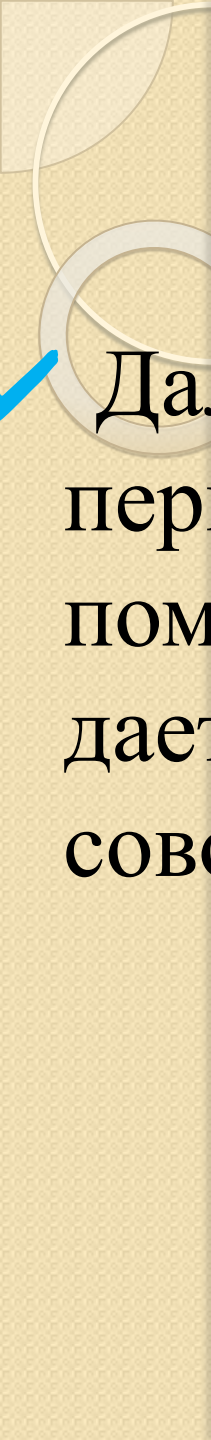




Сводка и группировка статистических данных

1. Понятие о статистической сводке. Программа и план сводки

Статистическая сводка – это второй этап статистического исследования. В результате проведения статистического наблюдения получают первичную информацию, характеризующую отдельные единицы изучаемой совокупности



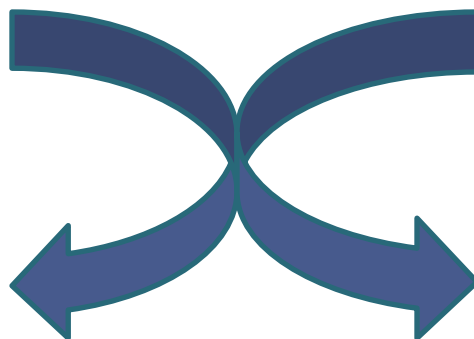
Дальнейшая задача статистики:
первичные данные систематизируются и с
помощью обобщающих показателей
дается сводная характеристика всей
совокупности

Статистическая сводка



научно-организованная обработка материалов наблюдения, включающая подсчет групповых и общих итогов, систематизацию, группировку данных, составление таблиц

Различают простую и сложную сводку



✓ При простой сводке производится подсчет только общих итогов по изучаемой совокупности

✓ При сложной сводке производится: группировка единиц наблюдения; подсчет итогов по каждой группе и по всей совокупности; представление результатов группировки в виде статистических таблиц

Содержание сводки:

Сводка состоит из следующих этапов:

1. Выбор группировочного признака
2. Определение порядка формирования групп
3. Разработка системы статистических показателей для характеристики отдельных групп и совокупности в целом
4. Разработка макетов статистических таблиц для представления результатов сводки

План статистической сводки

Содержит указания о последовательности и сроках выполнения отдельных частей сводки, ее исполнителях, о порядке изложения и представления результатов

2. Группировка как основа сводки. Задачи и виды группировок

Группировка представляет собой метод, при котором вся исследуемая совокупность разделяется на группы по какому-то существенному признаку

Основание группировки

- ✓ Признак, лежащий в основе группировки, называется группировочным или основанием группировки
- ✓ В зависимости от вида группировочных признаков различают группировки по количественным и качественным (атрибутивным) признакам

С помощью метода группировки решаются следующие задачи:

- ✓ выделение социально-экономических типов явлений
- ✓ изучение структуры явления и структурных сдвигов, происходящих в нем
- ✓ выявление связей и зависимостей между отдельными признаками явлений

Виды группировки

В зависимости от решаемых задач различают типологические, структурные и аналитические группировки

Типологическая группировка



Это разделение исследуемой совокупности на однородные группы, социально-экономические типы. Примером типологической группировки может служить группировка предприятий России по формам собственности в **2006** г.:

Всего	Число предприятий и организаций, тыс.	В процентах к итогу
	4771,9	100
в том числе по формам собственности:		
государственная	135,0	2,8
муниципальная	257,6	5,4
частная	3975,5	83,3
собственность общественных и религиозных организаций (объединений)	197,4	4,1
прочие формы собственности, включая смешанную российскую, иностранную, совместную российскую и иностранную	206,4	4,4

Таблица 1



Распределение предприятий и организаций РФ по формам собственности (на 1 января 2009г.)



Структурная группировка

Это группировка, в которой происходит разделение однородной совокупности на группы, характеризующие ее структуру по какому-то варьирующему признаку.

Примером структурной группировки может служить группировка населения России по размеру среднедушевого денежного дохода в 2007 г.:

Распределение населения РФ по величине среднедушевых денежных доходов в 2007 году
(в процентах к итогу)

Все население,	100
в том числе со среднедушевыми денежными доходами в месяц, руб.:	
До 2000,0	2,6
2000,1-4000,0	11,9
4000,1-6000,0	14,9
6000,1-8000,0	13,6
8000,1-10000,0	11,3
10000,1-15000,0	19,1
15000,1-25000,0	16,5
Свыше 25000,0	10,1

N п/п	Со средним душевым ден. доходом в мес., руб.	Численность населения	
		млн чел.	в % к итогу
1	Все население	142,2	100
2	до 2000,0	3,7	2,6
3	2000,1 - 4000,0	16,9	11,9
4	4000,1 - 6000,0	21,2	14,9
5	6000,1 - 8000,0	19,3	13,6
6	8000,1 - 10000,0	16,1	11,3
7	10000,1 - 15000,0	27,2	19,1
8	15000,1 - 25000,0	23,5	16,5
9	свыше 25000,0	14,3	10,1

Аналитическая группировка

позволяет выявить взаимосвязи между изучаемыми явлениями и их признаками. Примером аналитической группировки может служить – зависимость средней заработной платы от стажа работы:

Зависимость заработной платы от стажа

Стаж, лет	Заработная плата, руб.
Менее 5	11 000
5-10	20 500
10-15	32 900
15-20	33 600
20 и более	30 800

**Продолжительность
договорных
связей магазина с
поставщиками, лет**

Число поставщиков

абсолютное

в % к итогу

**Доля
качественной
стандартной
книжной
продукции, %**

До 2	3	14	65
3–5	8	38	69
5–8	6	29	74
Свыше 8	4	19	91
Итого	21	100	74,8

В зависимости от степени сложности массового явления и задач анализа группировки могут производиться по одному или нескольким признакам

Группировка по одному признаку называется простой. Группировка по двум и более признакам называется сложной (комбинационной)

От группировок следует отличать классификацию.

Классификация –

это систематизированное распределение явлений и объектов на определенные группы, классы, разряды с учетом их сходства и различия

Отличительные черты классификации:

- 1) в ее основу кладется качественный признак
- 2) классификации стандартны. Они устанавливаются органами государственной и международной статистики
- 3) классификации устойчивы. Они остаются неизменными в течение длительного времени

3. Определение количества выделяемых групп

Важнейшим вопросом является определение количества выделяемых групп.

Если в основании группировки лежит атрибутивный признак, то количество выделяемых групп определяется самим этим признаком. Например, производя группировку студентов, посещающих бассейн, по полу, выделяют две группы: мужчин и женщин

Если в основании группировки лежит количественный признак, то производят специальные расчеты для определения количества групп и величин интервалов группировки

Интервал группировки

это количественное значение, которое определяется как разность между максимальным и минимальным значениями признака в каждой группе

Интервалы группировки могут быть:



Равные



Неравные

Группировки с равными интервалами

применяются в тех случаях, когда вариация признака проявляется в сравнительно узких границах и распределение является практически равномерным

Для группировок с равными интервалами величина интервала h определяется как отношение разности между максимальным и минимальным значениями признака к количеству выделяемых групп:



где:

R – разность между максимальным и минимальным значениями признака (размах вариации)

X_{max} – максимальное значение признака в совокупности

X_{min} – минимальное значение признака в совокупности

n – количество выделяемых групп

✓ Оптимальное количество групп определяется по формуле Стерджеса:

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

где:

n – количество образуемых групп;

N – число единиц совокупности

Пример I

Если изучаемая совокупность состоит из 50 единиц, то число образуемых групп равно 7

Расчеты по формуле Стерджеса

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

Получаем следующее
соотношение:

Таблица 4

Определение количества групп по формуле Стерджеса

N	15-24	25-44	45-89	90-179	180-359	360-719
n	5	6	7	8	9	10

Источник: результаты расчетов по формуле Стерджеса

□ Интервалы групп могут быть открытые (указана одна из границ) и закрытые (указаны и верхняя и нижняя граница интервала). Величина открытого интервала приравнивается к величине смежного с ним интервала

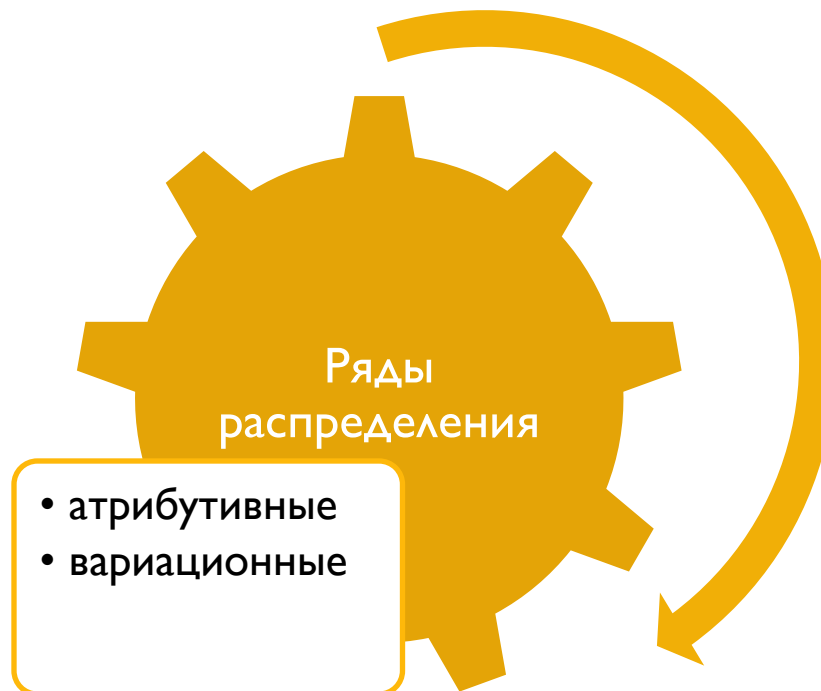
□ После определения группировочного признака, количества групп и интервалов группировки данные сводки и группировки представляются в виде рядов распределения и оформляются в виде таблиц

РЯДЫ РАСПРЕДЕЛЕНИЯ

Статистический ряд распределения представляет собой упорядоченное распределение единиц изучаемой совокупности на группы по определенному варьирующему признаку

Виды рядов распределения

В зависимости от признака, положенного в основу образования ряда распределения, различают:



Атрибутивными

называют ряды распределения, построенные по качественным признакам. Примерами атрибутивного распределения может служить распределение населения по полу, национальности, месту проживания

Вариационным

называются ряды **и** распределения,
построенные по количественному
признаку (в порядке возрастания или
убывания признака)

ПРИМЕ

Р:

Распределение студентов по возрасту, росту



варианты

The diagram illustrates the components of a variational distribution series. It features a central blue circle labeled 'Вариационный ряд распределения'. This central circle is connected by a yellow arc to an orange oval at the top labeled 'варианты' and by a red arc to a red oval at the bottom labeled 'частоты'.

Вариационный
ряд
распределения

частоты



Вариационный ряд распределения состоит из двух элементов: вариант и частот. Количественные значения признака в вариационном ряду распределения называются вариантами и обозначаются **x** . Частоты – это числа, показывающие: сколько раз в совокупности встречается данное значение признака, и обозначаются **f**

*Сумма всех частот равна
численности всей
совокупности*

Частоты –

это частоты, выраженные в процентах к
итогу или в долях. Сумма всех частотей,
выраженных в процентах, равна **100 %**, в
долях – **1**

✓ В зависимости от характера вариации признака вариационные ряды распределения подразделяются на дискретные и интервальные

✓ Если варианты признаков представлены в виде целых чисел (например, число детей в семье), то такой вариационный ряд называется *дискретным*

✓ Когда значения признака выражены в виде интервалов, это интервальный ряд

✓ Вариационные ряды распределения представляют в виде таблицы, состоящей из двух колонок. В первой колонке приводятся отдельные значения варьирующего признака, т.е. варианты. Во второй – числа, показывающие, сколько раз в совокупности встречается данная варианта, т.е. частоты

Пример

- имеются результаты обследования **20** семей, в которых выяснялось количество детей:
- **0 1 2 3 1 2 3 4 1 0 1 2 1 1 0 3 1 2 1 4**
- Поскольку значения признака представлены в виде целых чисел, построим дискретный ряд распределения

Таблица 5

Распределение 20 семей по количеству детей

Число детей	Число семей
0	3
1	8
2	4
3	3
4	2
Итого	20

Источник: рассчитано по примеру 2.

Микроперепись населения России в **1994** г.

Домохозяйства, состоящие из (чел):	Число домохозяйств в %
1	19,2
2	26,2
3	22,6
4	20,5
5	11,5
Итого	100,0

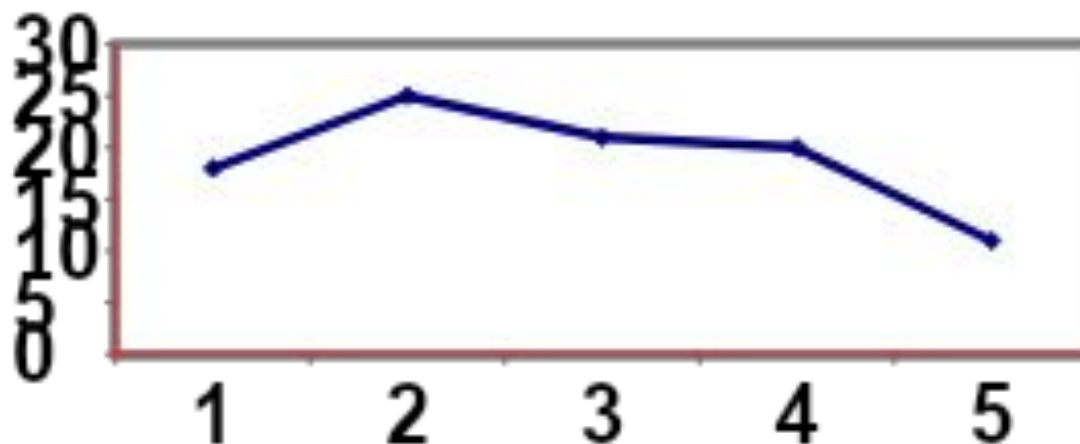
Источник: статистической информации - краткий статистический сборник "Россия в цифрах", Госкомстат, М., 1998.

❖ Для графического изображения дискретного вариационного ряда применяется полигон распределения

❖ В качестве интервального вариационного ряда рассмотрим распределение **30** рабочих бригад по размеру зарплаты (см. пример **3**)

Полигон распределения

Процент домохозяйств.



Размер домохозяйств, чел.

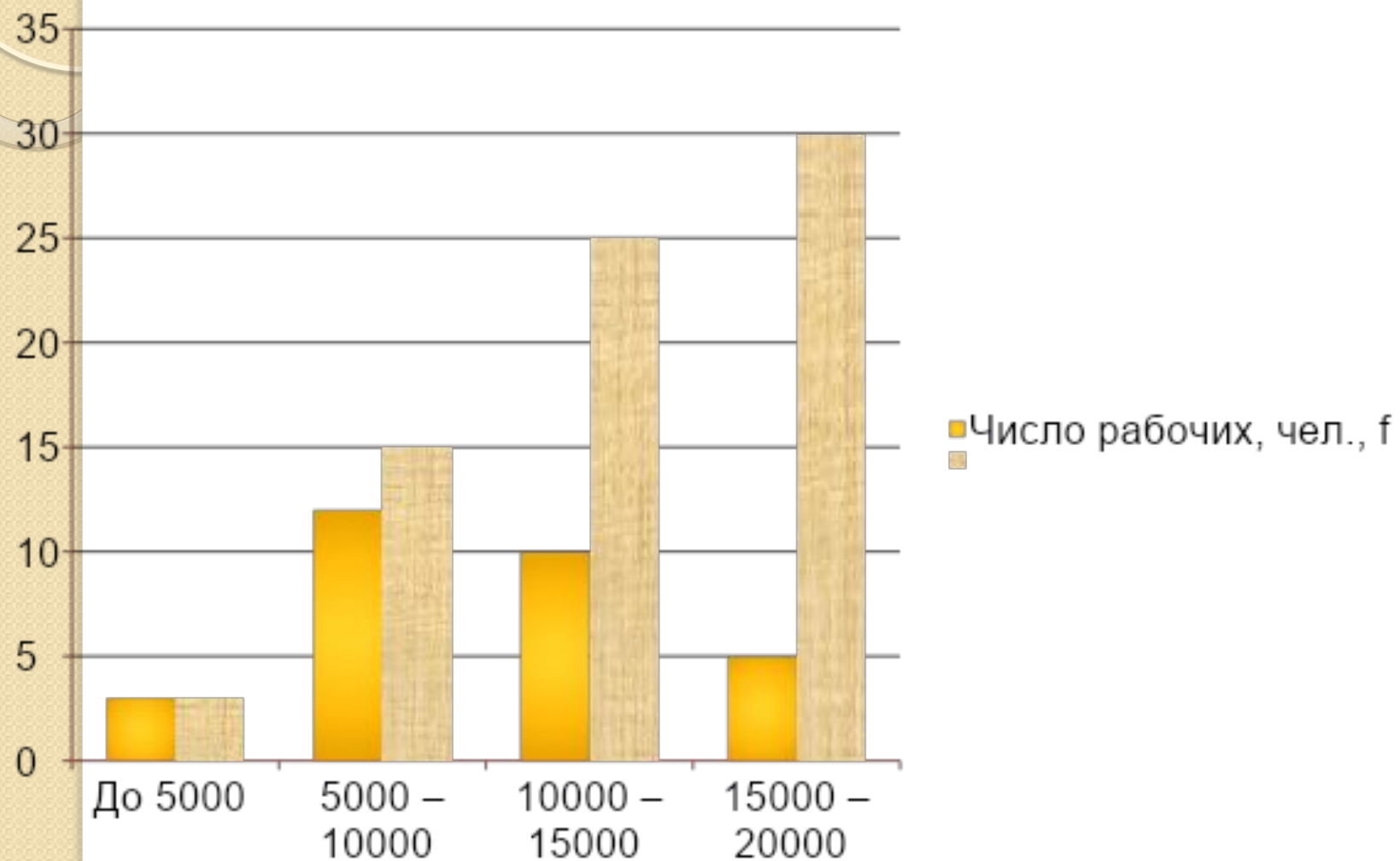
Таблица 7

Распределение **30** рабочих бригад по размеру месячной заработной платы

Заработная плата, руб. в месяц, x	Число рабочих, чел., f	Накопленные частоты, S
До 5000	3	3
5000 – 10000	12	15
10000 – 15000	10	25
15000 – 20000	5	30
Итого	30	

Источник: Числа условные

Гистограмма

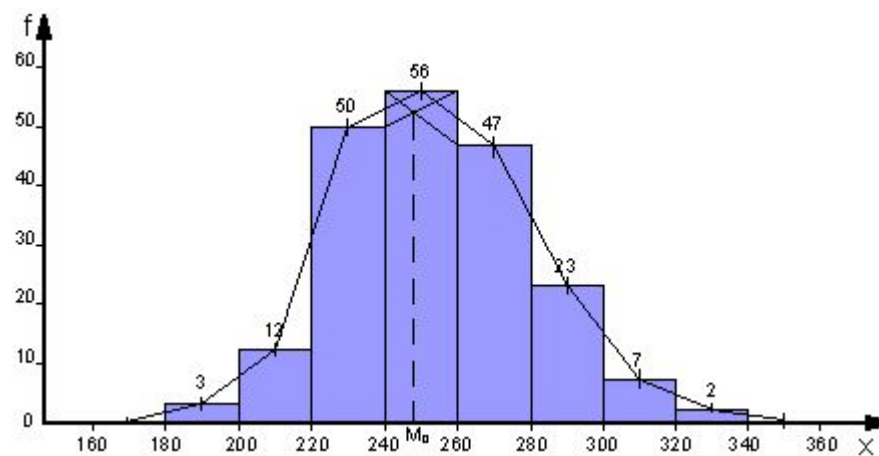




□ Для графического изображения интервального вариационного ряда применяется гистограмма

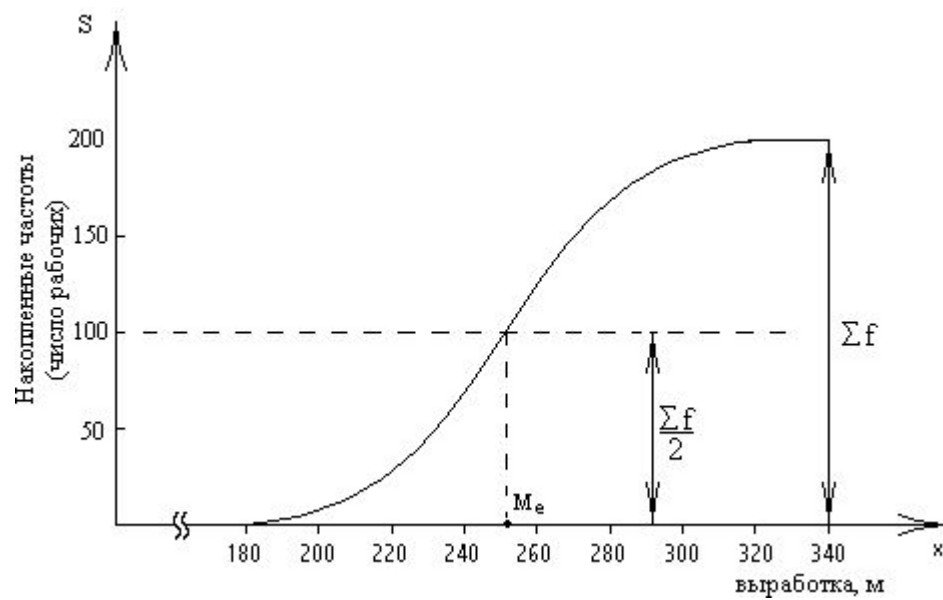
□ Любую гистограмму можно преобразовать в полигон распределения. Для этого достаточно последовательно соединить середины верхних оснований образованных прямоугольников

Гистограмма



В ряде случаев для графического изображения интервальных вариационных рядов применяется кумулята. Для ее построения сначала необходимо рассчитать накопленные частоты. Они определяются путем последовательного суммирования частот предшествующих интервалов и обозначаются **S**. Накопленные частоты показывают, сколько единиц совокупности имеют значение признака не больше, чем рассматриваемое. Рассчитаем частоты для примера **3**

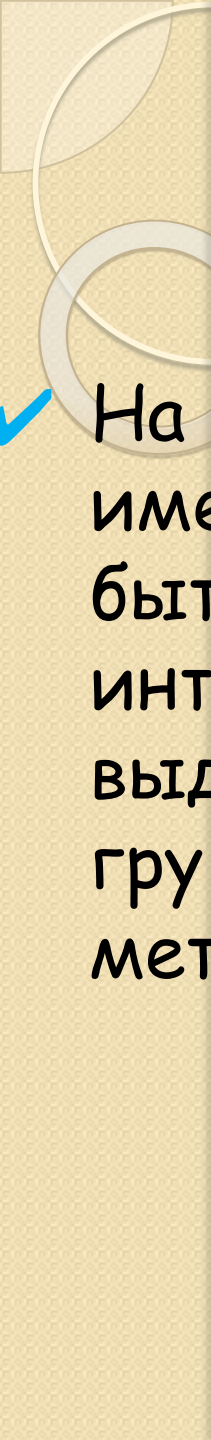
Кумулята



Рассмотрим, как подсчитываются параметры кумуляты на примере таблицы **7** о распределении **30** рабочих по размеру заработной платы.

Для первого интервала накопленная частота равна **3** (так как предшествующего интервала нет, **$3 + 0 = 3$**), для второго интервала – **15** (**$3 + 12 = 15$**), для третьего – **25** (**$3 + 12 + 10 = 25$** или **$15 + 10 = 25$**), для последнего интервала накопленная частота равна **30** (**$3 + 12 + 10 + 5 = 30$** или **$25 + 5 = 30$**)

Накопленная частота последнего интервала должна быть равна сумме частот, т.е. численности единиц совокупности. При построении кумуляты нижней границе первого интервала присваивается накопленная частота, равная **0**, и вся накопленная частота интервала присваивается его верхней границе. Для построения кумуляты на оси абцисс откладывают отрезки, соответствующие интервалам значений признака, на оси ординат — накопленные частота



✓ На практике приходится пользоваться уже имеющимися группировками, которые могут быть несопоставимы из-за неодинаковых границ интервалов или различного количества выделяемых групп. Для приведения таких группировок к сопоставимому виду используется метод вторичной группировки

5. Вторичная группировка

- ✓ Вторичная группировка – это образование новых групп на основе ранее произведенной группировки.
Применяют два способа образования новых групп на основе ранее произведенной группировки

Первый способ



состоит в укрупнении первоначальных интервалов. Это наиболее простой и распространенный способ

Второй способ



называется методом долевого перегруппировки и состоит в том, что за каждой группой закрепляется определенная доля единиц совокупности. Рассмотрим два способа на примере

Имеется группировка сотрудников двух управлений одного из московских банков по размеру месячной заработной платы тыс. руб.

Кредитное управление

Валютное управление

Группы

Размер
зарплатыЧисло
работников,
чел.

Группы

Размер
зарплатыЧисло
работников,
чел.

1

20-25

2

1

20-30

2

2

25-30

4

2

30-50

6

3

30-40

6

3

50-70

18

4

40-50

8

4

70 и более

4

5

50 и более

4

-

-

Итого

24

Итого

30

✓ Приведенные данные не позволяют сравнить распределение работников по размеру месячной заработной платы, так как величины интервалов различны, поэтому необходимо привести эти ряды распределения к сопоставимому виду.

✓ Произведем вторичную группировку, образовав группы с новыми укрупненными интервалами

Группировка сотрудников двух управлений одного из московских банков по размеру месячной заработной платы, тыс.руб.

N группы	Размер зарплаты	Число работников			
		Кредитное управления		Валютное управления	
		чел.	В % к итогу	чел.	В % к итогу
1	до 30	6	25,00	2	6,67
2	30-50	14	58,33	6	20,00
3	50 и выше	4	16,67	22	73,33
	Итого	24	100,00	30	100,00



❖ При вторичной группировке методом долевой перегруппировки устанавливаем новые интервалы распределения работников по размеру месячной заработной платы, за каждым интервалом закрепляем определенную долю единиц совокупности. В нашем примере одну из группировок (по Кредитному управлению) оставляем без изменений

По валютному управлению производим перегруппировку следующим образом. В первой группе с интервалом от **20** до **30** тыс.руб. частота равна **2**. Применительно к группировке по кредитному управлению этот интервал необходимо разбить на два равных интервала : от **20** до **25** тыс. руб. и от **25** до **30** тыс. руб., при этом исходная частота делится на поровну. Следующий интервал от **30** до **50** тыс. руб. соответственно надо разделить на два равных интервала: от **30** до **40** тыс. руб. и от **40** до **50** тыс. руб., при этом исходная частота делится поровну (**6 : 2 = 3**). Последние две группы необходимо объединить в одну с интервалом **50** тыс. руб. и более

Группировка сотрудников двух управлений одного из московских банков по размеру месячной заработной платы, тыс. руб.

Группы	Размер зарплаты	Число работников			
		Кредитное управления		Валютное управления	
		чел.	в % к итогу	чел.	в % к итогу
1	20-25	2	8,33	1	3,33
2	25-30	4	16,67	1	3,33
3	30-40	6	25,00	3	10,00
4	40-50	8	33,33	3	10,00
5	50 и выше	4	16,67	22	73,34
	Итого	24	100,00	30	100,00