

Сводка и группировка статистических данных

Лекция 2

Вопросы лекции

- Сводка и группировка материалов статистического наблюдения: понятие, содержание, основные задачи.
- Виды группировок: типологическая, структурная, аналитическая.
Методологические вопросы группировки.
- Понятие о рядах распределения.
Представление статистических данных: таблицы и графики.

1 Сводка и группировка материалов статистического наблюдения: понятие, содержание, основные задачи.

Получаемая в ходе статистического наблюдения информация характеризует единицы статистической совокупности с различных сторон и не позволяет сделать обобщающие выводы об объекте в целом (т.е. о всей статистической совокупности).

Статистическая совокупность –

это множество единиц явления, объединенных в соответствии с задачей исследования единой качественной основой (однородностью), но отличающиеся друг от друга признаками.

Единицей статистической совокупности является элементы данного множества, которые характеризуются общими свойствами, т.е. признаками.

Признаки бывают:

- *атрибутивными*, т.е. качественными;
- *количественными* (дискретными и непрерывными).

Вариация признаков обуславливается случайным характером реальных явлений и процессов и зависит от изменения факторов, влияющих на объект статистического исследования.

Статистическое наблюдение – это первый этап анализа.

Статистическая сводка –

это специальным образом организованная первичная обработка данных статистического наблюдения, включающая систематизацию, группировку данных, подсчет групповых, итоговых и относительных (средних показателей). (Это второй этап обработки данных).

Программа статистической сводки устанавливает следующие этапы:

- выбор группировочных признаков;
- определение порядка формирования групп;
- разработка системы статистических показателей для характеристик групп и объекта в целом;
- разработка макетов статистических таблиц или графиков.

В сводке отдельные единицы статистической совокупности объединяются в группы при помощи метода группировок.

С помощью метода группировок решаются задачи:

- выделение социально-экономических типов явлений;
- изучение структуры явления и структурных сдвигов, происходящих в нем;
- выявление связи и зависимости между явлениями.

Средства визуализации данных

Первая группа средств показывает, сколько котиков обладает тем или иным размером. Для их использования необходимо предварительно построить так называемые *таблицы частот*. В этих таблицах есть два столбика: в первом указывается размер (или любое другое котиковое свойство), а во втором – количество котиков при данном размере.

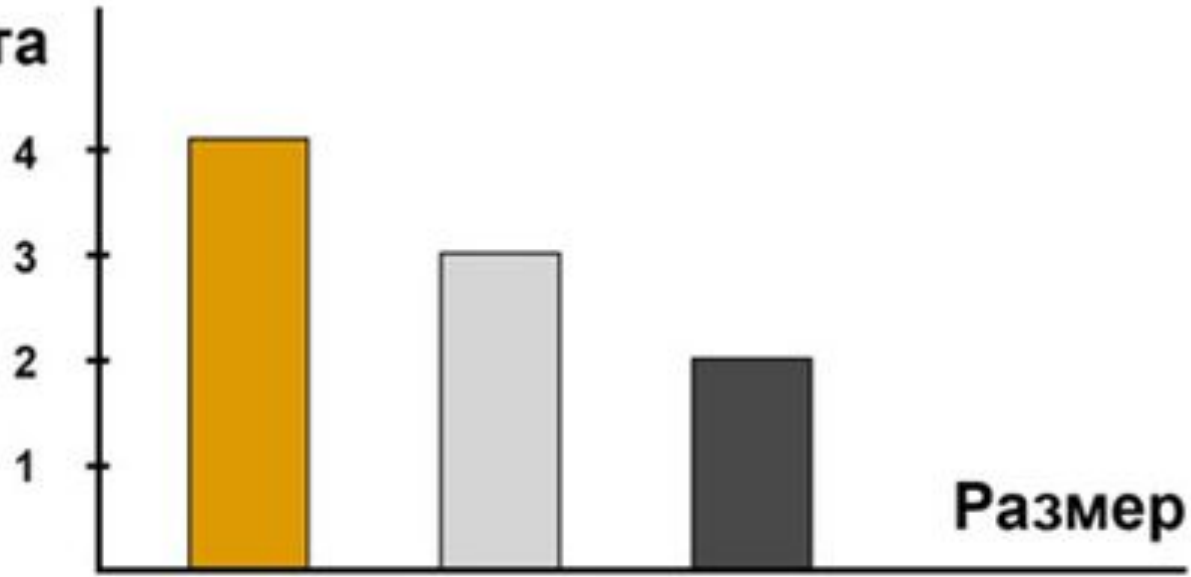
Это количество, кстати, и называется *частотой*. Эти частоты бывают *абсолютными* (в котиках) и *относительными* (в процентах).



Размер	Частота
	4
	3
	2

С таблицами частот можно делать много интересных вещей. Например построить *столбиковую диаграмму*. Для этого мы откладываем две перпендикулярных линии: горизонтальная будет обозначать размер, а вертикальная – частоту. А затем – рисуем столбики, высота которых будет соответствовать количеству котиков того или иного размера.

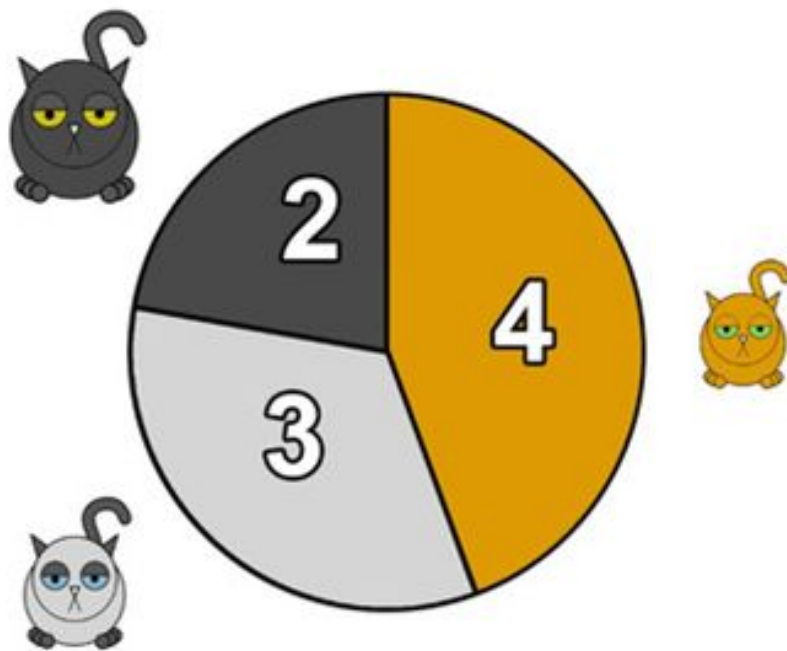
Частота



Размер



мы можем построить *круговую диаграмму*.
Величина каждого сектора такой
диаграммы будет соответствовать
проценту котиков определенного размера



2 Группировка – это

- процесс образования однородных групп на основе расчленения статистической совокупности на части или объединения изучаемых единиц в частные совокупности по существенным признакам.

Различают следующие **виды группировок**:

- *типологическая* группировка, т.е. разделение качественно разнородной совокупности на классы или однородные группы;
- *структурная* группировка, в которой происходит разделение однородной совокупности на группы, характеризующие ее структуру по какому-либо варьируемому признаку;
- *аналитическая* группировка, выявляющая взаимосвязи между изучаемыми явлениями и их признаками (факторными и результативными);
- *комбинированная* группировка, образованная по двум или более признакам.

Типологическая группировка

№ п/п	Группы предприятий по форме собственности	Число предприятий	
		единиц	В % к итогу
1	Федеральная собственность	26326	93,6
2	Муниципальная	420	1,5
3	Частная	1366	4,9
	Всего	28112	100,0

Структурная группировка

№ п/п	Группы населения по размеру среднедушевого дохода, руб.	Численность населения	
		всего, млн. чел.	в % к итогу
1	до 1000 руб.	2,4	2,0
2	1000–1800	24,8	18,0
3	1800–2600	34,2	25,0
4	2600–3400	29,4	21,5
5	3400–10000	45,7	33,5
Всего		136,5	100,0

Аналитическая группировка

№ п/п	Группы банков по сумме активов, млн.руб.	Количество банков	В среднем на 1 банк	
			Численность занятых, чел.	Балансовая прибыль, млрд. руб.
1	до 20	29	184	22,5
2	20 – 30	8	313	31,6
3	30 – 40	7	374	36,0
4	40 – 50	9	468	69,2
5	50 и более	7	516	205,6
Всего		50	1855	360,0

Принципы построения статистических группировок

1. **Выбор группировочного признака** – признака, по которому производится разбиение совокупности на отдельные группы. В качестве признака необходимо использовать существенные обоснованные признаки.

Группировочный признак – это основание (свойство объекта) для разделения объектов на группы.

Признаки различаются:

- по форме выражения (*атрибутивные* и *количественные*);
- по характеру колебания (*альтернативные* «да», «нет»; *множественные*);
- по роли во взаимосвязи явлений (*результативные* – могут меняться в зависимости от ситуации и целей анализа; *факторные* – воздействующие на другие признаки).

2. Определение количества групп

Если в основание группировки положен атрибутивный признак, то количество групп будет столько, сколько существует градаций (уровней) данного признака. Если основание группировки – количественный признак, то при определении количества групп в каждом конкретном случае следует исходить не только из степени колеблемости признака, но и из особенностей объекта и цели исследования.

$$m = 1 + 3,322 \cdot \lg N$$

Если совокупность состоит из большого числа единиц и распределение единиц по группировочному признаку близко к нормальному, для определения количества групп (m) используют формулу Стерджесса

где N – численность единиц совокупности.

3. Определение интервала группировки

- *Интервал* – это значение варьирующего признака, лежащее в определенных границах.
- Если вариация признака происходит в сравнительно узких границах и распределение носит равномерный характер, то строят группировку с **равными интервалами**:

$$h = \frac{x_{\max} - x_{\min}}{m}$$

где h – величина интервала;

x_{\max}, x_{\min} – максимальное и минимальное значения группировочного признака в совокупности;

m – число групп.

Величина интервала округляется до ближайшего целого числа, или же кратного 10, 50, 100.

Возможны и другие варианты определения интервала группировки.

- Интервалы могут быть двух видов:
- *закрытыми*, когда у интервала указаны обе границы (например, в таблице Структурная группировка, последняя – 10000 руб.);
- *открытыми*, когда у первого интервала указана верхняя граница, а у последнего – нижняя (например, в таблице Структурная группировка, 1-я группа населения по размеру среднедушевого дохода – до 1000 руб.;).

Возможно построение вторичных группировок.

Основные задачи, вторичной группировки:

- приведение данных к сопоставимым результатам;
- укрупнение интервалов;
- долевая перегруппировка (образование новых групп с меньшими интервалами).

Имеются первичные данные о количестве работников определенного возраста.

Возраст, лет	20	24	29	30	32	39	42	50	51	54	55	58	59	60
Число сотрудник ов	3	2	1	1	3	1	8	6	1	3	2	3	4	1

Произведем группировку работников предприятия по возрасту. Для этого рассчитаем число групп

- $m = 1 + 3,322 \cdot \lg 39 = 6,28 \approx 6$.

Определим интервал группировки по формуле

$$h = \frac{x_{\max} - x_{\min}}{m} = \frac{60 - 20}{6} = 6,67$$

Округлим величину интервала до ближайшего целого $h = 7$.

Тогда группировка будет следующей:

Возраст, лет	20	24	29	30	32	39	42	50	51	54	55	58	59	60
Число сотруднико в	3	2	1	1	3	1	8	6	1	3	2	3	4	1
Границы интервалов	20 – 27		27 – 33			33–40	40–47	47 – 54			54 – 61			
Число сотруднико в в интервале	5		5			1	8	10			10			

Этот понятие о рядах

распределения. Представление статистических данных: таблицы и графики

Вариационные ряды

При изучении совокупности интересующий нас признак у различных единиц совокупности принимает различные значения, т.е. он имеет некоторую вариацию.

Вариацией признака называется наличие различий в численных значениях признаков у отдельных единиц совокупности.

Чтобы выявить характер распределения единиц совокупности по варьирующим признакам, определить закономерности в этом распределении, строят ряды распределения единиц совокупностей по какому-либо варьирующему признаку.

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

При анализе вариационных рядов решают следующие задачи:

- 1) *Определение меры вариации*, т.е. количественное измерение степени колеблемости признака. Это позволяет сравнивать различные совокупности между собой по степени рассеяния и отслеживать уровень вариации признака одной и той же совокупности в различные периоды.
- 2) *Исследование закономерностей вариации* в статистических совокупностях для изучения причин, вызывающих вариацию.

Для описания статистических распределений

обычно используются следующие виды характеристик (показателей):

- 1) средние величины;
- 2) характеристики вариации (рассеяния);
- 3) характеристики дифференциации и концентрации;
- 4) характеристики формы распределения.

Графическое отображение вариационных рядов

Вариационный ряд по своей конструкции имеет 2 характеристики:

- значения варьирующего признака – **варианты** x_i , $i = 1, 2, \dots, m$;
- число случаев вариантов: абсолютные – **частоты** n_i (f_i), относительные – **частоты** w_i (относительные доли частот в общей сумме частот).

Тогда можно сказать, что **вариационный ряд** – это ранжированный (упорядоченный) в порядке возрастания или убывания ряд статистических частот (частостей).

Вариационные ряды по способу построения бывают дискретные и интервальные

Дискретный вариационный ряд можно рассматривать как такое преобразование ранжированного ряда, при котором перечисляются отдельные значения признака и указывается их частота.

Если число вариантов велико или признак имеет непрерывную вариацию, то строится **интервальный** вариационный ряд, в котором отдельные варианты объединяются в интервалы (группы).

Существуют следующие виды
графического отображения
вариационных рядов

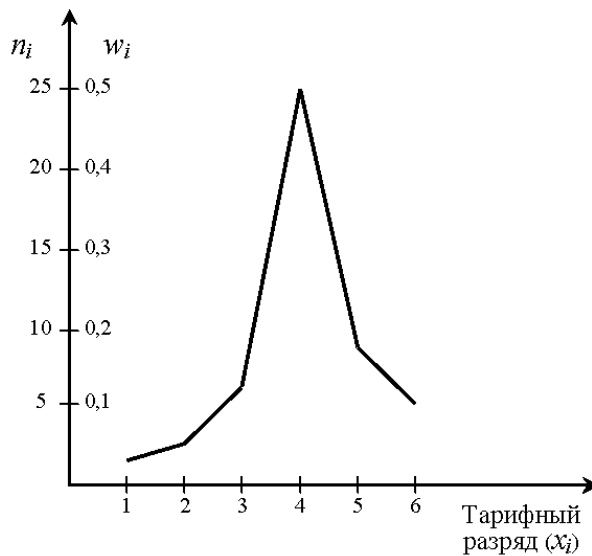
- **полигон** для отображения дискретных рядов, когда фиксируются значения (x_i ; n_i , $i = 1, 2, \dots, m$);
- **гистограмма** для отображения интервальных рядов ($k_i = x_{(i+1)} - x_i$, $n_i(w_i)$);
- **кумулята** (кумулятивный ряд) – кривая накопленных частот.

- Построить графическое отображение вариационного ряда. Дано распределение рабочих механического цеха по тарифному разряду:

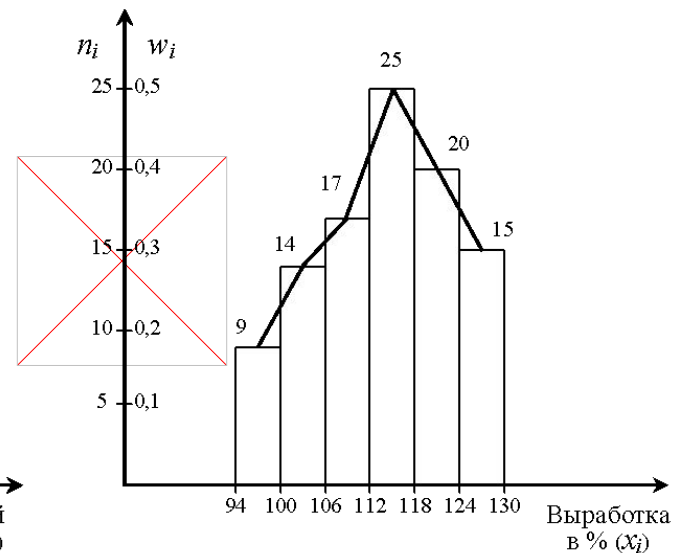
Тарифный разряд, x_i	1	2	3	4	5	6	Сумма
Количество рабочих (частота), n_i	2	3	6	25	9	5	50
Частость, $w_i = n_i/n$	0,04	0,06	0,12	0,5	0,18	0,1	1

дискретным, его графическое отображение представлено: полигон (на рис. 3.1, а), кумулята (на рис.3.2, а).

Данный вариационный ряд является дискретным, его графическое отображение представлено: полигон (на рис. 3.1, а), кумулята (на рис. 3.2, а).



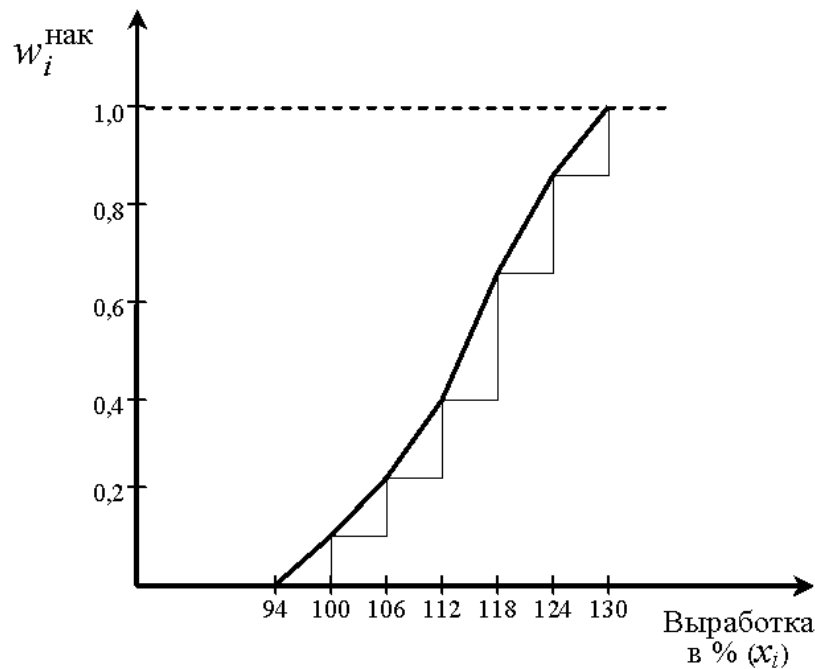
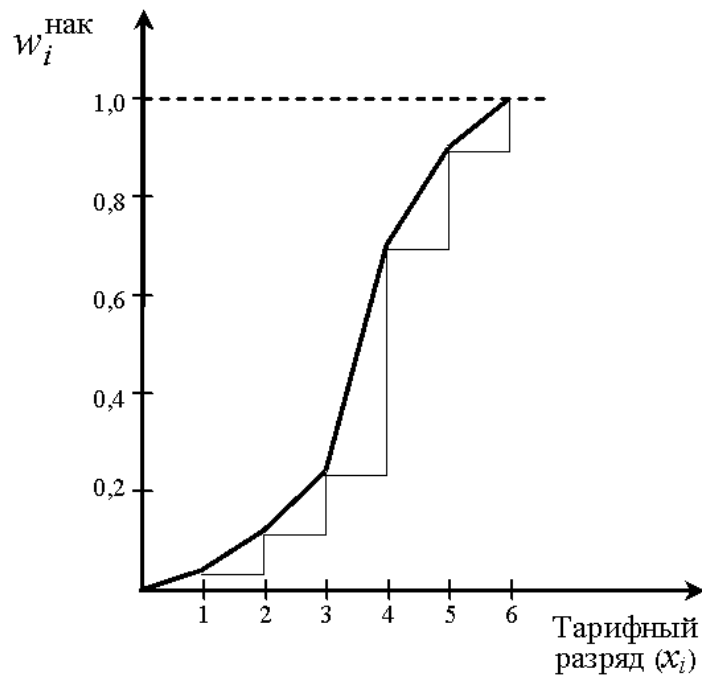
а) Дискретный вариационный ряд, (полигон)



б) Интервальный вариационный ряд, (гистограмма, полигон)

Графическое отображение

Рис. 3.1. Графическое отображение кумулятивного ряда

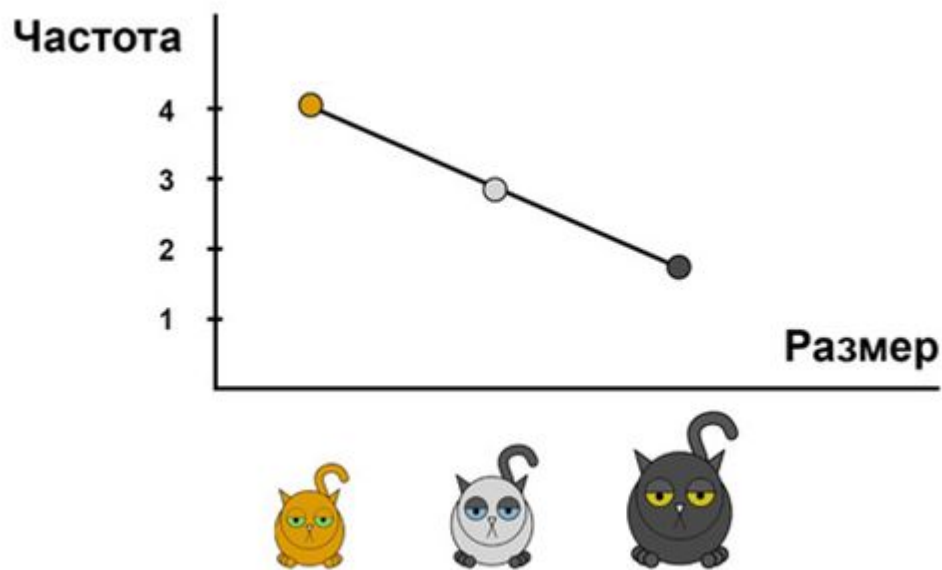


а) Дискретный вариационный ряд,
(кумулята)

б) Интервальный вариационный ряд,
(кумулята)

Рис. 3.2. Графическое отображение кумулятивного ряда

Вместо столбиков можно нарисовать точки и соединить их линиями. Результат называется *полигоном распределения*. Он довольно удобен, если котиковых размеров действительно много.



Домой вопросы:

- Понятие объекта наблюдения, единицы наблюдения, программы и субъекта наблюдения. Определение времени наблюдения и критического момента наблюдения.
- Организационные формы, виды и способы статистического наблюдения: по охвату единиц совокупности, по времени регистрации фактов, по отношению субъекта к объекту. Выборочное наблюдение – основной вид несплошного наблюдения.
- Способы формирования выборочной совокупности. Повторный и бесповторный отбор. Ошибка выборки.
- Статистический показатель, понятие и система показателей. Понятие статистического измерения. Типы шкал измерений.
- Статистическая методология: метод массовых наблюдений, метод группировок, методы обработки и анализа статистической информации.
- Этапы статистического исследования. Закон больших чисел, как основа статистической методологии.