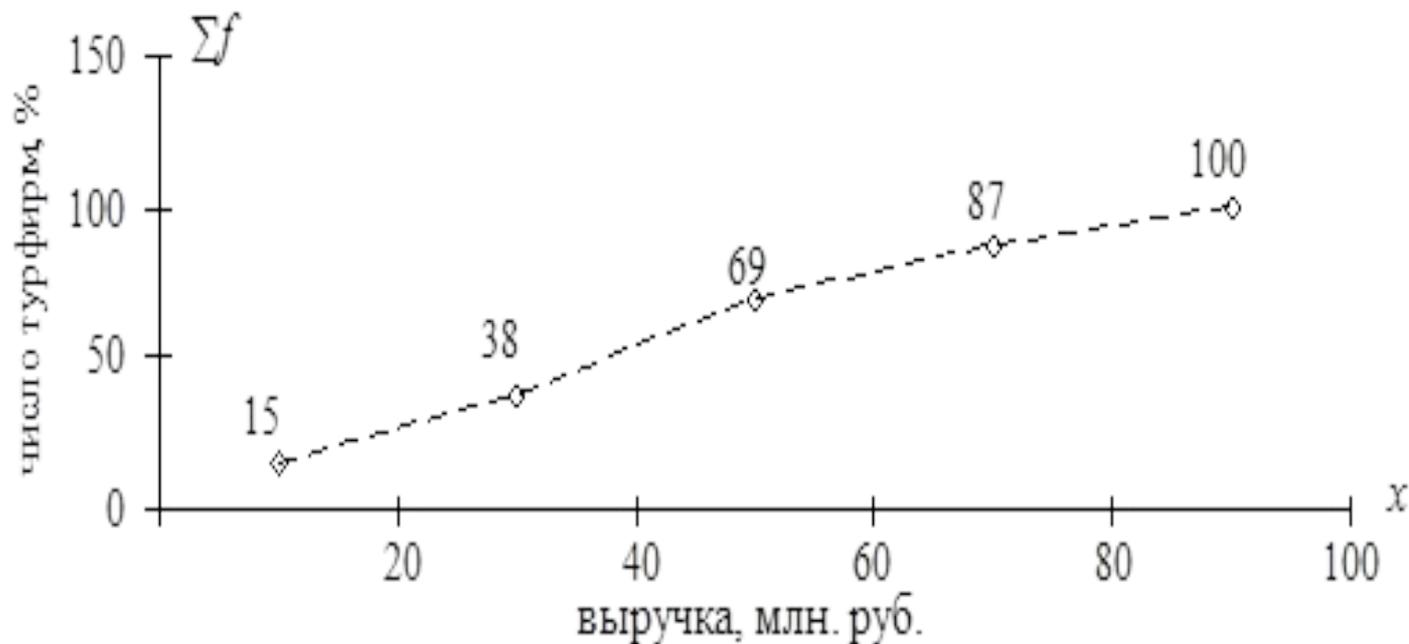
Группы турфирм по размеру выручки от реализации, млн. руб.	Число турфирм, % к итогу
0-20	15
20-40	23
40-60	31
60-80	18
80-100	13
Итого	100



Ряды распределения

Кумулята распределения – это видоизмененный полигон, на горизонтальной оси которого вновь откладываются значения вариант x , а на вертикальной – не просто частоты f , а накопленные частоты.

Любая точка кумуляты отражает, сколько единиц f имеет данное значение x и ниже. Например, 38% турфирм имеют выручку от реализации в размере до 30 млн. руб.



Ряды распределения

Огиба распределения – это полигон, на горизонтальной оси которого откладываются накопленные частоты, а на вертикальной – значения вариант x . Другими словами, если у кумуляты оси поменять местами, то получится огиба.