

# СТАТИСТИЧЕСКИЕ МЕТОДЫ НА ТРАНСПОРТЕ



# СТАТИСТИКА:

- **Характеризует** результаты производственной деятельности предприятия, наличие материальных и трудовых ресурсов и эффективность их использования.
- **Выявляет** неиспользованные резервы.
- **Дает оценку** конкурентоспособности предоставляемых услуг, состояния рынка транспортных услуг.
- **Обеспечивает** разносторонний материал для составления текущих и перспективных проектов.

# ПРЕДМЕТ СТАТИСТИКИ АВТОМОБИЛЬНОГО ТРАНСПОРТА



Термин **«статистика»** от латинского слова **«статус» (status)**, что означает **определенное положение вещей**

**Статистика** – отрасль практической деятельности людей, направленная на сбор, обработку и анализ данных, характеризующих социально-экономическое развитие страны, ее регионов, отраслей экономики, отдельных предприятий.

**Статистика** – наука, разрабатывающая теоретические положения и методы, используемые статистической практикой.

**Статистическая практика** применяет правила, выработанные наукой.

**Статистическая наука** опирается на материалы практики, обобщает ее опыт, разрабатывает новые положения.

**Особенность статистики** – статистические данные сообщаются в количественной форме, т.е. **статистика говорит языком цифр.**

# Предмет и объект статистики автомобильного транспорта

- **Объект** – автомобильный транспорт, т.е. совокупность предприятий любой формы собственности, производственный процесс которых заключается в перемещении грузов и пассажиров.
- **Предмет** – массовые экономические явления и процессы на автотранспортных предприятиях.

# **Этапы статистического исследования**

- **Массовое научно организованное наблюдение**
- **Группировка и сводка материала**
- **Обработка статистических показателей и анализ результатов**

# Группировка статистических приемов в зависимости от этапов исследования

<i>Этап статистического исследования</i>	<i>Группа статистических приемов исследования</i>
<b>Сбор данных</b>	<b>Статистическое наблюдение</b>
<b>Первичная обработка полученных данных</b>	<b>Статистическая группировка и сводка – статистические таблицы</b>
<b>Получение обобщающих статистических показателей и их анализ</b>	<b>Анализ средних величин Вариационный анализ Корреляционный и регрессионный анализ Построение динамических рядов Индексный анализ Выборочный метод и др.</b>

# **Задачи статистического исследования**

- **получение обобщающих характеристик исследуемой совокупности;**
- **выявление связи между признаками;**
- **изучение закономерностей развития явлений во времени и в пространстве;**
- **исследование изменений в структуре явлений;**
- **моделирование и прогнозирование развития социально-экономических явлений и процессов.**



# Основные термины статистики

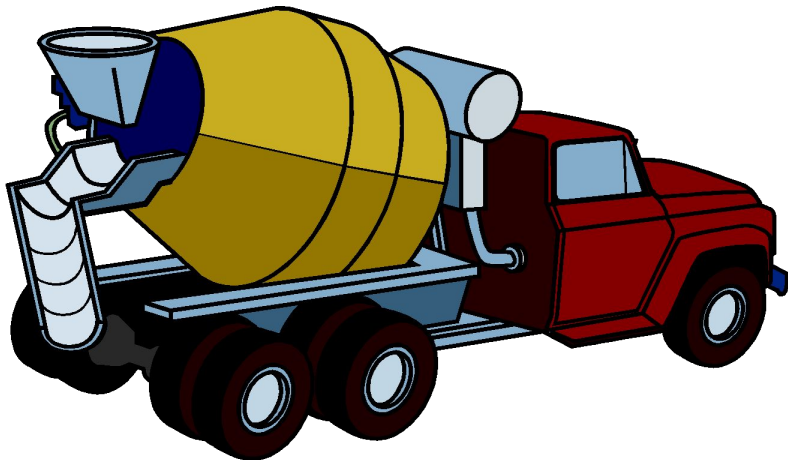
- Статистическая совокупность – это множество однокачественных варьирующих явлений.
- Единица совокупности (элемент) – неделимый первичный элемент совокупности, выражающий ее качественную однородность.
- Объем совокупности – это количество единиц в совокупности.
- Единицы совокупности обладают определенными свойствами, качествами, которые называются признаками.

# Основные термины статистики

- **Варианты** – значения, которые может принимать признак.
- **Вариация** – изменение значений признака при переходе от одной единицы совокупности к другой. Чем более однородна совокупность, тем меньше варьируют значения признаков.
- **Статистический показатель** – это понятие, отображающее количественные характеристики (размеры) соотношения признаков общественных явлений. Статистические показатели могут быть объемными (численность ПС) и расчетными (средние величины).
- **Признак** – это свойство, присущее единице совокупности.
- **Показатель** – это характеристика группы элементов или совокупности в целом.
- **Система статистических показателей** – это совокупность статистических показателей, отражающая взаимосвязи, которые объективно существуют между явлениями и процессами.

# **1 этап статистического исследования СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ НА АВТОМОБИЛЬНОМ ТРАНСПОРТЕ**

**Формирование информационной базы  
статистического исследования**



**Термин «информация» - от лат. «*informatio*» - сведения, передаваемые людьми устным, письменным или др. способом**

- **Статистическая информация** – первичный статистический материал, получаемый в результате статистического наблюдения, который подвергается систематизации, сводке, обработке и анализу.
- **Статистическое наблюдение** – планомерное, научно организованное получение первичной статистической информации об изучаемом явлении или процессе

# **Этапы статистического наблюдения**

- Подготовка наблюдения.**
- Непосредственный сбор первичных данных.**
- Контроль собранной информации.**

## **Требования к статистическому наблюдению:**

- Полученные при наблюдении данные должны быть достоверными**
- Полученные при наблюдении данные должны быть полными**
- Сведения, получаемые различными участниками наблюдения, аналогичные по содержанию, должны быть собраны единообразно**
- Первичная информация должна быть доставлена своевременно**

# **Программно-методологические вопросы плана статистического наблюдения:**

- **Определение цели статистического наблюдения**
- **Определение предмета исследования, т.е. указание явления и тех его сторон, которые подлежат изучению**
- **Определение объекта наблюдения, т.е. указание состава и границы той массы единиц, которые подлежат наблюдению**
- **Определение единицы наблюдения, т.е. составного элемента объекта наблюдения, на который составляется отдельная запись и признаки которого регистрируются при наблюдении**
- **Разработка программы наблюдения, т.е. перечня вопросов, на которые в процессе наблюдения нужно получить ответ по каждой единице наблюдения**
- **Разработка инструментария, т.е. формуляров и инструкций**

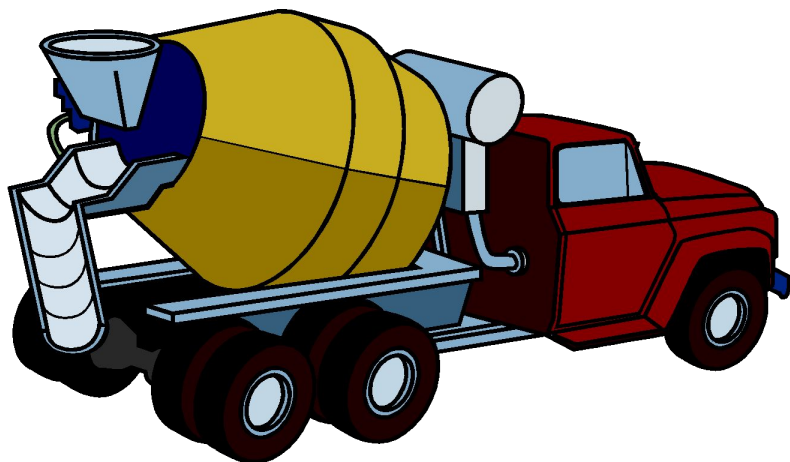
# Классификация форм, видов и способов наблюдения

Формы статистического наблюдения	Виды статистического наблюдения		Способы статистического наблюдения
	По времени регистрации фактов	По охвату единиц совокупности	
<p>1. Статистическая отчетность</p> <p>2. Специально организованное наблюдение</p> <p>3. Регистры</p>	<p>1. Непрерывное (текущее)</p> <p>2. Прерывное - периодическое - единовременное</p>	<p>1. Сплошное</p> <p>2. Не сплошное - выборочное - основного массива - монографическое - анкетное</p>	<p>1. Непосредственное</p> <p>2. Документальное</p> <p>3. Опрос - экспедиционный - саморегистрация - корреспондентский - явочный</p>



# **2 этап статистического исследования СВОДКА И ГРУППИРОВКА МАТЕРИАЛОВ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ**

**Статистические методы классификации и  
группировки**



**Статистическая сводка – это способ научной обработки первичных статистических сведений.**

**Последовательность статистической сводки:**

- **группировка статистических данных;**
- **расчет системы показателей;**
- **табличное и графическое представление результатов.**

**Статистическая группировка – это процесс разбиения единиц исследуемой совокупности на группы и подгруппы по существенным признакам.**

**Статистические группировки применяются для решения следующих задач:**

- **Разделение всей совокупности на качественно однородные группы. Такие группировки называют *типологическими*.**
- **Изучение взаимосвязей между отдельными признаками изучаемого явления. Такие группировки называют *аналитическими*.**
- **Характеристика структуры явления и структурных сдвигов. Такие группировки называют *структурными*.**

- **Этапы процесса группирования статистических данных:**

- 1. выбор группировочного признака;**
- 2. ранжирование совокупности по выбранному группировочному признаку;**
- 3. определение числа групп;**
- 4. определение величины интервала;**
- 5. распределение единиц совокупности по образованным группам.**

# 1. Выбор группировочного признака

***Группировочный признак*** – признак, на основе которого производится подразделение единиц наблюдения на группы

- **Качественные (атрибутивные) признаки** отражают состояние единицы наблюдения
- **Количественные признаки** имеют числовое выражение

# 1. Выбор группировочного признака

**Простая группировка** — выполняется по одному признаку.

**Сложная группировка** — выполняется по нескольким признакам, взятым в комбинации.

Сложная группировка может быть:

- ***комбинационной***
- ***многомерной***

# Пример комбинационной группировки: распределение АТС по типу и числу лет эксплуатации

<b>№ п/п</b>	<b>Тип АТС</b>	<b>В том числе подгруппы АТС по числу лет эксплуатации</b>	<b>Численность, ед.</b>
<b>1</b>	<b>Бортовые</b>	<b>0-2 лет</b>	<b>5</b>
		<b>3-5 лет</b>	<b>6</b>
		<b>5-10 лет</b>	<b>12</b>
		<b>10 лет</b>	<b>4</b>
<b>ИТОГО по группе</b>			<b>27</b>
<b>2</b>	<b>Самосвалы</b>	<b>0-2 лет</b>	<b>4</b>
		<b>3-5 лет</b>	<b>6</b>
		<b>5-10 лет</b>	<b>10</b>
		<b>10 лет</b>	<b>8</b>
<b>ИТОГО по группе</b>			<b>28</b>
	<b>ИТОГО по подгруппам</b>	<b>0-2 лет</b>	<b>9</b>
		<b>3-5 лет</b>	<b>12</b>
		<b>5-10 лет</b>	<b>22</b>
		<b>10 лет</b>	<b>12</b>
<b>ВСЕГО</b>			<b>55</b>

## 2. Ранжирование совокупности по выбранному группировочному признаку

- ***Ранжирование*** исследуемой статистической совокупности по группировочному признаку, т.е. все единицы наблюдения располагаются по возрастанию или убыванию значений выбранного признака.



# 3. Определение числа групп

**Число групп** определяется следующими факторами:

- задачами исследования;
- основанием группировки;
- численностью совокупности;
- степенью вариации (изменчивости) признака.

# 3. Определение числа групп

- При группировке по **качественному (атрибутивному) признаку** количество групп определяется числом градаций, видов, состояний или наименований этого признака, если это число не очень велико.
- При группировке, построенной по **количественному признаку** число групп определяется различно в зависимости от характера изменения признака и задач исследования.
- Группировка, построенная по количественному признаку может быть **дискретной и интервальной**.
- В **дискретной группировке** каждая группа представляет собой конкретное значение признака, в **интервальной** – интервал возможных значений.

# Пример дискретной группировки

Распределение водителей определенного класса,  
по количеству в АТО

<b>Группы водителей по классу</b>	<b>2011</b>	<b>2012</b>
<b>1</b>	<b>15</b>	<b>14</b>
<b>2</b>	<b>13</b>	<b>13</b>
<b>3</b>	<b>5</b>	<b>7</b>
<b>ВСЕГО</b>	<b>33</b>	<b>34</b>

# Пример интервальной группировки

**Возрастная структура парка легковых автомобилей  
в АТО (на конец года, в процентах к итогу)**

<b>Группы легковых автомобилей по числу лет эксплуатации</b>	<b>2010</b>	<b>2011</b>
<b>до 5</b>	<b>23,3</b>	<b>25,0</b>
<b>5-10</b>	<b>27,8</b>	<b>26,7</b>
<b>более 10</b>	<b>48,9</b>	<b>48,3</b>
<b>ВСЕГО</b>	<b>100</b>	<b>100</b>

# Приемы построения группировок

Для определения количества групп могут быть использованы **стандартные статистические процедуры**. Наиболее распространенная из них основана на использовании **формулы американского ученого Стерджесса**:

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

где  $m$  – число групп (округленное до целого);  
 $N$  – число единиц совокупности

# 4. Определение величины интервала

- **Интервалы группировки** – значения признака, лежащие в определенных границах.
- **Величина интервала** – это разница между верхней и нижней границами интервала, т.е. максимальным и минимальным значениями признака в каждой группе соответственно.
- Если в основание группировки положен **непрерывный признак**, то верхняя граница  $i$ -го интервала совпадает с нижней границей  $i+1$ -го.
- Если же группируется **дискретный признак**, то нижняя граница  $i+1$ -го интервала равна верхней границе  $i$ -го плюс 1.

## Пример группировки по непрерывному и дискретному признаку

Границы групп при распределении АТС по числу лет эксплуатации

№ группы	Вариант I	№ группы	Вариант II
	Число лет эксплуатации, лет		Число лет эксплуатации, лет
1	До 5	1	0 – 4
2	5 – 10	2	5 – 9
3	10 – 15	3	10 – 14
4	15 и выше	4	15 и выше
Границы групп, если число лет эксплуатации измеряется с точностью больше года		Границы групп, если под числом лет эксплуатации понимается число уже наступивших лет	

# По наличию границ различают интервалы

**Открытые** - интервалы, для которых определена только одна граница (верхняя или нижняя)

<b>Суточный пробег, км</b>	<b>До 100</b>	<b>100-150</b>	<b>150-200</b>	<b>200 и более</b>
<b>Число АТС</b>	<b>5</b>	<b>128</b>	<b>90</b>	<b>15</b>

**Закрытые** – интервалы, для которых определены обе границы (верхние и нижние)

<b>Суточный пробег, км</b>	<b>50-100</b>	<b>100-150</b>	<b>150-200</b>	<b>200-250</b>
<b>Число АТС</b>	<b>5</b>	<b>128</b>	<b>90</b>	<b>15</b>



## Группировка операторов диспетчерской по среднему времени обработки заказа

<b>№ группы</b>	<b>Среднее время обработки заказа, мин</b>	<b>Количество операторов, чел</b>
<b>1</b>	<b>До 3</b>	<b>30</b>
<b>2</b>	<b>3 – 7</b>	<b>60</b>
<b>3</b>	<b>7 и выше</b>	<b>10</b>
<b>ВСЕГО</b>		<b>100</b>

# По величине интервала группировки бывают

- **Равноинтервальные группировки** - величина интервала одинакова для всех групп
- **Неравноинтервальные группировки** - величина интервала разнится от группы к группе:
  - *прогрессивно возрастающие/убывающие;*
  - *равнонаполненные;*
  - *специализированные;*
  - *произвольные.*

# Определение величины интервала для равноинтервальной группировки

$$i = \frac{R}{m},$$

где  $R$  – размах вариации,  $R = X_{\max} - X_{\min}$ ;

$X_{\max}$  – максимальное значение группировочного признака;

$X_{\min}$  – минимальное значение группировочного признака;

$m$  – число групп.

**Полученный интервал округляется в большую сторону!**

# Схема определения границ каждого последующего интервала

<b>№ интервала</b>	<b>Границы интервала</b>
<b>1</b>	<i>от</i> $X_{\min}$ <i>до</i> $X_{\min} + i$
<b>2</b>	<i>от</i> $X_{\min} + i$ <i>до</i> $X_{\min} + 2i$
<b>...</b>	
<b><i>k</i></b>	<i>от</i> $X_{\min} + (k - 1)i$ <i>до</i> $X_{\max}$

# Пример группировки по количественному признаку с равными интервалами

- Пусть статистическая совокупность состоит из 40 автотранспортных компаний, показатели объемов перевозки груза которых варьируются от 50т до 650т, что является, соответственно, минимальным и максимальным значениями признака.

- Тогда по формуле Стерджесса получаем:

$$m = 1 + 3.322 \cdot \lg 40 \approx 6.322 \approx 6$$

- Величина интервала для построения равноинтервальной группировки определяется следующим образом:

$$i = (650 - 50) / 6 = 100$$

- Таким образом, совокупность автотранспортных предприятий будет разделена по показателю объема перевезенного груза на шесть равных групп: [50-150], [150-250], [250-350], [350-450], [450-550], [550-650] (т).

# Пример группировки по количественному признаку с равными интервалами

**Дано:** Данные о выполнении сменных норм выработки каждым рабочим за месяц, %: 100,8; 103,4; 105,2; 110,4; 108,7; 111,6; 101,9; 106,3.

**Необходимо:** построить равноинтервальную группировку

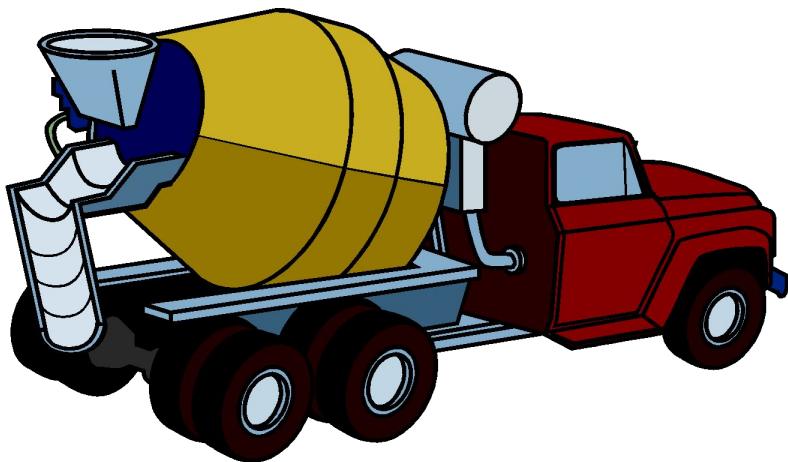
**Решение:**

- Учитывая немногочисленность данных, подразделяем рабочих на 3 группы.
- Рассчитаем величину интервала:
- $i = (111,6 - 100,8) / 3 = 3,6 = 4$

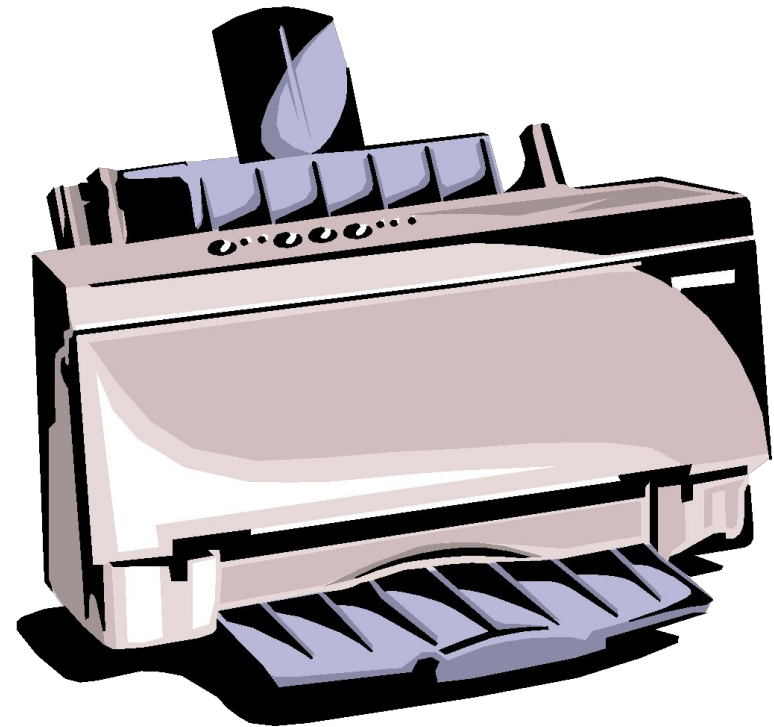
Процент выполнения сменных норм выработки за месяц	100-104	104-108	108-112
Число рабочих	3	2	3

# **3 этап статистического исследования**

## **Обработка статистических показателей и анализ результатов**



# Основные понятия ряда распределения





- **Ряд распределения** – это упорядоченное распределение единиц совокупности по группам по какому-либо варьирующему признаку.
- **Варианта ( $x_i$ )** – это конкретное значение варьирующего признака в ряду.
- **Частота ( $n_i$ )** – численность отдельных вариантов или каждой группы вариант, показывающая, как часто встречаются эти значения в ряду распределения.

$$\sum_{i=1}^k n_i = N,$$

*где  $n_i$  – число наблюдений в  $i$  – ой группе;*

*$k$  – число групп;*

*$N$  – число единиц совокупности.*

- **Частость**, – это частота, выраженная в долях единицы или в процентах к итогу.
- Сумма частостей по всем группам равна 1 или 100% соответственно, т.е.:

$$\sum_{i=1}^k p_i = 1, \quad p_i = \frac{n_i}{N},$$

*где  $p_i$  – частость в  $i$  – ой группе, выраженная в долях единицы,*

*$n_i$  – частота в  $i$  – ой группе;*

*$k$  – число групп;*

*$N$  – число единиц совокупности.*

$$\sum_{i=1}^k p_i = 100\%, \quad p_i = \frac{n_i}{N} \cdot 100\%,$$

*где  $p_i$  – частость в  $i$  – ой группе, выраженная в процентах к итогу;*

*$n_i$  – частота в  $i$  – ой группе;*

*$k$  – число групп;*

*$N$  – число единиц совокупности.*

- В зависимости от признака, лежащего в основании, различают **атрибутивные** и **вариационные** ряды распределения
- **Атрибутивный ряд распределения** – это ряд, построенный по качественному признаку.
- **Вариационный ряд распределения** – это ряд, построенный по количественному признаку
- Характер вариационного ряда может быть **дискретным** или **непрерывным**. Соответственно, различают дискретные и интервальные вариационные ряды

- **Абсолютная плотность распределения**, - это частота, рассчитанная на единицу интервала:

$$f_i = \frac{n_i}{\Delta_i},$$

где  $f_i$  – абсолютная плотность распределения в  $i$  – ой группе;

$n_i$  – частота в  $i$  – ой группе;

$\Delta_i$  – величина  $i$  – го интервала.

- **Относительная плотность распределения**, - это частота, рассчитанная на единицу интервала:

$$f'_i = \frac{p_i}{\Delta_i},$$

где  $f'_i$  – относительная плотность распределения в  $i$  – ой группе;

$p_i$  – частота в  $i$  – ой группе;

$\Delta_i$  – величина  $i$  – го интервала.

- **Накопленная частота/частость** - это число/доля единиц совокупности со значением признака не больше заданного:

$$F_k = \sum_{i=1}^k n_i \quad \forall k = 1, 2, \dots, m \quad \text{или} \quad F_1 = n_1, F_i = F_{i-1} + n_i \quad \forall i = 2, 3, \dots, m$$

где  $F_k$  – накопленная частота к концу  $k$  – ой группы;

$F_i$  – накопленная частота к концу  $i$  – ой группы;

$n_i$  – частота в  $i$  – ой группе.

$$F'_k = \sum_{i=1}^k p_i \quad \forall k = 1, 2, \dots, m \quad \text{или} \quad F'_1 = p_1, F'_i = F'_{i-1} + p_i \quad \forall i = 2, 3, \dots, m$$

где  $F'_k$  – накопленная частость к концу  $k$  – ой группы;

$F'_i$  – накопленная частость к концу  $i$  – ой группы;

$p_i$  – частость в  $i$  – ой группе.

- Кумулятивный характер накопленных частот/частостей подразумевает, что с возрастанием групповых значений их величины могут только увеличиваться. Таким образом:

$$\begin{array}{ll}
 0 < F_i \leq N & \forall i = 1, 2, \dots, k & 0 < F'_i \leq 1 & \forall i = 1, 2, \dots, k \\
 F_i > F_{i-1} & \forall i = 2, \dots, k & F'_i > F'_{i-1} & \forall i = 2, \dots, k \\
 F_k = N & & F_k = 1 & 
 \end{array}$$

- Обратная процедура – расчет частот/частостей через накопленные частоты/частости:

$$\begin{array}{ll}
 n_1 = F_1, & n_i = F_i - F_{i-1} & \forall i = 2, 3, \dots, k \\
 p_1 = F'_1, & p_i = F'_i - F'_{i-1} & \forall i = 2, 3, \dots, k
 \end{array}$$

# Графическое представление статистических данных



- **Статистический график** – это чертеж, отображающий характеристики той или иной статистической совокупности с помощью геометрических образов или знаков.
- Применительно к рядам распределения используют следующие графические изображения: **полигон, гистограмма, кумулята, огива.**

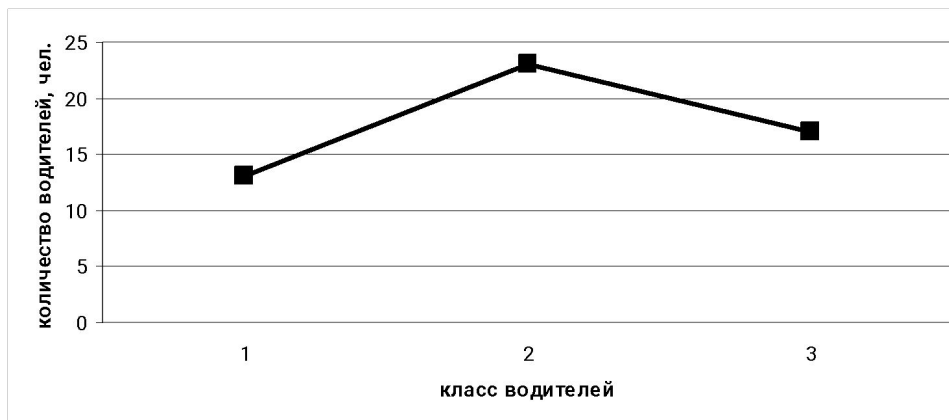


# Полигон

- **Полигон** – графическое изображение дискретного вариационного ряда распределения, дающее представление о характере изменения его частот.

Распределение водителей по классам в АТО на начало 2012 года

Группы водителей по классу	Количество водителей	
	всего, чел	в % к итогу
варианты, $x_i$	частота, $n_i$	частость, $p_i$
<b>1</b>	<b>13</b>	<b>24,5</b>
<b>2</b>	<b>23</b>	<b>43,4</b>
<b>3</b>	<b>17</b>	<b>32,1</b>
<b>ВСЕГО</b>	<b>53</b>	<b>100,0</b>



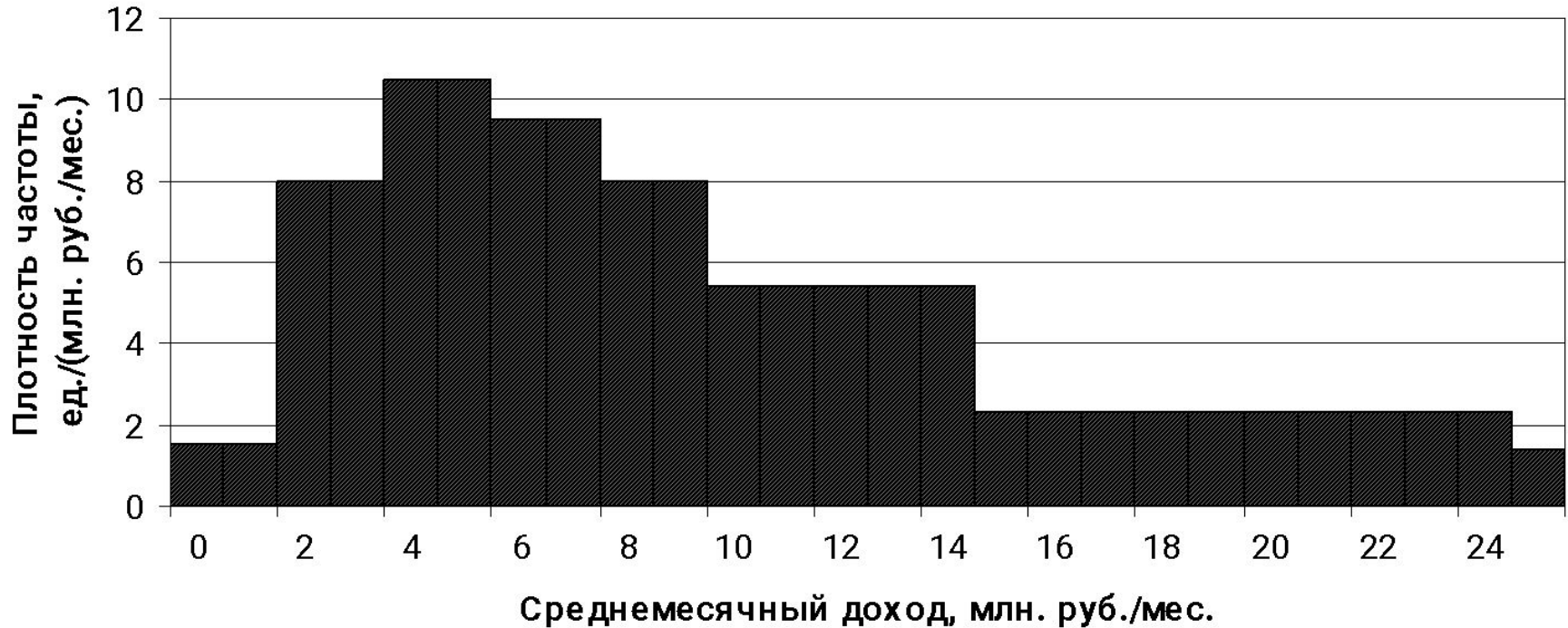
# Гистограмма

- **Гистограмма** (ленточная диаграмма) – графическое изображение интервального вариационного ряда распределения, дающее представление о характере изменения его частот.

Распределение АТО по величине среднемесячных доходов

Группы АТО по среднемесячному доходу, млн.руб./мес.	Количество АТО		Величина интервала, млн.руб./мес.	Плотность распределения	
	всего, ед.	в % к итогу		абсолютная	относительная
группы вариант, $x_i$	частота, $n_i$	частость, $p_i$	$\Delta_i$	плотность частоты, $f_i$	плотность частости, $f'_i$
до 2	3	2,2	2	1,5	1,1
2 – 4	16	11,5	2	8	5,8
4 – 6	21	15,1	2	10,5	7,6
6 – 8	19	13,7	2	9,5	6,8
8 – 10	16	11,5	2	8	5,8
10 – 15	27	19,4	5	5,4	3,9
15 – 25	23	16,5	10	2,3	1,7
свыше 25	14	10,1	10	1,4	1,0
<b>ВСЕГО</b>	<b>139</b>	<b>100</b>	-	-	-

# Гистограмма распределения АТО по величине среднемесячного дохода



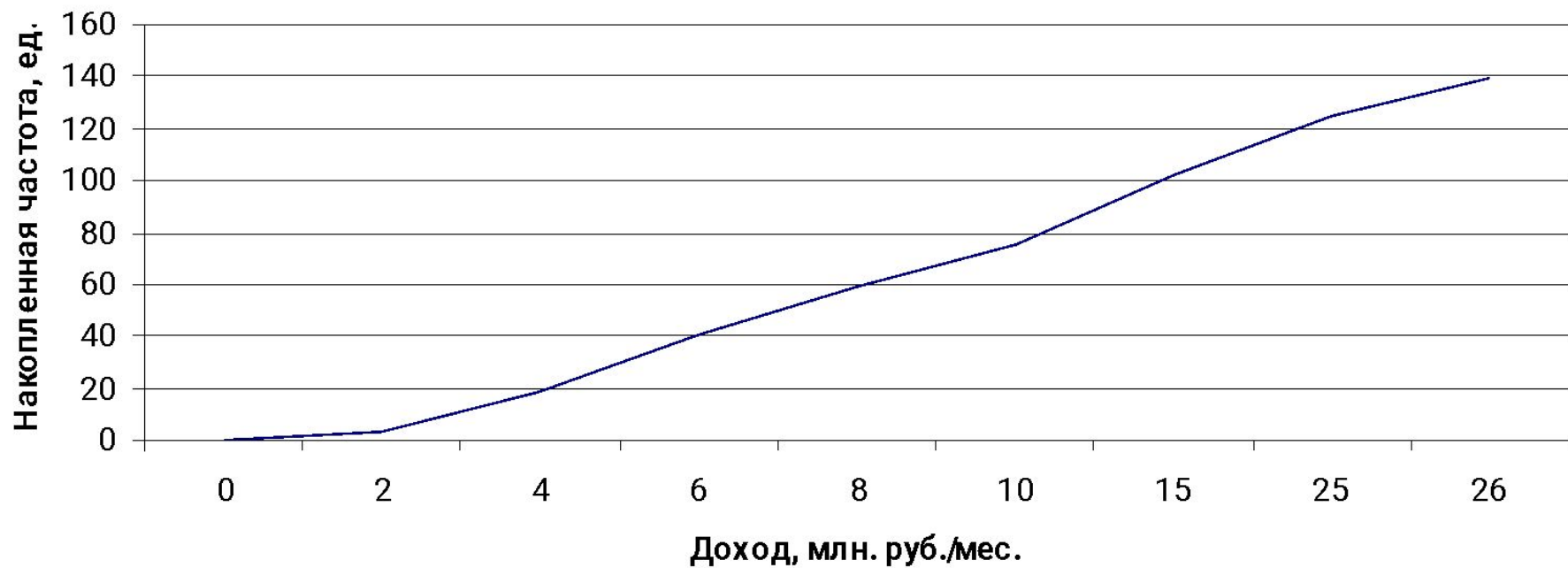
# Кумулята

- **Кумулята** – графическое изображение кумулятивной кривой, дающее представление о характере изменения накопленных частот/частостей.

Распределение АТО по величине среднемесячных доходов

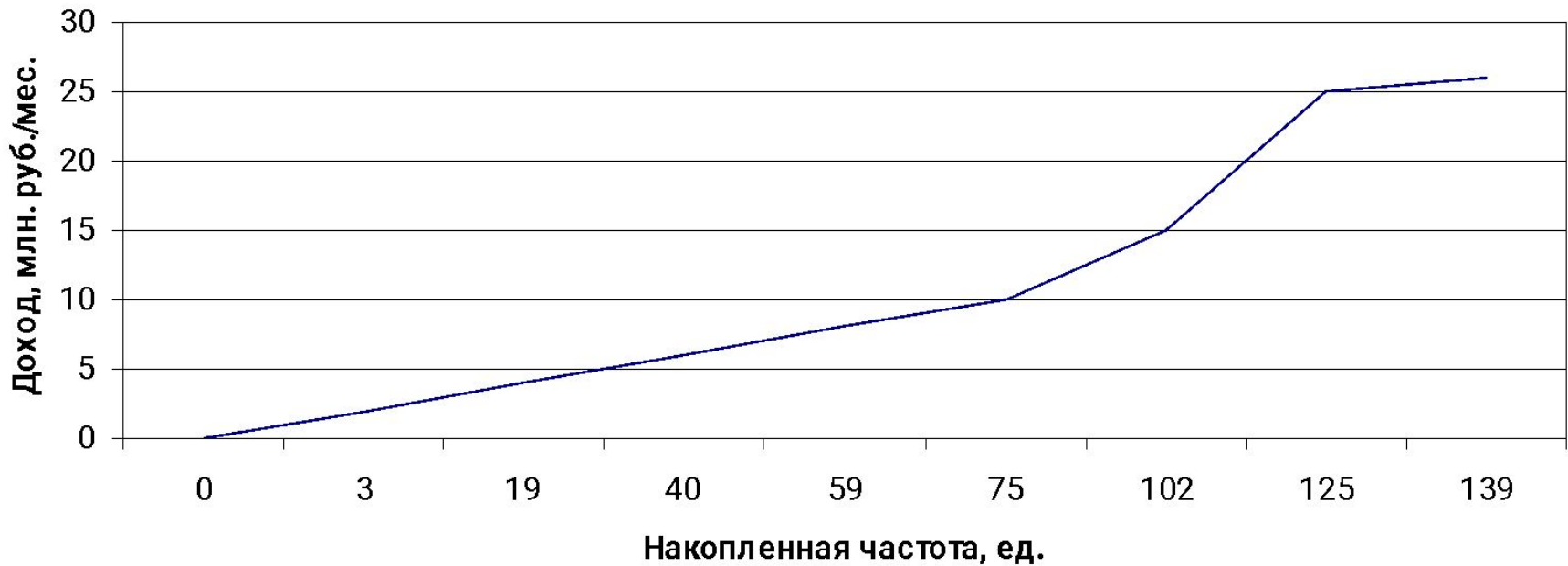
$x_i$ млн.руб./мес.	Численность			
	$n_i$ ед.	$P_i$ в % к итогу	накопленным итогом, ед.	накопленным итогом, в % к итогу
			Накопленная частота, $F_i$	Накопленная частость, $F'_i$
до 2	3	2,2	3	2,2
2 – 4	16	11,5	19	13,7
4 – 6	21	15,1	40	28,8
6 – 8	19	13,7	59	42,5
8 – 10	16	11,5	75	54
10 – 15	27	19,4	102	73,4
15 – 25	23	16,5	125	89,9
свыше 25	14	10,1	139	100
<b>ВСЕГО</b>	<b>139</b>	<b>100</b>	-	-

# Кумулята распределения АТО по величине среднемесячного дохода



# Огива

- **Огива** – это графическое изображение кумулятивной кривой, в котором оси кумуляты поменяны местами. На рис. представлена огива для распределения среднемесячного дохода.



**Огива распределения АТО по величине среднемесячного дохода**

# **Система статистических показателей.**

**Абсолютные, относительные  
и средние величины**

- **Абсолютными величинами** выражаются объемные статистические показатели. Они являются именованными величинами, имеющими **определенную размерность и единицы измерения**.
- 
- В зависимости от целей анализа применяются **натуральные, условно-натуральные, стоимостные и трудовые единицы измерения**.
- **Натуральные единицы** выражаются в физических мерах веса, длины и т.д. Они могут быть **простыми** (тонны, штуки, литры, метры) и **сложными**, являющимися комбинацией нескольких разноименных величин (например объем транспортной работы при перевозке грузов выражается в тонно-километрах).
- **Условно-натуральные** единицы используются для соизмерения разнородных, но взаимозаменяемых по какому-либо свойству объектов, причем мера этого свойства и становится средством соизмерения.
- **Стоимостные единицы** измерения используются, например, для выражения объема разнородной продукции в стоимостной (денежной) форме – рублях, долларах и т.п.
- В **трудовых единицах** измерения (человеко-днях, человеко-часах) учитываются общие затраты труда на предприятии, трудоемкость отдельных операций технологического цикла.
- **Абсолютные величины** могут быть **положительными и отрицательными**. Например, результат деятельности предприятия (прибыль/убыток).



- **Относительная величина** в статистике представляет собой результат сопоставления двух статистических показателей (отношение двух абсолютных или относительных величин). Относительными величинами в статистике выражаются качественные показатели.
- Относительные величины, получаемые при сопоставлении абсолютных величин, могут быть названы относительными величинами первого порядка, а полученные при сопоставлении относительных же величин – величинами высших (второго, третьего и т.д.) порядков.
- Величина, находящаяся в числителе называется сравниваемой. Величина, находящаяся в знаменателе называется **базой сравнения** или **основанием**.

# Средние величины

- **Средней величиной** в статистике называется обобщающая количественная характеристика признака в статистической совокупности, отражающая типичный уровень этого признака в расчете на единицу совокупности.
- Средняя величина всегда именованная, она имеет ту же размерность, что и признак у отдельных единиц совокупности.
- Существуют различные **категории средних величин**. Наиболее распространены **степенные средние** (средняя арифметическая, средняя гармоническая, средняя квадратическая, средняя геометрическая) и **структурные средние** (мода, медиана).

# Степенные средние. Веса усреднения

- **Формула простой степенной средней:**

$$\bar{X} = \left( \frac{\sum_{i=1}^N X_i^k}{N} \right)^{\frac{1}{k}} = \sqrt[k]{\frac{X_1^k + X_2^k + \dots + X_N^k}{N}}$$

- где  $k$ -показатель степени, определяющий вид степенной средней.
- **С изменением показателя степени  $k$  формула степенной средней меняется, и в каждом отдельном случае приходим к определенному виду средней (гармонической, геометрической, арифметической, квадратической и т.д.).**

# Степенные средние. Веса усреднения

- **Формула взвешенной степенной средней:**

$$\overline{X} = \left( \frac{\sum_{i=1} X_i^k \cdot f_i}{\sum_{i=1} f_i} \right)^{\frac{1}{k}}$$

- где  $f_i$ -вес усреднения.
- **Величина средней взвешенной зависит уже не только от величины индивидуальных значений признака (как в простой средней), но и от соотношения весов. Например, чем больше веса у малых значений вариант, тем величина средней меньше.**

## Средняя арифметическая, гармоническая, геометрическая и квадратическая

- Показатель степени  $k$  в формуле степенной средней определяет вид степенной средней:
- При  $k = -1$  имеем гармоническую среднюю;
- При  $k = 0$  - среднюю геометрическую;
- При  $k = 1$  - среднюю арифметическую;
- При  $k = 2$  - среднюю квадратическую.

# Простая средняя арифметическая

- **Простая средняя арифметическая** исчисляется путем деления суммы значений признака на число значений:

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$$

где  $x_i$  – значение признака у  $i$ -ой единицы совокупности,  $N$  – объем совокупности.

**Пример:** За 10 февраля суточный пробег восьми автомобилей следующий, км: 185, 192, 210, 170, 214, 175, 188, 218.

Тогда, средний суточный пробег автомобиля составит:  
 $(185 + 192 + 210 + 170 + 214 + 175 + 188 + 218)/8 = 194,0$   
км.

# Взвешенная средняя арифметическая

Если данные представлены в виде ряда распределения, то расчет средней проводится по формуле **средней арифметической взвешенной**:

$$\bar{x} = \frac{\sum_{j=1}^k x_j \cdot f_j}{\sum_{j=1}^k f_j}$$

где  $x_j$  – значение признака в  $j$ -ой группе ( $j=1;k$ );  $k$  – число групп;  $f_j$  – вес усреднения для  $j$ -ой группы. В качестве весов усреднения берут частоты (частости).

Тарифный разряд (x)	3	4	5	6
Число рабочих (f)	7	10	11	3

Средний уровень квалификации рабочих, определяемый средним тарифным разрядом, исчисляется по приведенной выше формуле:

$$(3 \cdot 7 + 4 \cdot 10 + 5 \cdot 11 + 6 \cdot 3) / (7 + 10 + 11 + 3) = 4,3 \text{ разряда.}$$

# Взвешенная средняя арифметическая

Если значение признака в группе задано интервалом, то в качестве варианты  $x_j$  берется середина интервала (центральное значение):

$$x_j = \frac{x_j^{\text{в}} + x_j^{\text{н}}}{2}$$

где  $x_{\text{н}j}$ ,  $x_{\text{в}j}$  - нижняя и верхняя граница интервала.

Величина открытых интервалов приравнивается к величине примыкающих к ним соседних интервалов

<b>Стаж работы, лет (x)</b>	<b>До 5</b>	<b>5-10</b>	<b>10-15</b>	<b>15 и более</b>
<b>Число рабочих (f)</b>	<b>2</b>	<b>6</b>	<b>15</b>	<b>7</b>

Таким образом, средний стаж работы рабочих участка следующий:  
 $(2,5 \cdot 2 + 7,5 \cdot 6 + 12,5 \cdot 15 + 17,5 \cdot 7) / (2 + 6 + 15 + 7) = 12$  лет,  
где  $(0 + 5) / 2 = 2,5$ ;  $(5 + 10) / 2 = 7,5$  и т.д.



***Показатели центра  
распределения, медиана, мода.  
Методы их расчета для  
различных видов рядов  
распределения***

- **Показатели центра распределения** позволяют определить типичное значение признака в совокупности.
- К структурным характеристикам ряда распределения относят **квантили распределения** (медиана, квартили, децили и др.) и **моду**.
- **Квантили распределения** - это обобщающие показатели, характеризующие структуру распределения признака в совокупности.
- **Квантиль распределения** – это значение признака, занимающее определенное место в упорядоченной по данному признаку совокупности.

# Виды квантилей

- **1) медиана** ( $Me$ ) - значение признака, приходящееся на середину упорядоченной совокупности
- 
- **2) квартили** ( $Q_{1/4}, Q_{2/4}=Me, Q_{3/4}$ ) – значения признака, делящие упорядоченную совокупность на 4 равные (по числу единиц) части
- **3) децили** ( $Q_{0,1}, Q_{0,2}, \dots, Q_{0,9}$ ) – значения признака, делящие упорядоченную совокупность на 10 равных частей,
- **4) проценти** ( $Q_{0,01}, Q_{0,02}, \dots, Q_{0,99}$ ) - значения признака, делящие упорядоченную совокупность на 100 равных частей.

# Медиана

**Медиана** - это численное значение признака, которое находится в середине ранжированного ряда (расположенного в порядке возрастания или убывания) и делит этот ряд на две равные по численности части.

Для определения медианы сначала определяют ее место в ряду по формуле:

$$(n + 1)/2$$

где  $n$  - число членов ряда.

А затем берут значение признака, стоящего на этом месте.

# Мода

- **Мода (Mo)** – это наиболее часто встречающееся значение признака в совокупности.
- Для **дискретного ряда мода** – это значение признака, которому соответствует наибольшая частота (частость) распределения.
- Для **интервального ряда** – это значение признака, которому соответствует наибольшая плотность распределения.
- Если все значения вариационного ряда имеют одинаковую частоту, то говорят, что этот вариационный ряд **не имеет моды**.
- Если две не соседних варианты имеют одинаковую доминирующую частоту, то такой вариационный ряд называют **бимодальным**; если таких вариант больше двух, то ряд – **полимодальный**.

# Пример исчисления моды и медианы для дискретного ряда

Тарифный разряд (x)	3	4	5	6
Число рабочих (f)	7	10	11	3

Здесь модальным значением будет пятый тарифный разряд, так как этому значению соответствует наибольшая частота, равная 11 ( $M_0 = 5$  разрядам).

Для определения медианы сначала выясняется ее место (номер члена):

$$(31 + 1)/2 = 16.$$

Затем для нахождения положения медианы рассчитываются накопленные частоты (нарастающий итог частот, начиная с первого интервала), т.е. для первой группы - 7, для второй - 17, для третьей - 28, для четвертой - 31.

Следовательно, медианным значением будет четвертый тарифный разряд, так как 16-й номер соответствует второй группе накопленных частот.

# Пример исчисления моды и медианы для интервального ряда

Суточный пробег, км (x)	100-130	130-160	160-190	190-210	210-240
Число автомобилей (f)	70	160	130	85	20
Накопленные частоты (F)	70	230	360	445	465

При исчислении моды и медианы в интервальном ряду сначала определяют интервал, в котором они находятся. Среднее значение этого интервала соответствует их приближенному значению.

- В данном примере мода лежит в интервале 130-160, так как этому интервалу соответствует наибольшая частота (160). Среднее значение интервала - 145 км ( $M_0 = 145$  км).
- Место медианы - 233-й член, или  $[(465+1) / 2]$ , она лежит в интервале 160-190 и приближенно равна 175 км ( $M_e = 175$  км).

# **Вариационный анализ.**

## **Показатели вариации**



- **Вариацией признаков** называется различие численных значений признака у отдельных единиц совокупности.
- Размеры вариации позволяют судить, насколько однородна изучаемая группа и, следовательно, насколько характерна средняя по группе.
- Для измерения рассеяния (вариации) признака применяются различные **абсолютные** и **относительные** показатели вариации.

## **Абсолютные показатели вариации:**

- **Размах вариации**
- **Среднее линейное отклонение**
- **Среднее квадратическое отклонение**
- **Дисперсия**

## **Относительный показатель вариации:**

- **Коэффициент вариации**

# Размах вариации

- **Размах вариации (колебания)** -  $R$  - разность между максимальным и минимальным значениями признака в совокупности:

$$R = X_{\max} - X_{\min}$$

- **Недостаток:** он опирается только на два крайних значения признака и не учитывает степени колеблемости основной массы членов ряда.
- Однако он имеет теоретическое и практическое значение, так как определяет размер максимальных колебаний, который затем может сопоставляться с допустимым (или установленным по норме)

# Среднее линейное отклонение

- **Среднее линейное отклонение** -  $d$  -это средняя арифметическая абсолютных значений отклонений отдельных вариант от их средней арифметической

- Для несгруппированных данных (для первичного ряда):

$$d = \frac{\sum_{i=1}^N |x_i - \bar{x}|}{N}$$

- Для сгруппированных данных (для ряда распределения):

$$d = \frac{\sum_{j=1}^m |x_j - \bar{x}| \cdot f_j}{\sum_{j=1}^m f_j}$$

где  $N$  – объем совокупности;  $m$ - число групп;  $f_j$  – частота (частость) в  $j$ -ой группе

# Среднее квадратическое отклонение

**Среднее квадратическое отклонение** -  $\sigma$  - это средняя квадратическая из отклонений отдельных вариантов от их средней арифметической.

- Для несгруппированных данных (для первичного ряда):

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}}$$

- Для сгруппированных данных (для ряда распределения):

$$\sigma = \sqrt{\frac{\sum_{j=1}^m (x_j - \bar{x})^2 f_j}{\sum_{j=1}^m f_j}}$$

# Дисперсия

- **Дисперсия** –  $\sigma^2$  - это квадрат среднего квадратического отклонения.
- Она представляет собой средний квадрат отклонений вариант от их средней величины.
- Она может быть также вычислена, как разность среднего квадрата значения признака и квадрата среднего арифметического значения признака:

$$\sigma^2 = \overline{X^2} - (\overline{X})^2$$

# Относительный показатель вариации.

## Коэффициент вариации

- **Относительные показатели вариации** применяют, если необходимо оценить интенсивность вариации, или сравнить вариацию признака в различных совокупностях, или сравнить вариацию различных признаков.
- **Коэффициент вариации** – это выраженное в процентах отношение среднего квадратического отклонения к средней арифметической:

$$V = \frac{\sigma}{X} 100\%$$

- По коэффициенту вариации можно оценить характеристику однородности совокупности. Совокупность считается количественно однородной, если коэффициент вариации не превышает 33%.
- Коэффициент вариации является критерием надежности средней: если он велик (превышает 40 %), то это свидетельствует о большой колеблемости в величине признака у отдельных единиц данной группы, следовательно, средняя недостаточно надежна.

# Пример расчета показателей вариации

Тарифный разряд (x)	Участок №1				Участок №2			
	Число рабочих (f)	$x-x_{cp}$	$ x-x_{cp} *f$	$(x-x_{cp})^2f$	Число рабочих	$x-x_{cp}$	$ x-x_{cp} *f$	$(x-x_{cp})^2f$
2	1	-2,4	2,4	5,76	6	-2	12	24
3	3	-1,4	4,2	5,88	3	-1	3	3
4	15	-0,4	6,0	2,40	6	0	0	0
5	13	+0,6	7,8	4,88	5	+1	5	5
6	3	+1,6	4,8	7,68	5	+2	10	20
Итого	35	-	25,2	26,40	25	-	30	52

## Участок № 1

$x_{cp} = (2*1+3*3 + 4*15 + 5*13 + 6*3) / 35 = 4,4$  разряда;

$d = 25,2/35 = 0,72$  разряда;

$\sigma = \sqrt{(26,4/35)} = 0,87$  разряда;

$v = 0,87/4,4 *100= 19,8 \%$

## Участок № 2

$x_{cp} = (2*6 + 3*3 + 4*6 + 5*5 + 6*5) / 35 = 4,0$  разряда;

$d = 30 / 25= 1,20$  разряда;

$\sigma = \sqrt{(52/25)} = 1,44$  разряда;

$v = 1,44/4,0 *100= 36,0 \%$

## Вывод:

Полученные показатели подтверждают, что средние по участкам достаточно надежны, так как коэффициенты вариации не превышают 40 %. Состав рабочих по квалификации более однороден на участке № 1.