

**Федеральное государственное образовательное бюджетное
учреждение -высшего образования «Финансовый университет
при Правительстве Российской Федерации»
(Финансовый университет)**

Благовещенский филиал Финуниверситета

ОП.03 Статистика

УЧЕБНОЕ ПОСОБИЕ

(для самостоятельной работы студентов специальности
38.02.03 Операционная деятельность в логистике)

Благовещенск, 2023

Разработчик:

Прикота Т.Б., преподаватель высшей квалификационной категории Благовещенского филиала Финуниверситета

Рассмотрено на заседании ПЦК «Экономика, финансы и учёт» и рекомендовано к утверждению на заседании методического совета Благовещенского филиала Финуниверситета

Протокол от «15» ноября 2023 г. № 3

Председатель ПЦК «Экономика, финансы и учёт»

 Т.Б. Прикота

УТВЕРЖДАЮ

председатель методического совета, зам. директора по учебно-методической работе

 О.В. Ладоня

«22» ноября 2023 г.

ВВЕДЕНИЕ

Дисциплина «Статистика» является профилирующей дисциплиной в системе подготовки студентов по специальности 38.02.03 Операционная деятельность в логистике.

Целью данного пособия является формирование у студентов теоретической базы и практических навыков в области экономики и анализа деятельности предприятий в современных рыночных условиях.

В пособии представлены основные темы дисциплины, в которых рассматриваются основные теоретические понятия, методы расчета статистических показателей, представлены примеры решения задач.

СОДЕРЖАНИЕ

| | |
|--|----|
| ВВЕДЕНИЕ..... | 3 |
| Тема 1: Сводка и группировка статистических данных. Ряды распределения..... | 5 |
| Тема 2: Абсолютные и относительные величины, их взаимосвязь... | 9 |
| Тема 3: Средние величины. Структурные средние величины. Показатели вариации..... | 15 |
| Тема 4: Ряды динамики и их виды..... | 27 |
| Тема 5: Корреляционно-регрессионный анализ | 37 |
| СПИСОК ЛИТЕРАТУРЫ..... | 49 |

Тема 1: Сводка и группировка статистических данных. Ряды распределения.

Собранный в ходе наблюдения материал необходимо систематизировать для дальнейшей обработки. Для этого в статистике применяются статистическая сводка и группировка данных.

Статистическая сводка – это научно организованная обработка материалов статистического наблюдения в целях получения обобщенных характеристик изучаемого явления по ряду существенных для него признаков.

Сводка бывает различных видов, в зависимости от характеризующих ее признаков:

1. По глубине и точности обработки:

- ***простая сводка*** - это подсчет общих итогов по совокупности единиц наблюдения;

- ***сложная сводка*** включает комплекс операций:

- 1) выбор группировочного признака;
- 2) определение порядка формирования групп;
- 3) разработка системы показателей для характеристики групп и совокупности в целом;
- 4) разработка макета статистической таблицы;
- 5) распределение единиц совокупности на группы по признакам;
- 6) подсчет итогов;
- 7) представление результатов сводки в виде статистических таблиц.

2. По форме обработки:

- ***централизованная сводка*** проводится в один прием одним стат. органом, в котором сосредотачиваются все собранные материалы. Данный вид сводки используется для обработки материалов единовременных обследований;

- ***децентрализованная сводка*** – это поэтапная сводка, применяемая для стат. отчетности. Сначала стат. отчетность сводится стат. органами субъектов РФ, затем ее итоги сводятся в Росстате в целом по стране.

3. По технике выполнения:

- **Механизированная** – все операции производятся с помощью компьютера;
- **ручная** – все операции осуществляются вручную.

Со статистической сводкой неразрывно связана группировка статистических данных.

Группировка – разделение единиц изучаемой совокупности на качественно однородные группы по значениям одного или нескольких признаков. Такой признак называется **группировочным** (или основанием группировки).

Группировочные признаки могут быть **атрибутивными** и **количественными**.

В зависимости от задач исследования различают три вида группировок.

1. **Типологическая** – это группировка, при которой происходит разделение всей совокупности на качественно однородные группы.

2. **Структурная** – группировка, которая предназначена для изучения состава однородной совокупности по какому-либо варьирующему признаку или нескольким признакам.

3. **Аналитическая** – группировка, выявляющая взаимосвязи между изучаемыми признаками.

Число групп, на которое разбивается исходная совокупность зависит от вида группировочного признака, от степени его вариации, от численности единиц изучаемой совокупности.

Для того чтобы точно построить группировку, необходимо следовать определенным правилам.

1. Если группировочный признак атрибутивный или количественный, но количество его значений не велико, то число групп равно числу значений признака.
2. Если группировочный признак непрерывный, или дискретный, но количество его значений велико, то число групп определяется с помощью расчета интервалов.

Интервал – разница между максимальным и минимальным значениями признака в каждой группе.

Ширина интервала определяется по формуле:

$$i = \frac{x_{\max} - x_{\min}}{n_{cp}},$$

где x_{\max} и x_{\min} – наибольшее и наименьшее значения признака у единиц изучаемой совокупности;

n – число групп.

Число групп зависит от численности единиц совокупности N и определяется по номограмме Стерджесса:

| | | | | | | |
|-------------------------------------|-------|-------|-------|--------|---------|---------|
| Численность единиц совокупности (N) | 15-24 | 25-44 | 45-89 | 90-179 | 180-359 | 360-719 |
| Число групп (n) | 5 | 6 | 7 | 8 | 9 | 10 |

Пример:

Необходимо разгруппировать совокупность работников торговых предприятий по уровню месячной заработной платы. Объем совокупности составляет 170 человек. Минимальная з/п 18000 руб., максимальная – 30000 руб.

Решение:

1. Определяем нужное количество групп по номограмме Стерджесса:

так как $N = 170$, то $n = 8$

2. Определяем ширину интервала: $i = \frac{30000 - 18000}{8} = 1500$ руб.

3. Определяем границы интервалов. Группировка примет следующий вид:

| № группы | заработная плата, руб. |
|----------|------------------------|
| 1 | 18000 - 19500 |
| 2 | 19500 – 21000 |
| 3 | 21000 – 22500 |
| 4 | 22500 – 24000 |
| 5 | 24000 – 25500 |
| 6 | 25500 – 27000 |
| 7 | 27000 – 28500 |
| 8 | 28500 – 30000 |

Результат группировки оформляется рядом распределения.

Ряд распределения – это упорядоченное распределение единиц совокупности на группы по изучаемому варьирующему признаку.

Ряд распределения может быть **вариационным** (изучаемый варьирующий признак измеряется количественно) и **атрибутивным** (варьирующий признак является качественным).

Каждый ряд распределения имеет два основных элемента:

- **варианту (x)** – конкретное значение признака. Если варианты представлены целыми значениями признака, то ряд **дискретный**. Если варианты представлены числовыми интервалами, то ряд **интервальный**;

- **частоту (f)** – число, показывающее, как часто встречается данная варианта в совокупности.

Ряд распределения может быть дополнен накопленной частотой.

Накопленная частота (S) – это сумма частот, образуемая от группы к группе путем суммирования предыдущих частот ряда распределения. Последняя накопленная частота равна объему совокупности.

Ряды распределения часто анализируют с помощью их графического изображения. Для дискретного ряда распределения строится *полигон распределения*, а для интервального – *гистограмма*.

Пример:

Имеются данные о возрасте 24 студентов учебной группы (лет).

| | | | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|----|------|
| 20,1 | 21,4 | 20,5 | 20 | 19,6 | 22,3 | 21 | 22,6 | 20,2 | 19 | 19 | 23 |
| 22 | 20,1 | 20,3 | 21,8 | 19 | 20,3 | 21,2 | 21,5 | 22,3 | 23,1 | 20 | 22,2 |

Построить ряд распределения студентов группы по полному числу лет, указать вид ряда, его элементы. Построить гистограмму распределения. Сделать выводы.

Решение:

1. Проранжируем значения признака по возрастанию:

| | | | | | | | | | | | |
|----|------|------|------|------|------|------|------|------|------|------|------|
| 19 | 19 | 19 | 19,6 | 20 | 20 | 20,1 | 20,1 | 20,2 | 20,3 | 20,3 | 20,5 |
| 21 | 21,2 | 21,4 | 21,5 | 21,8 | 21,9 | 22,2 | 22,3 | 22,3 | 22,6 | 23 | 23,1 |

2. По номограмме Стерджесса определяем число групп:

Так как $N = 24$, то $n = 5$

3. Рассчитываем ширину интервала:

$$i = (23,1-19) / 5 = 0,82 \text{ года.}$$

Округляем значение в большую сторону, то есть получаем 1 год.

4. Определяем границы интервалов и количество единиц совокупности в каждой группе. Строим ряд распределения:

| Полное число лет (x) | Количество студентов (f) |
|----------------------|--------------------------|
| 19 - 20 | 4 |
| 20 - 21 | 8 |
| 21 - 22 | 6 |
| 22 - 23 | 4 |
| 23 и более | 2 |

Вид ряда - вариационный интервальный.

Вариантой является возраст студентов.

Частотой является количество студентов определенного возраста



Вывод: Наибольшее количество студентов группы имеет возраст от 20 до 22 лет.

Тема 2: Абсолютные и относительные величины, их взаимосвязь

Первоначальным видом обобщающих статистических показателей являются **абсолютные величины**.

Абсолютные статистические величины получают при сборе первичных данных. Они могут быть *измеряемыми* или *расчетными*.

Абсолютные показатели могут иметь различные единицы измерения (см. рис. 1).

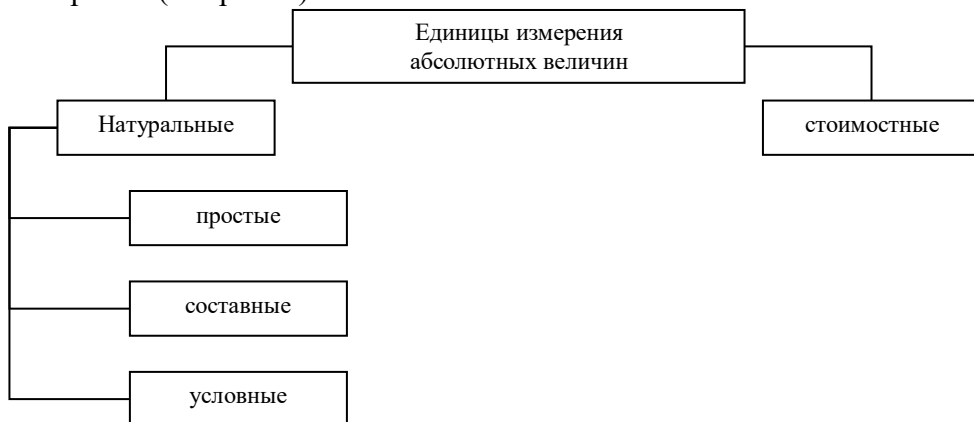


Рис. 1 – Единицы измерения абсолютных величин

При статистическом анализе проводится сопоставление статистических показателей во времени, в пространстве, изучается структура явления и ее изменения и т.д., поэтому наличия только абсолютных величин недостаточно.

Для решения этих задач используются относительные величины.

Относительные величины представляют собой меру количественного соотношения статистических показателей.

Если сравнивать *одноименные* величины и основание принять за единицу, то относительная величина выразится в форме коэффициента. Она покажет, во сколько раз сравниваемая величина больше основания.

Если основание принято за 100, то относительная величина будет выражена в процентах (%), если основание принято за 1000, то относительная величина будет выражена в промилле (‰).

Если сравниваемая величина больше основания, то ее удобнее выразить коэффициентом, если меньше, то в процентах, если значительно меньше, то в промилле.

Если сравниваемые величины *разноименные*, то их относительная величина будет иметь сложную единицу измерения, образуемую от наименований единиц измерения сравниваемых показателей (км/час, ц/га, руб./шт., м²/чел и т.д.).

В зависимости от целей статистического анализа различают несколько видов относительных показателей:

1. **Относительный показатель планового задания** показывает, на сколько процентов планируется увеличить показатель в текущем периоде по сравнению с предыдущим периодом.

$$\text{ОППЗ} = \frac{\text{Плановый уровень показателя на текущий период}}{\text{уровень показателя предшествующего периода}} \cdot 100$$

Пример:

Запланирован выпуск продукции на предстоящий период в объеме 375000 тыс. руб. В текущем периоде выпуск данной продукции составил 348000 тыс. руб.

$$\text{ОППЗ} = \frac{375000}{348000} \cdot 100 = 107,8 \%$$

Таким образом, запланировано увеличить выпуск продукции в планируемом периоде на 7,8 %

2. **Относительный показатель выполнения плана** показывает, на сколько процентов выполнен план.

$$\text{ОПВП} = \frac{\text{Фактический уровень показателя}}{\text{уровень показателя по плану}} \cdot 100$$

Пример:

Предприятием было запланировано продать продукции в январе на сумму 3450 тыс. руб. Фактически объем продаж за январь составил 3280 тыс. руб.

$$\text{ОПВП} = \frac{3280}{3450} \cdot 100 = 95,1 \%$$

Таким образом, предприятие недовыполнило план по продажам на 4,9%.

В плане предприятия так же могут иметься показатели, которые предусматривают не рост, а снижение уровней. Например, запланированное снижение себестоимости продукции, уменьшение управленческих расходов и т.д.

3. Относительный показатель динамики (темп) показывает, как быстро изменяется показатель в текущем периоде по сравнению с предыдущим.

$$\text{ОПД} = \frac{\text{Фактический уровень показателя в отчетном периоде}}{\text{Фактический уровень показателя в базисном периоде}} \cdot 100$$

Относительный показатель динамики называют **темпом роста**. Если из темпа роста вычесть 100, получится **темпы прироста**, который покажет, на сколько процентов изменился показатель.

Пример:

Себестоимость продукции, выпущенной предприятием за январь, составила 2120 тыс. руб. За февраль себестоимость выпущенной продукции зафиксирована в размере 2175 тыс. руб.

$$\text{ОПД} = \frac{2175}{2120} \cdot 100 = 102 \%$$

Таким образом, рост себестоимости продукции в феврале по сравнению с мартом составил 102 %, то есть показатель увеличился на 2 %.

Между показателями планового задания, выполнения плана и динамики существует взаимосвязь:

$$\text{ОПД} = \text{ОППЗ} \cdot \text{ОПВП}$$

4. Относительный показатель структуры характеризует состав изучаемой совокупности и отражает удельный вес (долю) каждой части в целом (в %).

$$\text{ОП структуры} = \frac{\text{Часть}}{\text{Целое}} \cdot 100$$

Пример:

В совокупности предприятий города, состоящей из 3450 единиц, по итогам года чистую прибыль от своей деятельности фиксировали 2830 предприятий. Остальные предприятия фиксировали чистый убыток. Определить долю убыточных предприятий.

Так как нужно рассчитать долю убыточных предприятий, сначала необходимо определить их число: $3450 - 2830 = 620$ единиц.

$$\text{ОП структуры} = \frac{620}{3450} \cdot 100 = 17\%$$

Таким образом, в совокупности предприятий города по итогам года зафиксировано 17 % убыточных предприятий.

5. Относительный показатель координации показывает, сколько единиц части, стоящей в числителе, приходится на единицу другой части, стоящей в знаменателе.

$$\text{ОП координации} = \frac{\text{Часть 1}}{\text{Часть 2}}$$

В некоторых случаях, показатель координации удобно рассчитывать не на 1, а на 100 или 1000 единиц.

Из предыдущего примера:

$$\text{ОП координации} = \frac{620}{2830} = 0,21 \text{ единицы}$$

Если оставить результат в виде десятичной дроби, то ответ будет звучать так: на каждое прибыльное предприятие города приходится 0,21 убыточного предприятия.

Для удобства примем основание за 100. Тогда можно сказать, что на каждые 100 прибыльных предприятий в городе приходится 21 убыточное предприятие.

6. Относительный показатель наглядности (сравнения).

Это отношение одноименных показателей, относящихся к одному моменту или периоду времени, но разным территориям или объектам.

$$\text{ОП сравнения} = \frac{\text{показатель 1}}{\text{показатель 2}}$$

С помощью этого показателя определяется, во сколько раз показатель в числителе больше показателя в знаменателе (например, во сколько раз численность населения одной страны превышает численность населения другой страны на один и тот же момент времени).

7. Относительные показатели интенсивности развития

Данные показатели характеризуют плотность распространения явления в определенной среде (рождаемость, смертность и т.д.).

В группу этих показателей относят:

- показатели рождаемости, смертности
- показатели потребления на душу населения
- показатели обеспеченности населения какими-либо благами
- показатели производства продукции на душу населения и т.д.

Например ВВП на душу населения рассчитывается следующим образом:

$$\text{ВВП на душу населения} = \frac{\text{Объем выпущенной продукции}}{\text{Среднегодовая численность населения}}$$

По данным Федеральной службы государственной статистики ВВП РФ в текущих ценах за 2015 год составил 80804,3 млрд. руб. Численность населения РФ на конец 2015 года составляла 146544,7 тыс. чел.

Соответственно

$$\text{ВВП на душу населения} = \frac{80804,3}{146544,7} = 551397 \text{ руб./чел.}$$

Тема 3: Средние величины. Показатели вариации.

Средняя величина является очень распространенным обобщающим показателем в статистике. Только с помощью средних можно охарактеризовать совокупность по количественному варьирующему признаку. Они показывают уровень признака в расчете на единицу совокупности.

Средняя – это обобщающая числовая характеристика совокупности по какому-либо количественно варьирующему признаку, определяющая уровень признака в расчете на единицу совокупности (средняя процентная ставка, средний балл студента, средняя заработная плата и т.д.).

В средних величинах погашаются индивидуальные различия единиц совокупности, обусловленные случайными факторами.

Виды средних величин

В статистике применяют *степенные* и *структурные* средние величины.

Общая формула степенной средней имеет следующий вид:

$$\bar{x} = \sqrt[m]{\frac{\sum x^m}{n}}$$

где \bar{x} – степенная средняя;
 x – варианта;
 n – число вариантов;
 m – показатель степени средней;
 \sum – знак суммирования.

Из степенных средних наиболее часто применяются средняя арифметическая, средняя гармоническая и средняя геометрическая.

Степенные средние могут быть рассчитаны различными способами, в зависимости от того, сгруппированы данные или нет.

Методы расчета средних величин.

1. Средняя арифметическая:

а) если данные не сгруппированы или сгруппированы, но все частоты равны, применяют формулу *средней арифметической простой*:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum x_i}{n},$$

где $x_1, x_2 \dots x_n$ – значение каждой варианты;
 n – число вариант.

б) если данные сгруппированы, а частоты не равны, применяют *среднюю арифметическую взвешенную*:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_n f_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum x f}{\sum f},$$

где $x_1, x_2 \dots x_n$ – значения варианты,
 $f_1, f_2 \dots f_n$ – частоты.

Взвешиванием в статистике называется умножение варианты на частоту. **Весом** является частота.

Пример: Расчет средней арифметической в дискретном ряду

| Заработная плата, руб. | Численность работающих, чел. |
|------------------------|------------------------------|
| 34000 | 6 |
| 32000 | 8 |
| 46000 | 6 |
| Итого | 20 |

$$\bar{x} = \frac{34000 \cdot 6 + 32000 \cdot 8 + 46000 \cdot 6}{20} = 36800 \text{ руб.}$$

При расчете средней в интервальном ряду определяют середины интервалов. Далее вычисления проводят как в дискретном ряду – варианты умножают на частоты (веса) и сумму произведений делят на сумму частот (весов).

Пример: Расчет средней арифметической в интервальном ряду

| Размер заработной платы, руб. | Количество рабочих, чел. | Средние значения в интервалах, руб. |
|-------------------------------|--------------------------|-------------------------------------|
| 34000-36000 | 15 | 35000 |
| 36000-38000 | 11 | 37000 |
| 38000-40000 | 8 | 39000 |
| 40000-42000 | 4 | 41000 |
| Итого | 38 | |

$$\bar{x} = \frac{35000 \cdot 15 + 37000 \cdot 11 + 39000 \cdot 8 + 41000 \cdot 4}{38} = 37053 \text{ руб.}$$

2. Средняя гармоническая.

Средняя гармоническая – это величина, вычисленная из обратных значений варьирующего признака.

Она применяется в тех случаях, когда неизвестны значения частот у вариант ряда, зато имеются для каждого x произведения этих вариант на соответствующие им частоты, то есть $W = xf$ (товарооборот по видам товаров при расчете их средней цены, фонд заработной платы по отдельным категориям работников при расчете средней заработной платы и т.д.)

а) если произведения вариант на соответствующие им частоты равны между собой, то есть $W_1 = W_2 = \dots = W_n$, то можно применять формулу **средней гармонической простой**:

$$\bar{x} = \frac{n}{\sum \frac{1}{x}}$$

где n – число единиц совокупности;

x – варианта;

б) если произведения вариант на соответствующие им частоты не равны между собой, то есть $W_1 \neq W_2 \neq \dots \neq W_n$, применяется формула **средней гармонической взвешенной**, которая имеет следующий вид:

$$\bar{x} = \frac{\sum W}{\sum \frac{W}{x}}$$

где W – произведения вариант на соответствующие им частоты.

Пример:

1) в магазине четыре отдела, продающих яблоки. Определить среднюю цену яблок.

| Дневная выручка отдела, руб. (W) | Цена яблок, руб. (x) |
|--------------------------------------|--------------------------|
| 8200 | 98,00 |
| 8200 | 112,00 |
| 8200 | 116,00 |
| 8200 | 125,00 |

Так как неизвестно количество проданных яблок, а выручка (то есть произведение цены на количество, то есть $xf = W$) одинакова во всех отделах, среднюю цену найдем по формуле средней гармонической простой:

$$\bar{x} = \frac{4}{\frac{1}{98} + \frac{1}{112} + \frac{1}{116} + \frac{1}{125}} = 111,88 \text{ руб}$$

2) заработная плата продавцов и фонд оплаты труда за месяц в торговом предприятии составили:

| Секции | Зарботная плата, руб. | Фонд оплаты труда, руб. |
|--------|-----------------------|-------------------------|
| 1 | 18000 | 54000 |
| 2 | 19600 | 78400 |
| 3 | 22400 | 67200 |
| 4 | 28200 | 141000 |
| Итого | | 340600 |

Вычислить среднюю заработную плату по четырем секциям в целом.

Так как изначально нам неизвестно количество продавцов, а известен фонд заработной платы (произведение заработной платы на количество работников) мы применяем формулу средней гармонической взвешенной:

$$\bar{x} = \frac{\sum W}{\sum \frac{W}{x}} = \frac{340600}{\frac{54000}{18000} + \frac{78400}{19600} + \frac{67200}{22400} + \frac{141000}{28200}} = 22707 \text{ руб.}$$

Вычисление средних из относительных величин

При расчете средней из относительных величин сначала составляется исходная формула относительной величины.

Если неизвестен числитель относительной величины, то расчет производится по формуле *средней арифметической взвешенной*.

Если при расчете средней из относительных величин неизвестен знаменатель относительной величины, используется формула *средней гармонической взвешенной*.

Пример:

Нужно оценить среднее выполнение плана по группе предприятий в первом и во втором кварталах.

Выполнение плана реализации товаров магазинами

| Магазин | 1 квартал | | 2 квартал | |
|---------|---------------------|---------------------------------|---------------------|-------------------------------------|
| | Выполнение плана, % | Товарооборот по плану, тыс. руб | Выполнение плана, % | Товарооборот фактический, тыс. руб. |
| 1 | 101,3 | 3500 | 102,0 | 4401,6 |
| 2 | 105,1 | 1200 | 104,8 | 1060,8 |
| Итого | | 4700 | | 5462,4 |

Решение:

Рассчитываемой средней величиной является процент выполнения плана. Исходная формула относительного показателя выполнения плана:

$$\text{ОПВП} = \frac{\text{Фактический уровень показателя}}{\text{уровень показателя по плану}} \cdot 100$$

В первом квартале нам неизвестен числитель отношения.

Значит, средний процент выполнения плана рассчитывается по формуле средней арифметической взвешенной, где за x примем процент выполнения плана, а за f – товарооборот по плану:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{101,3 \cdot 3500 + 105,1 \cdot 1200}{3500 + 1200} = \frac{480670}{4700} = 102,3 \%$$

Следовательно, в I квартале два магазина перевыполнили план по реализации в среднем на 2,3 %.

Во втором квартале нам неизвестен знаменатель исходной формулы. Следовательно, применим формулу средней гармонической взвешенной, где x примем процент выполнения плана, а за W – фактический товароборот:

$$\bar{x} = \frac{\sum W}{\sum \frac{W}{x}} = \frac{4401,6 + 1060,8}{\frac{4401,6}{102,0} + \frac{1060,8}{104,8}} = \frac{5462,4}{53,275} = 102,5 \%$$

Таким образом, во втором квартале план по реализации по двум магазинам в среднем перевыполнен на 2,5 %.

Структурные средние величины

Структурные средние величины – это вспомогательные характеристики изучаемой совокупности. В отличие от степенных средних они имеют не обобщающее значение признака, а вполне конкретное значение, то есть значение одной из вариантов.

К структурным средним величинам относятся *мода* и *медиана*.

Мода – это варианта с наибольшей частотой.

Мода характеризует наиболее часто встречающееся значение признака (размер одежды, пользующийся наибольшим спросом, самая распространенная заработная плата и т.д.).

Медиана – варианта, которая делит вариационный ряд на две равные части таким образом, что половина единиц совокупности имеет значение признака меньше, чем медиана, а половина – больше, чем медиана.

Расчет моды и медианы рассчитываются по-разному в дискретном и вариационном рядах.

1. Расчет моды и медианы в дискретном ряду.

Расчет моды в дискретном ряду.

Рассмотрим пример:

Группировка семей по количеству детей

| Группа семей по числу детей | Число семей |
|-----------------------------|-------------|
| 0 | 10 |
| 1 | 45 |
| 2 | 64 |
| 3 | 38 |
| 4 | 22 |
| 5 | 8 |
| 6 | 4 |
| Итого | 191 |

Модой здесь будет семья, имеющая 2 детей, так как этому значению варианты соответствует наибольшее число семей – 64.

Это значит, что чаще всего в данной совокупности встречаются семьи, имеющие 2-х детей.

Если все варианты встречаются одинаково часто, то они **одинаково модальны**.

Если две варианты имеют наибольшие частоты, то такое распределение будет **бимодальным**.

Расчет медианы в дискретном ряду

Чтобы найти медиану в дискретном ряду сначала определяют ее порядковый номер. Для этого необходимо сумму частот разделить пополам. Если сумма частот является нечетным числом, то к полученному результату необходимо прибавить 0,5.

$$\text{Номер } M_e = \left(\frac{\sum f}{2} + 0,5 \right).$$

В нашем примере медианой будет 96- вариант.

Чтобы определить ее значение необходимо накопить частоты.

Накопленная частота – это численность единиц, образуемая от группы к группе путем суммирования предыдущих частот

(нарастающим итогом). Последняя накопленная частота равна объему совокупности (семе частот).

| Группа семей по числу детей | Число семей | Накопленная частота |
|-----------------------------|-------------|---------------------|
| 0 | 10 | 10 |
| 1 | 45 | 55 |
| 2 | 64 | 119 |
| 3 | 38 | 157 |
| 4 | 22 | 179 |
| 5 | 8 | 187 |
| 6 | 4 | 191 |
| Итого | 191 | |

Находим группу, к которой относится 96-я варианта. Это третья группа.

Значит, медианой будет семья, имеющих двоих детей – половина семей имеет 2-х детей и более.

2. Расчет моды и медианы в интервальном вариационном ряду

В интервальном вариационном ряду для нахождения значения моды и медианы в пределах интервала необходимо производить дополнительные расчеты.

Рассмотрим пример:

Группировка рабочих по размеру годовой заработной платы.

| Группа рабочих по размеру годовой заработной платы, тыс. руб. | Число рабочих |
|---|---------------|
| 130-140 | 10 |
| 140-150 | 50 |
| 150-160 | 100 |
| 160-170 | 115 |
| 170-180 | 180 |
| 180-190 | 45 |
| Итого | 500 |

1) Для нахождения моды в интервальном ряду необходимо сначала определить **модальный интервал** данного ряда. То есть тот интервал, которому принадлежит наибольшая частота.

В нашем примере модальным будет интервал 170-180, так как максимальная частота – 180.

Для расчета моды этого интервала применяют следующую формулу:

$$M_o = x_{M_o} + i_{M_o} \cdot \frac{f_{M_o} - f_{M_{o-1}}}{(f_{M_o} - f_{M_{o-1}}) + (f_{M_o} - f_{M_{o+1}})}$$

где

x_{M_o} – нижняя граница модального интервала,

i_{M_o} – величина модального интервала,

f_{M_o} – частота модального интервала,

$f_{M_{o-1}}$ – частота домодального интервала,

$f_{M_{o+1}}$ – частота интервала, следующего за модальным.

В нашем примере получим:

$$M_o = 170 + 10 \cdot \frac{180 - 115}{(180 - 115) + (180 - 45)} = 170 + 10 \cdot \frac{65}{200} = 173,25 \text{ тыс. руб.}$$

Вывод: наибольшее число рабочих имеет заработную плату 173,25 тыс. руб. в год.

2) Для нахождения медианы в интервальном ряду нужно определить **медианный интервал**.

Медианным будет такой интервал, накопленная частота которого равна или превышает половину суммы частот.

| Группа рабочих по размеру годовой заработной платы, тыс. руб. | Число рабочих | Накопленная частота |
|---|---------------|---------------------|
| 130-140 | 10 | 10 |
| 140-150 | 50 | 60 |
| 150-160 | 100 | 160 |
| 160-170 | 115 | 275 |
| 170-180 | 180 | 455 |
| 180-190 | 45 | 500 |
| Итого | 500 | |

В нашем примере половина суммы частот равна 250. Следовательно медианным будет интервал 160-170.

Формула для исчисления медианы имеет следующий вид:

$$M_e = x_{M_e} + i_{M_e} \cdot \frac{\frac{f}{2} - S_{M_e}}{f_{M_e}},$$

где

x_{M_e} – нижняя граница медианного интервала,

i_{M_e} – величина медианного интервала,

$\sum f$ – сумма частот ряда,

S_{M_e} – сумма накопленных частот в интервалах, предшествующих медианному,

f_{M_e} – частота медианного интервала.

Подставляя значения из нашего примера получим:

$$M_e = 160 + 10 \cdot \frac{\frac{500}{2} - 160}{115} = 167,8 \text{ тыс. руб.}$$

То есть половина рабочих получает заработную плату 167,8 тыс. руб. в год и выше.

Показатели вариации

В процессе статистического анализа может сложиться ситуация, когда средние величины различных совокупностей совпадают, а значения признака у каждой единицы резко различаются между собой.

Измерить степень вариации признака и оценить достоверность средней величины можно с помощью показателей вариации.

Существуют *абсолютные* и *относительные* показатели вариации.

К абсолютным показателям вариации относятся:

- размах вариации
- среднее линейное отклонение

- дисперсия
- среднее квадратическое отклонение

К относительным показателям вариации относятся различные коэффициенты, в частности *коэффициент вариации*.

Расчет показателей вариации.

1. Размах вариации вычисляется как разность между максимальным и минимальным значениями признака:

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

2. Среднее линейное отклонение показывает индивидуальные отклонения значения признака от средней величины.

Для несгруппированных данных используют простое линейное отклонение:

$$\bar{d} = \frac{\sum |x - \bar{x}|}{n}$$

Для сгруппированных данных используют взвешенное линейное отклонение:

$$\bar{d} = \frac{\sum |x - \bar{x}| f}{\sum f}$$

3. Дисперсия (σ^2) (средний квадрат отклонений) – это средняя арифметическая величина, полученная из квадратов отклонений значений признака от их средней:

простая:

$$\sigma_{\text{прост}}^2 = \frac{\sum (x - \bar{x})^2}{n}$$

взвешенная:

$$\sigma_{\text{взвеш}}^2 = \frac{\sum (x - \bar{x})^2 f}{\sum f}$$

4. Среднее квадратическое отклонение
простое:

$$\sigma_{\text{прост}} = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

взвешенное:

$$\sigma_{\text{взвеш}} = \sqrt{\frac{\sum (x - \bar{x})^2 f}{\sum f}}$$

Абсолютные показатели вариации, кроме дисперсии, имеют те же единицы измерения, что и исследуемый показатель вариационного ряда.

Расчет коэффициента вариации производится по формуле:

$$v_{\sigma} = \frac{\sigma}{x} \cdot 100\%$$

Коэффициент вариации является критерием типичности средней.

Так, если $v > 33\%$, это означает, что средняя характеризует совокупность по признаку, который существенно изменяется у отдельных единиц. Типичность такой средней невелика, а совокупность считается неоднородной.

Если $v < 33\%$, то совокупность считается однородной, а типичность средней высокой.

Пример:

Определить средний размер построенных квартир. Рассчитать среднее квадратическое отклонение. Оценить типичность средней величины с помощью коэффициента вариации

| Метраж квартир | Число построенных квартир |
|----------------|---------------------------|
| 40-50 | 150 |
| 50-60 | 240 |
| 60-70 | 680 |
| 70-80 | 820 |
| 80-90 | 790 |
| 90-100 | 450 |
| итого | 3130 |

В задании дан интервальный вариационный ряд распределения, поэтому для определения средней величины

необходимо преобразовать этот ряд в дискретный путем нахождения середины интервалов.

Для решения построим вспомогательную таблицу:

| Метраж квартир, м ² | Число построенных квартир, ед. | Средины интервалов, м ² | $x \cdot f$ | $(x - \bar{x})^2 f$ |
|--------------------------------|--------------------------------|------------------------------------|-------------|---------------------|
| 40-50 | 150 | 45 | 6750 | 137310,1 |
| 50-60 | 240 | 55 | 13200 | 98469,35 |
| 60-70 | 680 | 65 | 44200 | 71520,46 |
| 70-80 | 820 | 75 | 61500 | 53,56797 |
| 80-90 | 790 | 85 | 67150 | 75013,27 |
| 90-100 | 450 | 95 | 42750 | 175428,8 |
| итого | 3130 | | 235550 | 557795,5 |

Так как данные сгруппированы, а частоты не равны, для расчета средней величины применяем формулы арифметической взвешенной:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{235550}{3130} = 75,3 \text{ м}^2$$

Среднее квадратическое отклонение:

$$\sigma_{\text{взвеш}} = \sqrt{\frac{\sum (x - \bar{x})^2 f}{\sum f}} = \sqrt{\frac{557795,5}{3130}} = 13,3 \text{ м}^2$$

Коэффициент вариации:

$$v_{\sigma} = \frac{\sigma}{\bar{x}} \cdot 100\% = \frac{13,3}{75,3} \cdot 100 = 17,7\%$$

Размер построенных квартир в среднем составляет 75,3 м² с отклонением в среднем на 13,3 м². Рассчитанный показатель вариации не превышает 33 %, следовательно, совокупность однородна по изучаемому признаку, а средняя величина для данной совокупности является типичной.

Ряды динамики используются для отражения действительности экономических явлений и процессов при переходе от одного момента или периода времени к другому. На основе анализа рядов динамики выявляют основные тенденции изменения явлений, строят модели для прогнозирования их развития.

Рядом динамики называется ряд значений статистического показателя, расположенных в хронологическом порядке и характеризующих развитие явления во времени.

Динамический ряд состоит из двух элементов:

- 1) **моментов** или **временных периодов** (t), по отношению к которым, приводятся статистические данные;
- 2) **уровней динамического ряда** (y) - значений **статистического показателя**, характеризующих размер рассматриваемого явления в соответствующие моменты или периоды времени.

Классификация рядов динамики

1. По признаку времени различают: **моментные** и **интервальные** ряды динамики.

В **моментном ряду** динамики представлены значения статистического показателя по состоянию на определенные последовательные моменты времени (**на** начало или конец рассматриваемого года, месяца, квартала). К таким рядам относятся: численность населения на начало года, запас материалов на конец месяца и т.д.

Особенности моментного ряда:

- уровни повторяются друг в друге (переходящий остаток)
- уровни не отражают продолжительность периода, в течение которого сохраняется их размер
- уровни не обладают свойством суммарности.

Интервальный ряд динамики – это совокупность значений статистического показателя, относящихся к нескольким последовательным периодам времени (**за** ряд лет, кварталов,

месяцев). К таким рядам относятся: ряд динамики произведенной продукции, отработанного времени, ВВП, инвестиций и т.д.

Особенности интервального ряда:

- уровни не повторяются друг в друге, то есть являются новыми по отношению к предыдущему периоду
- уровни отражают длину периода, за который они приведены
- уровни обладают свойством суммарности, то есть их можно складывать.

При суммировании показателей интервального ряда получается ряд динамики с нарастающим итогом (общий выпуск продукции за определенный период времени, общие затраты рабочего времени, общий объем продаж акций и т.д.).

Сумма же уровней моментного ряда не имеет смысла.

2. По расстоянию между датами или интервалами времени выделяют *полные* и *неполные динамические ряды*.

В *полных* динамических рядах даты или периоды времени указываются через равный интервал.

В *неполных* динамических рядах расстояние между датами или периодами неравно.

3. По числу показателей можно выделить *изолированные* и *комплексные (многомерные)* ряды динамики.

Если ведется анализ во времени одного показателя, то ряд динамики *изолированный* (например, число предприятий Российской Федерации).

В *многомерном* ряду представлена динамика нескольких показателей, характеризующих одно явление.

4. По форме представления уровня ряда различают ряды:

- *абсолютных* величин (количество предприятий, объем выпуска продукции, численность персонала и т.д.),

- ряды *относительных* величин (изменение во времени удельных весов, темпов роста, показателей интенсивности и т.д.),
- ряды *средних* величин (изменение во времени средней заработной платы, среднедушевых доходов и т.д.).

Цепные и базисные показатели рядов динамики

В зависимости от того, с какой базой, постоянной или переменной, исчисляются показатели, их делят на цепные и базисные.

При исчислении *цепных показателей* сравнение каждого уровня производится с предыдущим уровнем.

При исчислении *базисных показателей* сравнение каждого уровня производится с начальным уровнем (первым в ряду) или каким-то другим, принятым за базу сравнения.

При расчете показателей приняты следующие *условные обозначения*:

y_i – уровень любого периода (кроме начального), называемый уровнем текущего периода;

y_{i-1} – уровень периода, предшествующий текущему;

y_0 – базисный уровень, то есть принятый за постоянную базу сравнения (часто начальный уровень).

Расчет аналитических показателей ряда.

1. Абсолютный прирост показывает, на сколько абсолютных единиц изменился показатель:

а) *базисный*:

$$\Delta_{\text{б}} = y_i - y_0$$

б) *цепной*:

$$\Delta_{\text{ц}} = y_i - y_{i-1}$$

2. Темп роста показывает скорость изменения показателя. Может выражаться в виде коэффициента или процента:

а) *базисный*:

$$T_{p\delta} = \frac{y_i}{y_{\delta}} \cdot 100$$

б) *цепной*:

$$T_{pц} = \frac{y_i}{y_{i-1}} \cdot 100$$

3. Темп прироста характеризует абсолютный прирост в относительных величинах:

а) *базисный*:

$$T_{п\delta} = \frac{\Delta_{\delta}}{y_{\delta}} \cdot 100$$

$$T_{п\delta} = T_{p\delta} - 100$$

б) *цепной*:

$$T_{пц} = \frac{\Delta_y}{y_{i-1}} \cdot 100 \%$$

$$T_{пц} = T_{pц} - 100$$

4. Абсолютное значение 1% прироста обозначает насколько весом каждый процент прироста, какое содержание за ним скрывается:

а) *базисный*:

$$A_{\delta} = \frac{\Delta_{\delta}}{T_{п\delta}}, \quad A_{\delta} = \frac{y_{\delta}}{100}$$

б) *цепной*:

$$A_{ц} = \frac{\Delta_y}{T_{пц}}, \quad A_{ц} = \frac{y_{i-1}}{100}$$

Пример:

Имеются данные о выпуске продукции на предприятии:

| Год | 2014 | 2015 | 2016 | 2017 | 2018 |
|------------------------|------|------|------|------|------|
| Выпуск продукции, тонн | 718 | 891 | 1240 | 1621 | 1862 |

Рассчитать показатели динамики

Решение:

За базисный период принимается 2014 год.

Рассчитаем показатели в 2015 году:

- абсолютный прирост:

$$\Delta_{ц} = y_i - y_{i-1} = 891 - 718 = 173 \text{ т.}$$

$$\Delta_{б} = y_i - y_{б} = 173 \text{ т.}$$

- темп роста:

$$T_{рц} = \frac{y_i}{y_{i-1}} \cdot 100 = \frac{891}{718} \cdot 100 = 124,1 \%$$

$$T_{рб} = \frac{y_i}{y_{б}} \cdot 100 = 124,1 \%$$

- темп прироста:

$$T_{пц} = T_{рц} - 100 = 24,1 \%$$

$$T_{пб} = T_{рб} - 100 = 24,1 \%$$

- абсолютное значение 1 % прироста:

$$A_{ц} = \frac{y_{i-1}}{100} = \frac{718}{100} = 7,18 \text{ т.}$$

$$A_{б} = \frac{y_{б}}{100} = 7,18 \text{ т.}$$

Рассчитаем показатели в 2016 году:

- абсолютный прирост:

$$\Delta_{ц} = y_i - y_{i-1} = 1240 - 891 = 349 \text{ т.}$$

$$\Delta_{б} = y_i - y_{б} = 1240 - 718 = 522 \text{ т.}$$

- темп роста:

$$T_{рц} = \frac{y_i}{y_{i-1}} \cdot 100 = \frac{1240}{891} \cdot 100 = 139,2 \%$$

$$T_{рб} = \frac{y_i}{y_{б}} \cdot 100 = \frac{1240}{718} \cdot 100 = 172,7 \%$$

- темп прироста:

$$T_{пц} = T_{рц} - 100 = 39,2 \%$$

$$T_{пб} = T_{рб} - 100 = 72,7 \%$$

- абсолютное значение 1 % прироста:

$$A_{ц} = \frac{y_{i-1}}{100} = \frac{891}{100} = 8,91 \text{ т.}$$

$$A_{б} = \frac{y_{б}}{100} = 7,18 \text{ т.}$$

Аналогично рассчитываются показатели в остальных периодах.

| год | Выпуск продукции, т. | Абсолютный прирост, т. | | Темп роста, % | | Темп прироста, % | | Абсолютное значение 1% прироста, т. | |
|------|----------------------|------------------------|------|---------------|-------|------------------|-------|-------------------------------------|------|
| | | Цеп. | Баз. | Цеп. | Баз. | Цеп. | Баз. | Цеп. | Баз. |
| 2014 | 718 | - | 0 | - | 100 | - | 0 | - | - |
| 2015 | 891 | 173 | 173 | 124,1 | 124,1 | 24,1 | 24,1 | 7,18 | 7,18 |
| 2016 | 1240 | 349 | 522 | 139,2 | 172,7 | 39,2 | 72,7 | 8,91 | 7,18 |
| 2017 | 1621 | 381 | 903 | 130,7 | 225,8 | 30,7 | 125,8 | 12,4 | 7,18 |
| 2018 | 1862 | 241 | 1144 | 114,9 | 259,3 | 14,9 | 159,3 | 16,21 | 7,18 |

Выпуск продукции на данном предприятии в течение пяти лет возрастал. Максимальный прирост выпуска по сравнению с предшествующим годом наблюдался в 2017 году и был зафиксирован на уровне 381 т., что на 30,7 % выше показателя 2016 года. Однако в 2018 году наблюдалось заметное снижение темпов роста выпуска продукции. Так по сравнению с 2017 годом прирост выпуска составил лишь 241 т. (14,9 %).

По сравнению с базисным 2014 годом выпуск продукции увеличивался стабильными темпами.

Для анализа рядов динамики кроме абсолютного прироста, темпов роста и темпов прироста исчисляются *средний уровень ряда, средний абсолютный прирост, средние темп роста и темп прироста*, которые дают обобщающую характеристику динамики явления или процесса на протяжении длительного временного интервала.

Расчет среднего уровня ряда динамики

Выбор формулы для расчета среднего уровня динамического ряда зависит от его разновидности.

1. Расчет среднего уровня в интервальном динамическом ряду:

а) в полном интервальном ряду средний уровень ряда рассчитывается по формуле средней арифметической простой:

$$\bar{y} = \frac{\sum y_i}{n},$$

где y_i – значение показателя в i -м интервале времени;
 n – количество интервалов.

б) в неполном интервальном ряду средний уровень ряда рассчитывается по формуле средней арифметической взвешенной, в которой в качестве весов используют длину временных периодов, соответствующих уровням ряда динамики:

$$\bar{y} = \frac{\sum y_i t_i}{\sum t_i}$$

где t_i – длина временного периода, соответствующего уровню ряда;

Пример:

Известен среднемесячный тоннаж отгрузки овощей. Необходимо рассчитать, сколько в среднем за месяц отгружается овощей в течение года.

| Месяц | январь | февраль | март | II квартал | III квартал | IV квартал |
|---|--------|---------|------|------------|-------------|------------|
| Среднемесячный объем отгрузки фруктов, тонн | 12,3 | 12,1 | 11,5 | 12,8 | 13,3 | 13,8 |

Данный динамический ряд является интервальным, неполным, поэтому расчет среднего уровня ведется по формуле арифметической взвешенной:

$$\bar{y} = \frac{12,3 \cdot 1 + 12,1 \cdot 1 + 11,5 \cdot 1 + 12,8 \cdot 3 + 13,3 \cdot 3 + 13,8 \cdot 3}{12} = 12,97 \text{ тонн}$$

2. Расчет среднего уровня ряда в моментном ряду

а) в полном моментном динамическом ряду (с равными промежутками времени между датами) средний уровень ряда рассчитывается по формуле средней хронологической простой:

$$\bar{y} = \frac{\frac{1}{2}y_1 + y_2 + \dots + \frac{1}{2}y_n}{n-1},$$

где n – количество моментов времени, на которые зафиксированы значения показателя.

б) в неполном моментном ряду (с неравными промежутками времени между датами) средний уровень ряда рассчитывается по средней хронологической взвешенной:

$$\bar{y} = \frac{(y_1 + y_2) \cdot t_1 + (y_2 + y_3) \cdot t_2 + \dots + (y_{n-1} + y_n) \cdot t_{n-1}}{2 \sum t_i},$$

где t_i – длина временного периода между двумя соседними датами.

Пример:

По данным об остатках материалов на начало месяца, определим средний размер остатков в 2015 году.

| Дата | 01.01. 2017 | 01.02. 2017 | 01.03. 2017 | 01.07. 2017 | 01.09. 2017 | 01.12. 2017 | 01.01. 2018 |
|-------------------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Остаток материалов, тыс. руб. | 135 | 142 | 138 | 136 | 145 | 148 | 154 |

Данный ряд является моментным, неполным, поэтому средний уровень рассчитывается по формуле средней хронологической взвешенной:

$$\bar{y} = \frac{(135 + 142) \cdot 1 + (142 + 138) \cdot 1 + (138 + 136) \cdot 4 + (136 + 145) \cdot 2 + (145 + 148) \cdot 3 + (148 + 154) \cdot 1}{2 \cdot 12} = 141,5$$

тыс. руб.

В среднем в течение года остаток материалов на складе был в сумме 141,5 тыс. руб.

Расчет среднегодовых показателей динамики

1. Среднегодовой абсолютный прирост – обобщающий показатель скорости абсолютного изменения уровней динамического ряда:

$$\bar{\Delta} = \frac{y_n - y_0}{n - 1},$$

где y_n – конечный уровень ряда;
 n – число уровней ряда.

2. Среднегодовой темп роста – обобщающая характеристика индивидуальных темпов роста ряда динамики:

$$\bar{T} = \sqrt[n-1]{\frac{y_n}{y_0}} \cdot 100$$

где n – число уровней ряда;

или

$$\bar{T}_p = \sqrt[n]{\prod T_{цп}},$$

где $\prod T_{цп}$ – произведение цепных темпов роста;
 n – число цепных темпов роста.

3. Средний темп прироста (среднегодовой) определяется на основе взаимосвязи между темпами роста и прироста. Если средний темп роста выражен коэффициентом, то от него отнимают 1, если средний темп роста выражен в процентах, то от него отнимают 100.

$$\bar{T}_{\Pi} = \bar{T}_p - 100 \%$$

4. Средняя величина 1 % прироста

$$\bar{A} = \frac{\bar{\Delta}}{\bar{T}_{\Pi P}}$$

Для примера, представленного на стр. 27, рассчитаем средний объем выпуска, средний абсолютный прирост и средние темп роста и прироста.

Так как представленный ряд динамики является полным интервальным рядом, то средний уровень (объем выпуска) рассчитывается по средней арифметической простой:

$$\bar{y} = \frac{\sum y_i}{n} = \frac{718 + 891 + 1240 + 1621 + 1862}{5} = 1266,4 \text{ т.}$$

Среднегодовой прирост:

$$\bar{\Delta} = \frac{y_n - y_6}{n - 1} = \frac{1862 - 718}{4} = 286 \text{ т.}$$

Среднегодовой темп роста:

$$\bar{T} = \sqrt[n-1]{\frac{y_n}{y_6}} \cdot 100 = \sqrt[4]{\frac{1862}{718}} \cdot 100 = 126,9 \%$$

Среднегодовой темп прироста:

$$\bar{T}_{\Pi} = \bar{T}_p - 100\% = 126,9 - 100 = 26,9 \%$$

Таким образом, выпуск продукции в среднем за каждый год возрастал на 286 т., что составляет 26,9 %.

Тема 5: Корреляционно-регрессионный анализ

Общественная жизнь состоит из большого количества сложных явлений, которые формируются под влиянием многочисленных, разнообразных и взаимосвязанных факторов.

Понять и изучить какие-либо явление можно, исследуя его во взаимосвязи с окружающими признаками.

В статистике различают *факторные* и *результативные признаки*.

Факторные признаки (x) – это признаки, обуславливающие изменение других, связанных с ними признаков.

Результативные признаки (y) – это признаки, изменяющиеся под действие факторных признаков.

Между явлениями и их признаками различают два вида связей: *функциональные* и *корреляционные*.

Функциональной называется такая связь, при которой определенному значению факторного признака соответствует одно значение результативного признака.

Такие связи являются **полными** и часто встречаются в физике, химии, математике. Функциональная зависимость с одинаковой силой проявляется у всех единиц совокупности, независимо от изменения других признаков данного явления.

Корреляционные связи обнаруживаются не в единичных случаях, а в массе и требуют для своего исследования массовых наблюдений. Лишь в достаточно большом числе фактов индивидуальные особенности и второстепенные факторы сглаждаются, и зависимость проявится достаточно отчетливо. Особенность таких связей состоит в том, что они являются **неполными**.

В зависимости от *направления действия* связи делятся на *прямые* и *обратные*.

Прямая связь – с изменением значений факторного признака, изменение результативного признака происходит в том же направлении.

Обратная связь – с изменением значений факторного признака, значения результативного признака изменяются в противоположном направлении.

По *аналитическому* выражению связи делятся на *прямолинейные* и *криволинейные*.

При прямолинейной связи – с возрастанием величины факторного признака происходит равномерное возрастание или убывание величин результативного признака. Эти связи выражаются уравнением прямой линии.

При криволинейной связи – с возрастанием величины факторного признака возрастание или убывание результативного признака происходит неравномерно или направление его изменения меняется на обратное. Эти связи выражаются уравнениями кривых линий: гиперболы, параболы и др.

В зависимости от *количества признаков*, включенных в модель, связи делят на *парные* и *множественные*.

Парные (однофакторные) связи - связь между одним признаком-фактором и результативным признаком (без учета влияния других).

Множественные (многофакторные) связи – связь между несколькими факторными признаками и результативным признаком (факторы действуют одновременно и взаимосвязано).

В статистике применяются различные методы изучения взаимосвязей.

Методы изучения взаимосвязей

1. Метод параллельных рядов.

После того, как будет установлено, что между изучаемыми явлениями возможна связь, необходимо установить ее наличие. Для этого факторы, характеризующие результативный признак располагают в возрастающем порядке, параллельно записывая значения результативных признаков. Путем сопоставления расположенных таким образом рядов значений выявляют наличие связи и ее направление.

Например, мы знаем, что количество пропущенных занятий по статистике влияет на оценку на экзамене.

Пример:

| № студента | Балл на экзамене, <i>y</i> | Количество пропущенных занятий, <i>x</i> | Приведенные данные | |
|------------|----------------------------|--|--------------------|----------|
| | | | <i>x</i> | <i>y</i> |
| 1 | 5 | 1 | 1 | 5 |
| 2 | 3 | 8 | 2 | 5 |
| 3 | 4 | 3 | 2 | 5 |
| 4 | 4 | 5 | 3 | 4 |

| | | | | |
|----|---|----|----|---|
| 5 | 3 | 8 | 4 | 4 |
| 6 | 2 | 10 | 5 | 4 |
| 7 | 5 | 2 | 6 | 3 |
| 8 | 4 | 4 | 8 | 3 |
| 9 | 5 | 2 | 8 | 3 |
| 10 | 3 | 6 | 10 | 2 |

Мы видим, что с увеличением пропущенных занятий успеваемость студентов имеет тенденцию к снижению. То есть связь между этими явлениями обратная.

2. Графический метод выявления корреляционных зависимостей.

Данный метод дает наглядное представление о существующей связи между явлениями.

Так выявленную в рассмотренном примере связь можно изобразить графически, если по оси абсцисс отложить значения признака x , а по оси ординат – соответствующие значения признака y .

Нанеся на график точки, мы получили корреляционное поле, где по характеру расположения точек можно судить о силе и направлении связи.

Так, если точки беспорядочно разбросаны по всему полю, то связи между явлениями нет. Если точки концентрируются вокруг оси, идущей от начала координат к верхнему правому углу, то имеется прямая зависимость между признаками. Если точки концентрируются вокруг оси, идущей от левого верхнего угла в правый нижний, то связь обратная.

3. Метод аналитических группировок.

При данном методе производится сравнение не индивидуальных данных, а групповых средних.

С этим методом мы познакомились в теме «Статистическая группировка».

4. Однофакторный корреляционно-регрессионный анализ

Данный анализ заключается в построении и анализе статистической модели в виде уравнения регрессии (уравнения корреляционной связи), приближенно выражающей зависимость результативного признака от одного или нескольких признаков-факторов.

5. Многофакторный корреляционно-регрессионный анализ

Однофакторный корреляционно-регрессионный анализ

Для построения уравнения регрессии необходимо правильно выбрать тип функции.

Если на корреляционном поле соединить точки отрезками, то получится ломаная линия, которая называется эмпирическая ломаная линия регрессии. Изломы этой линии говорят о том, что помимо признака x на признак y оказывают влияние другие факторы.

Чтобы абстрагироваться от данного влияния, необходимо прибегнуть к выравниванию эмпирической линии регрессии.

Если присмотреться к расположению точек на корреляционном поле, то можно увидеть, что наилучшим образом связь отображает прямая линия. Поэтому для выравнивания применяется уравнение прямой, то есть в нашем примере связь линейная.

Уравнение прямой линии: $y_x = a_0 + a_1x$

Данная линия связи называется линией регрессии, а ее поиск, построение, анализ и практическое применение - регрессионным анализом.

Параметры уравнения регрессии имеют большое практическое значение, так как применяются в плановых и экономических расчетах.

Для нахождения теоретического уравнения связи необходимо определить параметры прямой с помощью способа наименьших квадратов, который дает систему нормальных уравнений:

$$\begin{aligned} a_0n + a_1\sum x &= \sum y; \\ a_0\sum x + a_1\sum x^2 &= \sum xy \end{aligned}$$

где n – численность совокупности.

Построим корреляционную таблицу:

| № п/п | x | y | x^2 | y^2 | xy | y_x |
|-------|-----|-----|-------|-------|------|-------|
| 1 | 1 | 5 | 1 | 25 | 5 | 5,079 |
| 2 | 2 | 5 | 4 | 25 | 10 | 4,751 |
| 3 | 2 | 5 | 4 | 25 | 10 | 4,751 |
| 4 | 3 | 4 | 9 | 16 | 12 | 4,423 |

| | | | | | | |
|-------|----|----|-----|-----|-----|--------|
| 5 | 4 | 4 | 16 | 16 | 16 | 4,095 |
| 6 | 5 | 4 | 25 | 16 | 20 | 3,767 |
| 7 | 6 | 3 | 36 | 9 | 18 | 3,439 |
| 8 | 8 | 3 | 64 | 9 | 24 | 2,783 |
| 9 | 8 | 3 | 64 | 9 | 24 | 2,783 |
| 10 | 10 | 2 | 100 | 4 | 20 | 2,127 |
| Итого | 49 | 38 | 323 | 154 | 159 | 37,998 |

В систему уравнений подставим найденные значения:

$$10 a_0 + 49 a_1 = 38;$$

$$49 a_0 + 323 a_1 = 159$$

Решая полученную систему определяем параметры a_0 и a_1 :

$$a_1 = -0,328$$

$$a_0 = 5,407$$

Следовательно, теоретическое уравнение связи приобретает вид:

$$y_x = 5,407 - 0,328 x$$

Параметр a_1 в уравнении регрессии называется **коэффициентом регрессии** и имеет большое практическое значение. Он характеризует, в какой мере изменяется y_x с изменением величины x . В нашем примере с увеличением количества пропусков на одно занятие, балл на экзамене снижается на 0,328 единиц.

Для определения тесноты связи, то есть степени ее приближения к функциональной, определяется **линейный коэффициент корреляции**.

$$r = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n} \right] \cdot \left[\sum y^2 - \frac{(\sum y)^2}{n} \right]}}$$

$$r = \frac{159 - \frac{49 \cdot 38}{10}}{\sqrt{\left[323 - \frac{(49)^2}{10} \right] \cdot \left[154 - \frac{(38)^2}{10} \right]}} = \frac{-27,2}{\sqrt{82,9 \cdot 9,6}} = -0,96$$

Значение коэффициента корреляции изменяется от -1 до +1. Близость значения коэффициента к единице характеризует близость к функциональной зависимости.

В нашем примере связь между количеством пропущенных занятий и баллом, полученным на экзамене очень тесная и обратная, то есть увеличение количества пропущенных занятий ведет к снижению балла на экзамене.

Чтобы оценить, сколько процентов изменения результативного признака обусловлено изменением факторного признака, рассчитывают коэффициент детерминации:

$$d^2 = r^2$$

$$d^2 = 0,92$$

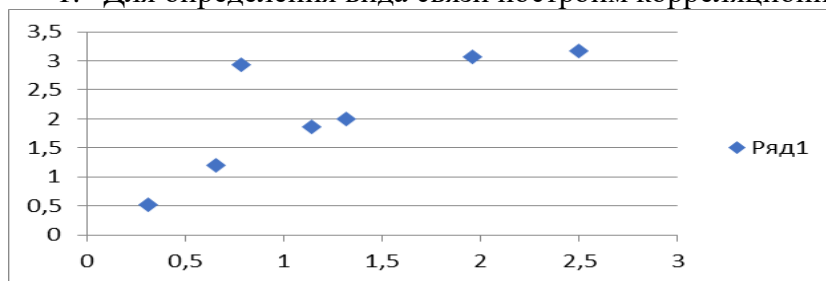
Это говорит о том, что количество баллов, получаемых на экзамене, на 92 % зависит от количества посещенных занятий. Остальные 8 % влияния приходится на другие факторы.

Пример 1:

По данным о сумме активов и кредитных вложений коммерческих банков одного из регионов РФ на 01.01.2012 г. необходимо определить направление и тесноту связи между признаками.

| № банка | Кредитные вложения, млн. руб., x | Сумма активов, млн. руб., y |
|---------|----------------------------------|-----------------------------|
| 1 | 311 | 518 |
| 2 | 658 | 1194 |
| 3 | 783 | 2941 |
| 4 | 1142 | 1865 |
| 5 | 1319 | 1997 |
| 6 | 1962 | 3066 |
| 7 | 2496 | 3176 |
| итого | 8671 | 14757 |

1. Для определения вида связи построим корреляционное поле.



На графике видно, что точки концентрируются вдоль прямой линии, идущей от левого нижнего угла к правому верхнему. Значит, существует прямая линейная связь между признаками.

2. Построим уравнение регрессии:

Так как связь между явлениями линейная, то применяем уравнение прямой: $y_x = a_0 + a_1x$

Для решения уравнения составим систем нормальных уравнений:

$$\begin{aligned} a_0n + a_1\sum x &= \sum y; \\ a_0\sum x + a_1\sum x^2 &= \sum xy \end{aligned}$$

где $n = 14$

Построим корреляционную таблицу (млрд. руб.):

| № п/п | x | y | x^2 | y^2 | xy | y_x |
|-------|-------|--------|--------|--------|--------|--------|
| 1 | 0,311 | 0,518 | 0,097 | 0,268 | 0,161 | 1,140 |
| 2 | 0,658 | 1,194 | 0,433 | 1,426 | 0,786 | 1,502 |
| 3 | 0,783 | 2,941 | 0,613 | 8,649 | 2,303 | 1,633 |
| 4 | 1,142 | 1,865 | 1,304 | 3,478 | 2,130 | 2,007 |
| 5 | 1,319 | 1,997 | 1,740 | 3,988 | 2,634 | 2,192 |
| 6 | 1,962 | 3,066 | 3,849 | 9,400 | 6,015 | 2,862 |
| 7 | 2,496 | 3,176 | 6,230 | 10,087 | 7,927 | 3,419 |
| итого | 8,671 | 14,757 | 14,266 | 37,297 | 21,956 | 14,755 |

$$\begin{aligned} 7 a_0 + 8,671 a_1 &= 14,757; \\ 8,671 a_0 + 14,266 a_1 &= 21,956 \end{aligned}$$

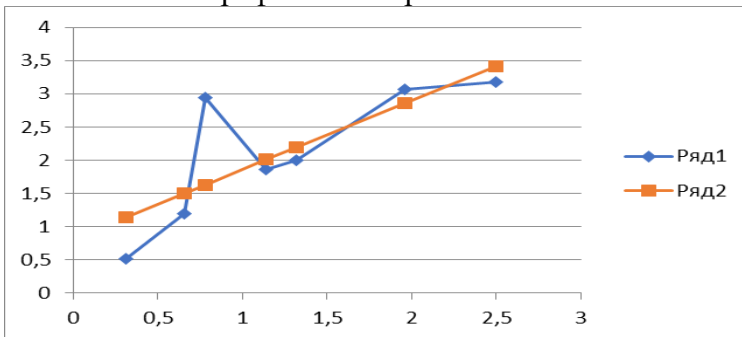
Решая полученную систему определяем параметры a_0 и a_1 :

$$\begin{aligned} a_1 &= 1,0429 \\ a_0 &= 0,816 \end{aligned}$$

Следовательно, теоретическое уравнение связи приобретает вид:

$$y_x = 0,816 + 1,0429 x$$

Строим линию на графике по выровненным значениям.



Коэффициент регрессии говорит о том, что при увеличении кредитов на 1 млн. руб. сумма активов возрастет на 1,043 тыс. руб.

3. Рассчитаем линейный коэффициент корреляции:

$$r = \frac{\Sigma xy - \frac{\Sigma x \cdot \Sigma y}{n}}{\sqrt{\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right] \cdot \left[\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right]}}$$

$$r = \frac{21956214 - \frac{8671 \cdot 14757}{7}}{\sqrt{\left[14266159 - \frac{(8671)^2}{7}\right] \cdot \left[37297007 - \frac{(14757)^2}{7}\right]}}$$

$$= \frac{3676507}{4670260} = 0,787$$

Расчеты показали, что между кредитными вложениями и суммой активов разрыву существует прямая и тесная связь.

4. Рассчитаем коэффициент детерминации

$$d^2 = r^2 = 0,62$$

Это говорит о том, что увеличение суммы активов коммерческих банков обусловлено увеличением суммы кредитных вложений всего на 62 %, а остальные 38 % составляет влияние других факторов.

СПИСОК ЛИТЕРАТУРЫ

1. Замедлина Е. А. Статистика : учебное пособие / Е.А. Замедлина - Москва : РИОР : ИНФРА-М, 2019. - 160 с. - (Среднее профессиональное образование). - Текст : электронный. - URL: <https://znanium.com/catalog/product/945157> (дата обращения: 20.08.2022). – Режим доступа: по подписке.

2. Сергеева И. И. Статистика : учебник / И.И. Сергеева, Т.А. Чекулина, С.А. Тимофеева. — 2-е изд., испр. и доп. — Москва : ФОРУМ : ИНФРА-М, 2021. — 304 с. — (Среднее профессиональное образование). - Текст : электронный. - URL: <https://znanium.com/catalog/product/1141798> (дата обращения: 20.08.2022). – Режим доступа: по подписке.

3. Шумак О. А. Статистика: Учебное пособие / О.А. Шумак, А.В. Гераськин. – Москва : ИЦ РИОР: НИЦ Инфра-М, 2019. - 311 с.: ил.; - (Высшее образование: Бакалавриат). - Текст : электронный. - URL: <https://znanium.com/catalog/product/1002740> (дата обращения: 20.08.2022). – Режим доступа: по подписке.

Интернет ресурсы:

<http://www.gks.ru> Федеральная служба государственной статистики

<http://www.amurstat.gks.ru> Федеральная служба гос. статистики по Амурской области

<http://www.amurobl.ru> Правительство Амурской области

<http://www.minfin.ru> Министерство финансов РФ

<http://www.economy.gov.ru> Минэкономразвития РФ

<http://www.minregion.ru> Министерство регионального развития

<http://www.cbr.ru> Центральный банк РФ

<http://www.garant.ru> Справочно-правовая система Гарант

<http://www.consultant.ru> Справочно-правовая система Консультант