1. ВВЕДЕНИЕ В СТАТИСТИКУ. ПОНЯТИЕ СТАТИСТИКИ

1.1. Предмет и метод статистической науки

Статистика – отрасль знаний, изучающую количественную сторону явлений жизни общества в неразрывной связи с их качественным содержанием.

Предмет ее исследований – массовые явления социально-экономической жизни.

Объект изучения – это количественная сторона этих явлений, отражающая их качественное содержание в конкретных условиях места и времени.

Статистика создает фундамент из точных и бесспорных фактов для анализа общественных явлений в конкретных исторических условиях. Экономическая теория, опираясь на статистику, формулирует законы развития социально-экономических явлений.

Явления и процессы в жизни общества изучаются статистикой посредством статистических показателей.

Статистический показатель – это количественная оценка свойства изучаемого явления.

Учетно-оценочные показатели – это количественные характеристики изучаемых явлений на определенный момент времени (объемы, уровни).

Аналитические показатели – это показатели (относительные, средние величины, показатели вариации и динамики и др.), используемые для характеристики развития изучаемого явления (соотношение его отдельных частей, скорость развития во времени и т.д.)

Признак – характерное свойство изучаемого явления, отличающего его от других явлений, и выражается смысловыми понятиями и числовыми значениями.

Статистическая совокупность – это множество единиц изучаемого явления, объединенных в соответствии с задачей исследования.

Цель наблюдения – основной результат статистических исследований.

Объект статистического наблюдения – совокупность единиц изучаемого явления, о которых должны быть собраны статистические данные.

Единица наблюдения – это первичная ячейка, от которой должны быть получены необходимые статистические сведения.

Программа статистического наблюдения – перечень показателей, подлежащих изучению, регистрации по каждой единице наблюдения.

Статистическая методология – совокупность правил, приемов и методов статистического исследования. Статистическое исследование состоит из трех стадий:

1. Сбор первичной статистической информации (применяется **метод статистического наблюдения**).
2. Статистическая сводка и обработка первичной информации (собранная информация обрабатывается **методом статистических группировок** для выделения в изучаемой совокупности социально-экономических **типов**).

3. Анализ статистической информации. Обобщение и интерпретация (проводится **анализ статистической информации** на основе применения обобщающих статистических показателей: средних величин, вариации, скорости изменения явлений во времени, индексов и др).

Статистическое наблюдение – сбор первичных данных о изучаемом явлении. Осуществляется с помощью оценки и регистрации значений признаков единицы изучаемой совокупности в учетных документах (формы отчетности предприятий, записи счетчиков в переписных листах ответов, и др.).

Статистическая информация (статистические данные) – первичный статистический материал, формирующийся в процессе статистического наблюдения, который подвергается систематизации, сводке, обработке, анализу и обобщению.

1.2. Статистическая сводка

Статистическая сводка – обработка и систематизация единичных фактов, позволяющая перейти к упорядоченной системе обобщающих статистических показателей всей совокупности и ее частей и осуществить анализ и прогнозирование изучаемых явлений и процессов.

Статистическая сводка осуществляется по составленной **программе** с этапами**:**

- выбор группировочных признаков;
- определение порядка формирования групп;
- разработка системы статистических показателей для характеристики групп и объекта в целом;
- разработка макетов таблиц занесения данных, характеристик объекта и его отдельных частей, для представления результатов сводки различными показателями;
- выбор способов сводки данных статистического наблюдения.

Макет – незаполненная таблица. **Подлежащее** таблицы показывает, о каком явлении идет речь в таблице и представляет собой группы и подгруппы. **Сказуемое** таблицы – показатели, с помощью которых изучается объект, т.е. численные значения характеристик явления.

1.3. Статистические группировки

Статистическая группировка – это процесс образования качественно однородных групп на основе расчленения статистической совокупности на части или объединение изучаемых единиц в частные совокупности по существенным признакам.

Признаки, по которым производится деление совокупности на группы, называются **группировочными** или **основанием** группировки.

1. Типологическая группировка – выявление и характеристика классов, типов, однородных групп единиц совокупности. Группировка осуществляется по **атрибутивному признаку** (по **количественному**), если его отдельным значениям, или группам значений, (его отдельным интервалам) соответствуют качественно однородные группы изучаемого явления.

2. Структурная группировка – типологическая группировка, с последующим вычислением удельных весов групп в совокупности по некоторому другому признаку (пропорции хозяйств по объему продукции).

3. Аналитические (факторные) группировки – исследуют взаимосвязи между явлениями (у–зарплата, х – квалификация, группировочный признак).

Комбинационные группировки – образование групп по двум и более признакам, взятым в определенном сочетании.

Пример 1. Распределение промышленной продукции, произведенной в различных формах хозяйствования за отчетный период.

Группы предприятий по формам хозяйствования	Объем промышленной продукции, млн. руб.	В % к итогу
Государственные, с традиционными формами управления	405,0	89,2
Арендные Кооперативные	19,0 30,0	4,19 6,61
Всего:	454,0	100,0

Пример 2. Группировка торговых предприятий района по объему товарооборота.

Группы магазинов по Объему товарооборота, тыс. руб.	Число магазинов	Розничный Товарооборот	Торговая Площадь
До 1700	21,87	11,22	18,05
1700-2000	28,13	19,04	21,38

1.4. Принципы построения статистических группировок

Построение группировок начинают с выбора группировочного признака. Признаки различаются:

- по форме (атрибутивные и количественные);
- по степени и характеру колеблемости:
- по взаимосвязи (факторные признаки (х) воздействуют на другие, результативные (у) – зависят от других).

Пример 3. Прибыль – результативный признак при оценке предприятия; – факторный при планировании развития предприятия.

Количество групп для количественного признака задают по формуле Стерджесса $n = 1+3{,}322 \lg N$, где **n** –число групп, N – численность совокупности.

Величину интервала группировки с равными интервалами вычисляют по формуле

$i = \dfrac{X_{\max} - X_{\min}}{n}$, где **n** –число групп, $R = X_{max} - X_{min}$ – размах вариации, X_{max} – наибольшее, а X_{min} – наименьшее значение признака.

В зависимости от степени колеблемости группировочного признака, характера распределения статистической совокупности устанавливают интервалы **равные** или **неравные**.

Пример 4. Произведем группировку с выделением пяти групп продавцов. Наибольшая производительность труда составила 180 тыс. руб., а наименьшая – 80 тыс. руб. Определим размах вариации $R = X_{max} - X_{min} = 180 - 80 = 100$ тыс. руб. Определим величину интервала – 100 : 5 = 20 тыс. руб. Получим интервалы групп:

80-100; 100-120; 120-140; 140-160; 160-180.

Интервалы бывают **открытые** (указана одна граница) и **закрытые**. Зная значение интервала группировки открытые интервалы можно дополнять до закрытых.

1.5. Статистические ряды распределения

После определения группировочного признака и границ групп строится ряд распределения.

Статистический ряд распределения – это упорядоченное распределение единиц совокупности на группы по группировочному признаку. Ряды распределения характеризуют структуру изучаемого явления, закономерности развития и т.д.

Ряды распределения образованные по атрибутивным признакам называются **атрибутивными** (распределение населения по полу), а по количественному – **вариационными** (распределение населения по возрасту).

Вариационные ряды, в зависимости от группировки по дискретному или непрерывному признаку, бывают **дискретными** или **интервальными.** Вариационные ряды состоят из 2-х элементов: **варианты и частоты**.

Варианта – числовое значение количественного признака в ряду распределения.

Частота – численности отдельных вариант или групп вариационного ряда (f_i). Сумма частот составляет **объем ряда** распределения $\sum f_i = n$. Частоты, выраженные в долях единицы или в процентах к объему ряда, называются **частостями**.

Полигоном графически изображаются дискретные ряды ($x=x_i$, $y(x) = f_i$). Интервальные ряды распределения графически отображаются **гистограммами**. **Гистограмма** – это график в виде вертикальных прямоугольников, высота которых соответствует их частотам, а ширина – величине интервала.

Кумулятивные ряды распределения строятся по накопленным частотам и показывают, сколько единиц совокупности, или какая их доля не превышает данное значение ($x=x_i$, $y(x) = \sum_{j: x_j \le x} f_j$). Они позволяют определить структурные средние, проследить за процессом концентрации изучаемого явления.

Пример 5. Распределение магазинов района по числу товарных секций (дискретный ряд распределения).

Число	**на 1. 01. 99 г.**	**на 1. 01. 2000 г.**

Товарных секций	Число магазинов	%	Число Магазинов	%
1	3	6	6	10
2	10	20	16	27
3	15	30	20	33
4	12	24	12	20
5	7	14	4	7
6	3	6	2	3
Итого	50	100	60	100

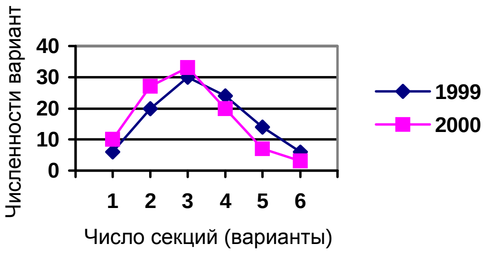


Рис. 1. Полигон распределения магазинов по числу товарных секций.

Пример 6. Распределение продавцов магазина по выработке (интервальный ряд распределения).

Выработка продавца, тыс. руб.	Число продавцов	% к итогу (частость)	Кумулятивная (накопительная) численность Продавцов
80-100	5	10	5
100-120	10	20	15 (5+10)
120-140	20	40	35 (15+20)
140-160	10	20	45 (35+10)
160-180	5	10	50 (45+5)
Итого	50	100	

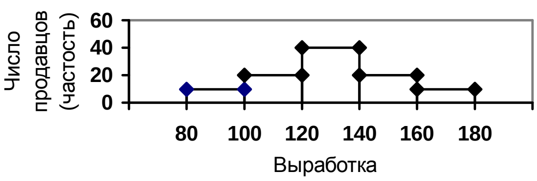
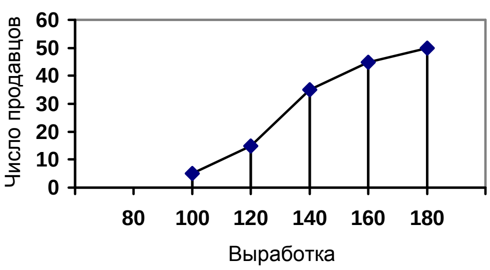


Рис.2. Гистограмма распределения продавцов по выработке.

Частота 3-ей группы показывает число продавцов или их долю с размером выработки 120-140 тыс. руб. (35 продавцов).

Для сравнения исчисляют плотность распределения – число единиц группы, приходящихся на единицу интервала, т.е. p= n/i, n – число магазинов; i – величина интервала.

Рис. 3. Кумулянта распределения 50 продавцов по выработке.



Задача 1. Имеются данные о размеры кредитов, предоставленных коммерческими банками пред-

приятиям и организациям.

№ банка	Кредит, млн. руб.				
1	2	1	2	1	2
1	9,55	11	7,60	21	17,9
2	13,58	12	25,52	22	12,30
3	22,33	13	2,50	23	5,40
4	27,50	14	13,24	24	12,18
5	13,54	15	20,15	25	17,10
6	11,60	16	6,10	26	1,00
7	8,90	17	13,36	27	12,12
8	3,25	18	19,62	28	16,45
9	21,20	19	11,90	29	26,50
10	13,50	20	5,20	30	23,98

Построить интервальный ряд, характеризующий распределение банков по сумме выданных кредитов, образовав 5 групп с равными интервалами. Определить какой процент банков выдал кредиты на сумму свыше 11,6 руб.

Решение. Определим величину интервала

$$h = \frac{X_{max} - X_{min}}{n} = \frac{27{,}5 - 1}{5} = 5{,}3 \text{ млн. руб.}$$

Построим интервальные группы распределения банков по размеру выданных кредитов.

№ Группы	**Группы банков по размеру кредита**	**Число банков**	
		в абсолютном выражении	**в относительных единицах, %**
I	1-6,3	6	20,0
II	6,3-11,6	3	10,0
III	11,6-16,9	11	36,6
IV	16,9-22,2	5	16,7
V	22,2-27,5	5	16,7
Итого:		30	100

Данные группировки показывают, что 70% банков выдали кредиты на сумму свыше 11,6 млн. руб.

2. АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ

2.1. Абсолютные величины

Результаты статистического наблюдения регистрируются в форме первичных **абсолютных величин**, именованных и измеренных в конкретных единицах. Абсолютные величины выражают количественную сторону той или иной сущности явления или свойства явления.

Виды абсолютных величин: индивидуальные, групповые и общие. **Индивидуальные** – выражают размеры количественных признаков и отдельных единиц (число вкладов в сберегательной кассе, количество тракторов в колхозе, производственный стаж сотрудника). **Групповые** и **общие** абсолютные статистические величины выражают суммарную величину того или иного признака всех единиц совокупности или единиц отдельных групп.

Единицы измерения абсолютных величин: натуральные, стоимостные и трудовые единицы измерения. Сумма объемов единиц совокупности – **объем совокупности**.

Натуральные единицы измерения – выражают величину предметов, вещей и т.п. в физических мерах: веса, объема, длины и т.д. в соответствии с их физическими свойствами ($м^2$, $м^3$, кВт·час).

Условные натуральные единицы измерения используют для сведения воедино нескольких разновидностей одной и той же потребительской стоимости (консервы в условных банках)

$$У = с + кх,$$

где: у – условные единицы измерения; с – количество эталонных единиц; х – количество единиц, отличающихся от эталонных; к – коэффициент пересчета.

Денежные единицы используют для характеристики в стоимостном выражении статистических показателей (объема продаж, товарооборота, доходов). Денежные единицы позволяют соизмерять, обобщать и сопоставлять разнородные величины.

Трудовые единицы (человеко-час, человеко-день и т.п.) – для измерения затрат труда на производство продукции, выполнения какой-либо работы, производительности труда, величины трудовых ресурсов.

2.2. Относительные величины

Относительными величинами называются величины, выражающие количественные соотношения между социально-экономическими явлениями или их признаками. Величина, с которой производится сравнение (знаменатель), называется **основанием относительной величины** или **базисной** (x_0), а та, которая сравнивается, называется **текущей**, **сравниваемой** или **отчетной величиной** (x_1)**.** С помощью относительных величин выражаются многие факты общественной жизни: процент выполнения плана, темп роста и прироста и др.

1. **Относительные величины динамики** – $i = \frac{x_1}{x_0} = \frac{текущий\ уровень}{базисный\ уровень}$. Характеризуют изменение уровня показателя по сравнению с предыдущим периодом.

2. **Относительная величина планового задания** - $i_{пл.з} = \frac{x_{1пл}}{x_0} = \frac{плановый\ уровень}{базисный\ уровень}$.

3. **Относительная величина выполнения плана** – $i_{вып.пл} = \frac{x_1}{x_{1пл}} = \frac{фактический}{плановый}$.

Справедлива взаимосвязь $i = i_{пл.з} \times i_{вып.пл}$

4. **Относительные величины структуры** – характеризуют долю отдельных частей в общем объеме совокупности. Рассчитываются по сгруппированным данным (категории: рабочие, служащие) $d_i = \frac{f_i}{\sum f_i} \times 100\%$, где: di – удельный вес частей.

4. **Относительные величины координации** – характеризуют соотношение частей целого между собой (отношение численности сельского населения к городскому). Одну из частей принимают за базу сравнения и находят соотношение с ней других частей или число единиц одной части, приходящееся на 1, 10, 100, 1000 единиц другой части.

5. **Относительные величины сравнения** – отражают результаты сопоставления одноименных показателей, относящихся к одному и тому же периоду времени, но к разным объектам или территориям. Применяют для сравнения уровня развития стран и регионов, а также при оценке результатов деятельности отдельных предприятий отрасли.

6. **Относительные величины интенсивности** – характеризуют степень насыщенности или развития явления – показатели жизненного уровня населения (потребление продуктов питания, непродовольственных товаров на душу населения); уровни экономического и социального развития; технической оснащенности труда (фондо-, машино-, и энерговооружения); уровня технического развития.

Пример 1. (Показатель интенсивности). Среднегодовая численность населения в 1996 г. составила 147,7 млн. человек, число родившихся – 1304,6 тыс. человек. Определить число родившихся на каждую 1000 человек населения.

Решение. Коэффициент рождаемости $= \frac{\text{Число родившихся}}{\text{Среднегод. числ. населения}} \times 1000 = = \frac{1304{,}6}{147700} \times 1000 = 8{,}9$ чел.

На каждую 1000 человек населения рождалось 8,9 человек.

Пример 2. (Показатель координации).

Данные об экономически активном населении на начало 1997г., млн. чел.

Экономически активное население, в т.ч.	- 72,7
занятые в экономике	- 65,9
безработные	- 6,8

Определить сколько безработных приходится на 1000 занятых.

Решение. На каждые 1000 занятых в экономике приходится $\text{ОПК} = \frac{6{,}8}{65{,}9} \times 1000 = 103{,}2$ безработных.

Задача 1. Мыловаренный завод должен был по плану выработать 750 т мыла разных сортов (в пересчете на мыло 40% жирности). Определить процент выполнения плана заводом по выработке мыла.

Сорт мыла	Выпуск в натуральных единицах, т	Коэффициент перевода, к	Выпуск в условных единицах
80%-ное	90	80 : 40 = 2	180
60%-ное	150	60 : 40 = 1,5	225
40%-ное	450	40 : 40 = 1	450
Стиральный порошок 10%-ный	60	10 : 40 = 0,25	15
Итого:			870

Решение. Условная единица – мыло 40%-ной жирности. Переведем другие сорта в условные единицы. Для этого процентную жирность каждого сорта мыла сопоставим с условной, то есть получим коэффициент перевода к. Рассчитаем выпуск продукции в условных единицах по формуле у = с + кх. Результаты вычислений в таблице.

Процент выполнения плана $ОПВП = \frac{870 \times 100}{750} = 116\%$.

Задача 2. Расход топлива на нужды предприятия характеризуется следующими данными

Вид топлива	Единица измерения	Расход		Коэффициент Перевода в условное топливо, калорийные эквиваленты
		по плану	факт	
Мазут топочный	т	500	520	1,37
Уголь	т	320	300	0,9
Газ природный	тыс. $м^3$	650	690	1,2

Определить:
1) общее потребление условного топлива по плану и фактически;
2) процент выполнения плана по общему расходу топлива;
3) удельный вес фактически израсходованного топлива по видам.

Решение. 1) Израсходовано:

Вид топлива	План, усл. ед.	Факт, усл. ед.	Удельный вес в общем объеме, %
Мазут топочный	$500 \times 1,37 =$ 685,0	$520 \times 1,37 =$ 712,4	39,4
Уголь	$320 \times 0,9 = 288,0$	$300 \times 0,9 = 270,0$	14,9
Газ природный	$650 \times 1,2 = 780,0$	$690 \times 1,2 = 828,0$	45,7
Итого:	1753,0	1810,4	100

2) Процент выполнения плана по общему расходу топлива

$$ОПВП = \frac{1810,4}{1753} \times 100 = 103,27\%.$$

Расход топлива превышает плановый на 3,27%.
3) Удельный вес фактически израсходованного топлива по видам (структура топлива)

$$d = \frac{\text{Уровень части совокупности}}{\text{Суммарный уровень совокупности}} = \frac{712}{1810} \times 100 = 39,4\%$$

Задача 3. По объединению имеются следующие данные о выпуске продукции за год Определить процент выполнения плана выпуска продукции в целом.

№ фирмы	Фактический выпуск Продукции, Млн. руб.	Процент выполнения плана	План выпуска, млн. руб.
1	29,4	105	28,0
2	42,6	100	42,6
3	24,0	96,0	24,0
Σ	96,0		95,6

Решение. Определим плановый выпуск для каждой фирмы и суммарный

$$у_{пл} = \frac{29,4 \times 100}{105} = 28,0 \text{ и т.д.}$$

Процент выполнения плана по фирме ОПВП $= \dfrac{96}{95{,}6} \times 100 = 100{,}4\%$

3. СРЕДНИЕ ВЕЛИЧИНЫ

Средняя величина – обобщающая характеристика изучаемого признака в качественно однородной совокупности. Она отражает его типичный уровень для единицы совокупности в конкретных условиях места и времени: средний доход; средняя выработка; и тд .

Средняя по статистической совокупности называется **общая средняя**. Средняя по группе – **групповая средняя**.

Признак x по которому находится средняя $\overline{x}$, называется **усредняемым признаком**. Величина усредняемого признака у каждой единицы совокупности (x_1, x_2, x_3, ….) называется **индивидуальным** его значением (**вариантой**). **Частота** (или вес f) – повторяемость (количество) индивидуальных значений.

3.1. Виды средних и методы их расчета

1. **Средняя арифметическая** ($\overline{x}$)– исчисляется в тех случаях, когда объем усредняемого признака образуется как сумма его значений у отдельных единиц изучаемой статистической совокупности x_1, x_2, x_3, .

В зависимости от характера исходных данных средняя арифметическая $\overline{x}$ определяется следующим образом:

Средняя арифметическая простая: $$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + \mathbf{K} + x_n}{n}.$$

Например, средний стаж работников предприятия.

Средняя арифметическая взвешенная

Среднее значение по ряду распределения заданного вариантами x_i и частотами f_i	$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$
Частоты отдельных вариантов могут быть выражены относительными величинами – частостями (w_i).	$\overline{x} = \frac{\sum x_i w_i}{\sum w_i}$

Например, средняя урожайность – взвешивание производиться по площади посевов, а не по количеству участков.

Средняя арифметическая взвешенная интервального ряда распределения вычисляется по правилу:

1) в каждом варианте определить срединное значение x', как полусумму значений нижней и верхней границ интервала x'=(x0+x1)/2, т.е. образуем дискретный ряд;
2) произвести взвешивание $x'f$ и вычислить среднее $\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$.

Мы предположили, что отдельные варианты равномерно распределены внутри интервала, что позволило нам образовать дискретный ряд с вариантами x'=(x0+x1)/2.

Свойства средней арифметической.

1. Если все веса (f) увеличить или уменьшить в одинаковое число раз К, то величина средней не изменится

$$\bar{x} = \frac{\sum \frac{f_i}{K} x_i}{\sum \frac{f_i}{K}} = \frac{\frac{1}{K} \sum f_i x_i}{\frac{1}{K} \sum f_i} = \frac{\sum f_i x_i}{\sum f_i} = \bar{x}.$$

2. Если каждую варианту (x_i) увеличить или уменьшить на одну и ту же величину А, то средняя увеличится или уменьшится на эту же величину

$$\frac{\sum f_i (x_i - A)}{\sum f_i} = \frac{\sum f_i x_i - A \sum f_i}{\sum f_i} = \bar{x} - A.$$

3. Если каждую варианту (x_i) увеличить или уменьшить в одно и то же число раз (h), то средняя увеличится или уменьшится в то же число раз

$$\frac{\sum f_i \frac{x_i}{h}}{\sum f_i} = \frac{\frac{1}{h} \sum f_i x_i}{\sum f_i} = \frac{\bar{x}}{h}.$$

4. Сумма отклонений вариант от средней, взвешенных их частотами, равна нулю

$$\sum f_i (x_i - \bar{x}) = 0.$$

Способ моментов (способ отсчета от условного нуля). С учетом этих свойств формула средней арифметической взвешенной будет иметь вид

$$\bar{x} = m_1 i + A, \text{ где } m_1 = \frac{\sum \left(\frac{x - A}{i} \right) f}{\sum f}.$$

Здесь m_1 – момент 1-го порядка; А – одна из центральных вариант ряда.

Пример 1. Дан интервальный ряд распределения предприятий торговли по объему товарооборота. Найти средний объем товарооборота.

Группы предпр. По Объему Товарооборот., x	**Число предпр. f**	**Серед. зн-ние интервала,** x'	$x'f$	$x' - A$	$\frac{x' - A}{i}$	$\frac{x' - A}{i} f$
до 400	9	350	3150	-200	-2	-18
400-500	12	450	5400	-100	-1	-12
500-600	8	550	4400	0	0	0
600-700	9	650	5850	100	1	9
свыше 700	2	750	4500	200	2	4
Итого $\sum$	40		20300	0	0	-17

Открытые интервалы дополнили до закрытых и рассчитали серединное значение для каждого интервала. Промежуточные расчеты приведены в таблице.

$$\bar{x} = \frac{20300}{40} = 507{,}5 \text{ млн. руб.};$$

Расчет средней арифметической способом моментов. А = 550.

$$\bar{x} = \frac{\sum\left(\frac{x-A}{i}\right)f}{\sum f} i + A = \frac{-17}{40} \times 100 + 550 = 507{,}5 \text{ млн. руб.}$$

Пример 2. Имеются следующие данные о продаже товара. Рассчитать среднюю цену реализации.

Город	**Цена,** x_i	**Сумма реализации** $w_i = f_i \times x_i$.	**Частота,** $f_i = \frac{w_i}{x_i}$
А	30	600	20
Б	20	1000	50
В	35	350	10
Итого	–	1950	80

Рассчитаем среднюю цену реализации по формуле средней гармонической взвешенной. Количество реализованных единиц – частота (f)

$$\bar{x} = \frac{\sum w_i}{\sum \frac{w_i}{x_i}} = \frac{600+1000+350}{\frac{600}{30}+\frac{1000}{20}+\frac{350}{35}} = 24{,}3 \text{ руб.}$$

2.**Средняя гармоническая** используется, когда статистическая информация не содержит частот (f_i) по отдельным вариантам (x_i) совокупности, а представлена как их произведение ($f_i \times x_i$), т.е. в виде объема явления $w_i = f_i \times x_i$. Тогда среднее значение можно вычислить как **среднее гармоническое взвешенное**

$$\bar{x}_{гар} = \frac{\sum w_i}{\sum \frac{1}{x_i} w_i} = \frac{\sum f_i x_i}{\sum \frac{f_i x_i}{x_i}} = \frac{\sum f_i x_i}{\sum f_i} = \bar{x}\,.$$

Объемы явления $w_i = f_i \times x_i$ можно выражать в долях и процентах, формула вычисления остается неизменной. При равных значениях объемов $w_1 = w_2 = \ldots w_n$ средняя вычисляется, как **средняя гармоническая простая** $\bar{x} = n / \sum \frac{1}{x}$.

Пример. $\text{Средняя цена яблок} = \frac{\text{Сумма реализации}}{\text{Количество реализов единиц}} = \frac{\text{Сумма реализации}}{\sum\limits_{\text{по магазинам}} \frac{\text{выручка}}{\text{цена}}}$.

3. **Средняя геометрическая** – используется как средняя относительных величин динамики, построенных в виде цепных величин (отношения текущего к предыдущему уровню) в рядах динамики (например, в расчетах среднегодовых темпов роста).

Средняя геометрическая простая $\bar{x} = \sqrt[n]{x_1 \times x_2 \times x_3 \times \mathbf{K} \times x_n}$ или $\bar{x} = \sqrt[n]{\Pi(x)}$, где П – символ произведений; n – число вариантов (коэффициент прироста национального дохода).

3.2. Структурные средние величины

Для характеристики структуры совокупности применяются структурные средние: **мода и медиана**.

Модой (Мо) называется наиболее часто встречающееся или типичное значение признака, т.е. то значение варианты, которое соответствует максимальной точке теоретической кривой распределения вариационного ряда. Мода часто используется при изучении покупательского спроса.

В дискретном ряду мода – это варианта с наибольшей частотой. **В интервальном ряду мода** – центральный вариант **модального интервала**, то есть того интервала, который имеет наибольшую частоту (частость). В пределах интервала находят значение признака, которое является модой по формуле

$$\mathrm{Mo} = x_{Mo} + i_{Mo}\frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})},$$

где: x_{Mo} – нижняя граница модального интервала; i_{Mo} – величина модального интервала; f_{Mo} – частота, соответствующая модальному интервалу; f_{Mo-1} – частота, предшествующая модальной; f_{Mo+1} – частота интервала, следующего за модальным.

Пример 3. Дан дискретный ряд распределения купленных пар обуви

Размер обуви	34	35	36	37	38	39	40
Число купленных пар	2	10	20	88 Мо	19	9	1

Определим значение моды. Варианта с наибольшей частотой равна 88. Этой частоте соответствует варианта, значение которой Мо = 37 размер.

Пример 4. Дан интервальный ряд распределения числа рабочих по стажу работы.

Стаж (лет)	**Число рабочих**
До 2	4
2-4	23
4-6	20
6-8	35
8-10	11
свыше 10	7

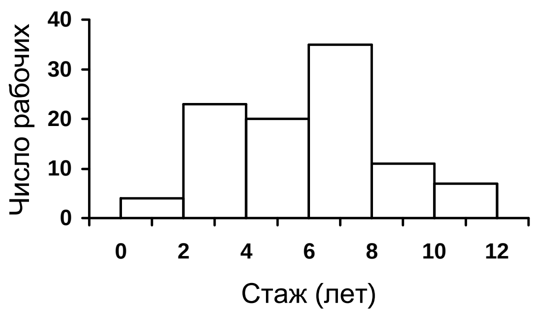


Рис. 4. Гистограмма распределения рабочих по стажу

Модальным интервалом является интервал 6-8 соответствующий наибольшей частоте равной 35. Значение моды находится внутри этого интервала.

$$\mathrm{Mo} = 6 + 2 \times \frac{35 - 20}{(35 - 20) + (35 - 11)} = 6{,}77 \text{ лет.}$$

Медиана (Ме) – это величина, которая делит численность упорядоченного вариационного ряда на две равные части: одна часть имеет значения варьирующего признака меньшие, чем средний вариант, а другая – большие.

$$\mathrm{Me} = \frac{x_{Me} + x_{Me+1}}{2} \text{ – для четного ряда.}$$

Для ранжированного ряда с нечетным числом членов медианой является варианта, расположенная в центре ряда.

Для ранжированного ряда с четным числом членов ряда медианой будет среднее арифметическое из двух смежных вариант.

В интервальном ряде $$Me = x_{Me} + i_{Me}\frac{\frac{\sum f}{2} - s_{Me-1}}{f_{Me}}$$

где: x_{Me} – нижняя граница медианного интервала; i_{Me} – величина медианного интервала; $\frac{\sum f}{2}$ – полусумма частот ряда; s_{Me-1} – сумма накопленных частот, предшествующих медианному интервалу; f_{Me} – число наблюдений (частота) медианного интервала.

Свойство медианы: сумма абсолютных отклонений членов ряда от медианы минимальна $\sum |x - Me| = \min$.

Пример 5. Имеются данные о расположении магазинов от базы. Вычислить отклонения от среднего и медианных значений.

№	Расположение магазина от базы, км (x)	Отклонение от среднего значения, $(x - \bar{x})$	Отклонение от медианного значения, $(x - Me)$
1 2 3 4 5	2 3 4 6 10	3 2 1 1 5	2 1 0 2 6
Итого	25	13	11

Среднее значение $\bar{x} = 25 : 5 = 5$ км. Медианой нечетного ряда является центральный вариант, находящийся на 3-ем месте, равный Me = 4 км.

Пример 6. Рассчитать среднюю заработную плату в целом по трем предприятиям в зависимости от имеющихся данных

Предприятие	Численность персонала, чел. f_i	Месячный ФЗП, тыс. руб., w_i	Средняя заработная плата, $\bar{x}_i$
А	1	2	3
1 2 3	540 275 458	564,84 332,75 517,54	1046 1210 1130
Итого	1273	1415,13	

Среднюю заработную плату рассчитаем разными способами в зависимости от того, какие данные нам будут известны.

1. Если имеются данные групп 1 и 2, формула средней агрегатной

$$\bar{x} = \frac{\sum w_i}{\sum f_i} = \frac{1415130}{1273} = 1112 \text{ руб.},$$

где: $w_i = \bar{x}_i \times f_i$; $\bar{x}_i$ – вариант определяемого признака; f_i – вес i-го варианта.

2. Если есть данные групп 1 и 3, то общая средняя может быть рассчитана по формуле средней арифметической взвешенной

$$\bar{x} = \frac{\sum \bar{x}_i f_i}{\sum f_i} = \frac{1046 \times 540 + 1210 \times 275 + 1130 \times 458}{540 + 275 + 458} = 1112 \text{ руб.}$$

3. Если имеются данные 3 группы, а частоты f_i равны между собой или отсутствуют, то используются формула средней арифметической простой

$$\bar{x} = \frac{\sum \bar{x}_i}{n} = \frac{1046+1210+1130}{3} = \frac{3386}{3} \approx 1128{,}6 \text{ руб.}$$

4. Если имеются данные групп 2 и 3, то расчет ведется по средней гармонической взвешенной

$$\bar{x} = \frac{w_i}{\sum \frac{w_i}{\bar{x}_i}} = \frac{564840+332750+517540}{\frac{564840}{1046}+\frac{332750}{1210}+\frac{517540}{1130}} = 1112 \text{ руб.}$$

Задача 1. Имеются следующие данные о заработной плате рабочих участка. Определить среднюю заработную плату рабочих участка.

Профессия	Количество рабочих	Заработная плата каждого рабочего участка за месяц, руб.
Токари	5	1700, 1208, 917, 1620, 1400
Фрезеровщики	2	1810, 1550
Слесари	3	1210, 1380, 870

Решение. Воспользуемся формулой простой средней арифметической

$$\bar{x} = \frac{\sum x}{n} = \frac{1700+1208+917+1620+1400+1810+1550+1210+1380+870}{10} = \frac{13665}{10} = 1366{,}5 \text{ руб.}$$

Задача 2. Распределение рабочих по стажу работы следующее. Определить средний стаж работы рабочих участка.

Стаж рабочих, лет	до 5 лет	5 - 10	10 - 15	15 и более
Количество рабочих, f	2	6	15	7
Серединное значение интервала, x	2,5	7,5	12,5	17,5

Решение. Воспользуемся формулой расчета средней арифметической взвешенной для интервального ряда. Предварительно вычислим серединное значение интервального признака x, дополнив открытые интервалы значениями недостающих границ:

$$x'_1 = \frac{0+5}{2} = 2{,}5;\ x'_2 = \frac{5+10}{2} = 7{,}5;\ x'_3 = \frac{10+15}{2} = 12{,}5;\ x'_4 = \frac{15+20}{2} = 17{,}5$$

Средний стаж рабочих участка

$$x' = \frac{\sum x'f}{\sum f} = \frac{2{,}5\times 2 + 7{,}5\times 6 + 12{,}5\times 15 + 17{,}5\times 7}{2+6+15+7} = 12{,}0 \text{ лет}$$

Задача 3. Имеются следующие данные об экспорте продукции металлургического комбината. Определить средний удельный вес продукции на экспорт.

Вид продукции	Удельный вес Продукции на экспорт, % x	Стоимость продукции на экспорт, тыс. руб. w
Сталь арматурная	40	32100
Прокат листовой	32	42500

Решение. Уд. вес прод. наэкспорт$= \frac{\text{Стоимость продукции на экспорт}}{\text{Стоимость всей продукции}} \times 100$,

или $\frac{w}{x} \times 100$ – стоимость всей продукции. Воспользуемся формулой средней гармонической взвешенной. Средний удельный вес продукции на экспорт

$$\bar{x}=\frac{\sum w}{\sum \frac{w}{x}\times 100}=\frac{32100+42500}{\frac{32100}{40}\times 100+\frac{42500}{32}\times 100}\times 100=\frac{32100+42500}{80250+132812{,}5}\times 100=\frac{74600}{213062{,}5}\times 100=35{,}0\,\%$$

Задача 4. Проведена малая выборка из партии электролампочек для определения продолжительности их службы. Определить значения моды и медианы.

№ лампочки	**Срок горения, час.**
1	1450
2	1400
3	1370
4	1430
5	1400
6	1380
7	1270
8	1420
9	1400

Произведем ранжирование данных по возрастанию 1270; 1370; 1380; 1400; 1400; 1400; 1420; 1430; 1450. Мода Мо = 1400, так как это значение признака встречается три раза.

Место медианы $N_{Me}=\frac{n+1}{2}=\frac{9+1}{2}=5$. Me = 1400 – значение признака на 5-ом месте в ранжированном ряду.

Задача 5. Перевозка грузов по автотранспортному предприятию такова. Определить среднемесячный темп роста объема грузовых перевозок.

Месяц	**Январь**	**Февраль**	**Март**
Перевезено грузов, тыс. т	37,0	40,5	42,0

Решение. Определим цепные коэффициенты роста объема грузовых перевозок

$$T_{P_1}=\frac{40{,}5}{37{,}0}=1{,}095;\; T_{P_2}=\frac{42{,}0}{40{,}5}=1{,}037.$$

Среднемесячный темп роста определим по формуле средней геометрической

$$\overline{T}_P=\sqrt{T_{P_1}\times T_{P_2}}=\sqrt{1{,}095\times 1{,}037}=1{,}066\text{ или }106{,}6\%$$

Задача 6. Имеются следующие данные о товарообороте магазина. Определить среднюю цену мужского костюма.

Товары	**Товарооборот, тыс. руб. w**	**Цена за единицу, руб. x**
Костюм мужской, х/б	50	50
Костюм мужской, п/ш	120	75
Костюм мужской, ч/ш	108	120

Решение. Применим формулу расчета средней гармонической взвешенной

$$\bar{x}=\frac{\sum w_i}{\sum \frac{w_i}{x_i}}=\frac{(50+120+108)\times 1000}{\frac{50000}{50}+\frac{120000}{75}+\frac{108000}{120}}=\frac{278000}{1000+1600+900}=\frac{278000}{3500}=79{,}43\text{ руб.}$$

4. ПОКАЗАТЕЛИ ВАРИАЦИИ

Различие индивидуальных значений признака внутри изучаемой совокупности в статистике называется **вариацией** признака. Она возникает в результате того, что его индивидуальные значения складываются под совокупным влиянием разнообразных факторов. Степень близости данных отдельных единиц X_i к средней измеряется рядом абсолютных, средних и относительных показателей.

Средняя величина отражает тенденцию развития, то есть действие главных причин (факторов), **показатели вариации** измеряют силу воздействия прочих факторов.

4.1. Абсолютные и средние показатели вариации и способы их расчета

Размах вариации $R = X_{max} - X_{min}$ устанавливает только крайние отклонения, то есть пределы выборки.

Среднее абсолютное линейное отклонение $\bar{d} = \frac{\sum|x-\bar{x}|}{n}$ или $\bar{d} = \frac{\sum|x-\bar{x}|f}{\sum f}$

учитывает различия всех единиц изучаемой совокупности

Средний квадрат отклонений (дисперсия) $\sigma^2 = \frac{\sum(x-\bar{x})^2}{n}$ или $\sigma^2 = \frac{\sum(x-\bar{x})^2 f}{\sum f}$. Расчет дисперсии можно производить по формуле: $s^2 = \frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2 = \frac{\sum x^2 f}{\sum f} - \bar{x}^2$

Среднее квадратичное отклонение $\sqrt{\sigma^2} = \sigma$. Чем меньше среднее квадратичное отклонение, тем лучше средняя величина отражает собой всю представленную совокупность. Между средним абсолютным и средним квадратичным отклонением существует соотношение $\sigma \approx 1{,}25\bar{d}$.

Свойства дисперсии.

1. Уменьшение всех значений признака на одну и ту же величину не меняет величины дисперсии	$s^2(x_i - A) = s^2$
2. Уменьшение всех значений признака в k раз уменьшает дисперсию в k^2	$\sigma^2\left(\frac{x_i}{k}\right) = \frac{\sigma^2}{k^2}$
3. Дисперсия от средней меньше дисперсии, исчисленной от любой величины A	$s^2 \le \frac{\sum(x-A)^2 f}{\sum f}$

Альтернативный признак. Обозначим 1 – наличие признака; 0 – отсутствие; p – долю единиц, обладающих данным признаком; q – долю единиц, не обладающих данным признаком (p+q=1).

Среднее значение альтернативного признака $\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{1 \cdot p + 0 \cdot q}{p+q} = p$

Дисперсия альтернативного признака

$$s_p^2 = \frac{\sum (x-\bar{x})^2 f}{\sum f} = \frac{(1-p)^2 p + (0-p)^2 q}{p+q} = \frac{q^2 p + p^2 q}{p+q} = pq = p(1-p)$$

Среднее квадратичное отклонение альтернативного признака $s_p = \sqrt{pq} = \sqrt{p(1-p)}$.

Пример. Доля брака 2 % (p=0.002, q= 0.98). Дисперсия доли брака s_p^2=0.02×0.98=0.0196. Среднее квадратичное отклонение доли брака $s_p = \sqrt{0.0196} = 0.14$.

4.2. Виды дисперсий

Меры вариации для сгруппированных данных определяют следующие показатели.

Общая дисперсия характеризует вариацию признака, который зависит от всех условий в данной совокупности ($\bar{x}_0$ – общая средняя для всей совокупности).	$\sigma_0^2 = \frac{\sum (x_i - \bar{x}_0)^2 f_i}{\sum f_i}$
Внутригрупповая дисперсия отражает вариацию изучаемого признака внутри группы ($\bar{x}_i$ – групповая средняя, вычисляется по правилу (x0+x1)/2).	$\sigma_i^2 = \frac{\sum (x - \bar{x}_i)^2 f_i}{\sum f_i}$
Средняя внутригрупповых дисперсий характеризует случайную вариацию в каждой отдельной группе. Возникает от не учитываемых факторов и не зависит от условий группировки.	$\bar{s}_i^2 = \frac{\sum s_i^2 f_i}{\sum f_i}$
Межгрупповая дисперсия измеряет вариацию групповых средних $\bar{x}_i$ вокруг общей средней. Она измеряет вариацию, обусловленную признаком, положенным в основу группировки.	$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x}_0)^2 f_i}{\sum f_i}$
Общая дисперсия равна сумме средней из внутригрупповых и межгрупповых дисперсий	$\sigma_0^2 = \bar{\sigma}_i^2 + \delta^2$

В сгруппированных данных истинные значения признака заменяются центральными значениями интервалов (x0+x1)/2.

4.3. Показатели относительного рассеивания

Для характеристики меры колеблемости изучаемого признака исчисляются показатели колеблемости в относительных величинах. Они позволяют сравнивать характер рассеивания в различных распределениях.

Относительное линейное отклонение характеризует долю усредненного значения абсолютных отклонений от средней величины.	$K_d = \frac{\bar{d}}{\bar{x}} \times 100$
Коэффициент вариации используется для оценки типичности средних величин. Совокупность считается количественно однородной, если v<33%.	$v = \frac{s}{\bar{x}} \times 100$
Эмпирическое корреляционное отношение показывает тесноту связи группировочного признака и результативного. $\eta = 1 -$ группировочный	$\eta = \sqrt{\frac{\delta^2}{\sigma_0^2}}$

признак целиком определяет вариацию результативного; $\eta = 0 -$ влияние группировочного признака на результативный равно 0. Соотношения Чеддока дают качественную оценку тесноты связи.	

Пример 1. Результаты опроса членов бригады приведены в таблице. Провести группировку бригады по факторному признаку «тех. обучение». Рассчитать значения средних показателей выработки дисперсий по группам и в целом по совокупности.

Табельный №	**Обучение**	**Выработка, шт.**
1	Да	9
2	Нет	8
3	Нет	6
4	да	7
5	нет	7
6	да	8
7	Да	8
8	Нет	7

Решение. Сгруппируем данные по признаку «тех. обучение»

Группы рабочих	**Число рабочих**	**Выработка изделий, шт.**	**Групповая Средняя** $\bar{x}_i$
1. Прошли тех. обучение	4	9,8,8,7	8
2. Не прошли тех. обучение	4	8,7,7,6	7

Общая средняя $\bar{x} = 7{,}5$ шт. Расчет общей дисперсии

Выработка изделий, шт. x	**Частота f**	$(x-\bar{x})$	$(x-\bar{x})^2$	$(x-\bar{x})^2 f$
9	1	1,5	2,25	2,25
8	3	0,5	0,25	0,75
7	3	-0,5	0,25	0,75
6	1	-1,5	2,25	2,25
Итого	8	–	–	6,0

Общая дисперсия выработки $\sigma_0^2 = \dfrac{\sum(x-\bar{x})^2 f}{\sum f} = \dfrac{6{,}0}{8} = 0{,}75$. Рассчитаем дисперсию каждой группы. Расчет дисперсии выработки у группы рабочих, прошедших тех. обучение.

Выработка изделий, шт. x	**Частота f**	$(x-\bar{x})$	$(x-\bar{x})^2$	$(x-\bar{x})^2 f$
9	1	1	1	1
8	2	0	0	0
7	1	-1	1	1
Итого	4	–	–	2

Расчет дисперсии выработки у группы рабочих, не прошедших тех. обучение.

Выработка изделий, шт. x	**Частота f**	$(x-\bar{x})$	$(x-\bar{x})^2$	$(x-\bar{x})^2 f$
8	1	1	1	1
7	2	0	0	0
6	1	-1	1	1
Итого	4	–	–	2

Рассчитаем среднюю из внутригрупповых дисперсий по формуле средней арифметической взвешенной

$$\sigma_1^2 = \frac{2}{4} = 0{,}5, \quad \sigma_2^2 = \frac{2}{4} = 0{,}5, \qquad \bar{\sigma}_i^2 = \frac{\sum \sigma_i^2 f}{\sum f} = \frac{0{,}5 \times 4 + 0{,}5 \times 4}{8} = 0{,}5.$$

Этот показатель характеризует влияние на результативный признак (выработки) всех прочих факторных признаков за исключением признака, положенного в основу группировки (тех. обучение).

Найдем межгрупповую дисперсию

$$\delta^2 = \frac{\Sigma(\bar{x}_i - \bar{x}_0)^2 f}{\sum f} = \frac{(8{,}0 - 7{,}5)^2 \times 4 + (7 - 7{,}5)^2 \times 4}{8} = 0{,}25.$$

Различие в выработке двух групп вызывается тем, что рабочие 1-ой группы прошли обучение, а 2-ой группы не прошли.

Общая дисперсия $\sigma^2 = \bar{\sigma}_i{}^2 + \delta^2$ $0{,}75 = 0{,}5 + 0{,}25$. Общая дисперсия, возникающая под воздействием всех факторов, должна быть равна сумме дисперсий, возникающих под влиянием всех прочих факторов, и дисперсии, возникающей за счет фактора группировки.

Влияние группировочного признака на результативный оценим корреляционным отношением

$$\eta = \sqrt{\frac{\delta^2}{\sigma_0{}^2}}; \qquad \eta = \sqrt{\frac{0{,}25}{0{,}75}} = 0{,}58.$$

Следовательно, фактор, положенный в основу группировки, существенно влияет на производительность труда, но существуют и другие факторы, влияние которых тоже значительно.

Задача 1. Имеются данные о возрастном составе рабочих цеха (лет): 18, 38, 28, 29, 26, 38, 34, 22, 28, 30, 22, 23, 35, 33, 27, 24, 30, 32, 28, 25, 29, 26, 31, 24, 29, 27, 32, 25, 29, 29. Построить интервальный ряд распределения для 7 групп и графическое изображение ряда. Вычислить показатель центра распределения, т.е среднее значение возраста ($\bar{x}$) и показатели вариации.

Решение. 1) Определим величину интервала группировки

$$i = \frac{R}{n} = \frac{38-18}{7} = \frac{20}{7} \approx 3$$

Построим интервальный ряд распределения рабочих по возрасту.

Группы рабочих по возрасту, (x)	**Число Рабочих, F**		x'	$x'f$	$(x-\bar{x})^2$	$(x-\bar{x})^2 f$
18-21	1	1	19,5	19,5	84,64	84,64
21-24	3	4	22,5	67,5	38,44	115,32
24-27	6	10	25,5	153,0	10,24	61,44
27-30	10	20	28,5	285,0	0,04	0,4
30-33	5	25	31,5	157,5	7,84	39,2
33-36	3	28	34,5	103,5	33,64	100,92
36-39	2	30	37,5	75,5	77,44	154,88
Итого	30	–	–	861,0	–	556,8

2) Расчет показателя среднего возраста рабочих $\bar{x} = \frac{\sum x'f}{\sum f} = 28{,}4$ года

3) Расчет структурных средних:

а) Модальный интервал 27-30 лет, так как $f_{max} = 10$. Значение моды

$$Mo = x_{Mo} + i\frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})} = 27 + 3\frac{10-6}{(10-6)+(10-5)} = 28{,}33$$

б) Место медианы $N_{Me} = \frac{n+1}{2} = 15{,}5.$

Медианный интервал приходится на группу 27-30 лет находящуюся в середине нечетного ряда

$$Me = x_{Me} + i\frac{\frac{n+1}{2} - s_{Me-1}}{f_{Me}} = 27 + 3\frac{\frac{30+1}{2} - 10}{10} = 28{,}65.$$

3) Показатели вариации

Среднее абсолютное линейное отклонение $d = \frac{\sum(x' - \bar{x})f}{\sum f} = \frac{116{,}0}{30} = 3{,}87$ года.

Среднее квадратическое отклонение $\sigma = \sqrt{\frac{\sum(x' - \bar{x})^2 f}{\sum f}} = \sqrt{\frac{556{,}8}{30}} = 4{,}31$ года.

Коэффициент вариации $v = \frac{\sigma}{\bar{x}}100 = \frac{4{,}31}{28{,}7} \times 100 = 15\%$.

Вариация незначительна, следовательно, совокупность однородна.

5. ВЫБОРОЧНЫЙ МЕТОД В СТАТИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ

5.1. Понятие о выборочном исследовании. Способы отбора единиц из генеральной совокупности

Статистическое наблюдение можно организовать **сплошное** и **несплошное**. В **выборочном** методе несплошного наблюдения обобщающие показатели совокупности устанавливаются по некоторой ее части (сравнительно небольшой –5-10% всей совокупности).

Совокупность, из которой производится отбор части единиц, называется **генеральной**. Отобранная часть единиц называется **выборочной совокупностью**, или **выборкой**.

При использовании выборочного метода исследования осуществляется в более короткие сроки и с минимальными затратами труда и средств. Это повышает оперативность и уменьшает ошибки регистрации. Выборочный метод является единственно возможным при разрушающем контроле качества продукции. Он распространен в государственной и ведомственной статистике (бюджетные обследования семей рабочих, крестьян, служащих; обследование жилищных условий), в торговле (спрос населения на товар; эффективность новых форм торговли) и т.д.

В основе отбора единиц положен **принцип равных возможностей попадания в выборку каждой единицы** генеральной совокупности.

М**етоды отбора единиц в выборку**. **Повторный отбор** – выбранная единица после фиксации значения возвращается в генеральную совокупность, где может быть опять выбрана. **Бесповторный отбор** – не возвращается.

Способы отбора определяются правилами формирования выборочной совокупности.

Собственно-случайный – образуется в результате непреднамеренного отбора отдельных единиц из генеральной совокупности. Количество отбираемых единиц определяется долей выборки.

Механическая выборка – отбор единиц производится из генеральной совокупности, разбитой на равные интервалы (группы). Генеральная совокупность механически разбивается на равновеликие группы, из каждой группы в выборку отбирается лишь одна единица. Для определения ошибки механической выборки используют формулы случайной безповторной выборки.

Типическая выборка – генеральная совокупность разделяется на однородные типические группы. Затем из каждой группы отбираются единицы в выборку (производительность труда рабочих по группам квалификации).

Серийная или гнездовая выборка – выбираются из генеральной совокупности целые серии и внутри серии обследуются все единицы (коробка, пачка, ящик, …). Отбор серий производится любым способом по схеме бесповторного отбора.

Комбинированная – отбор групп, а внутри отбор отдельных единиц.

Соблюдение правил выборочного метода гарантирует репрезентативность выборочной совокупности (ее воспроизводство генеральной совокупности по изучаемым признакам).

Основные характеристики параметров генеральной и выборочной совокупностей:

N – объем генеральной совокупности (число входящих в нее единиц);

n – объем выборки (число обследованных единиц);

$\bar{x}$ – генеральная средняя (среднее значение признака генеральной совокупности);

$\tilde{x}$ – выборочная средняя;

p – генеральная доля (доля единиц, обладающих данным значением признака в общем числе единиц генеральной совокупности;

w – выборочная доля;

σ^2 – генеральная дисперсия (дисперсия признака в генеральной совокупности;

s^2 – выборочная дисперсия этого же признака;

σ – среднее квадратическое отклонение в генеральной совокупности;

s – среднее квадратическое отклонение в выборке.

Задача выборочного обследования состоит в том, чтобы на основе характеристик выборочной совокупности (частости w или средней $\tilde{x}$) получить достоверные суждения о показателях (доли p или средней $\bar{x}$) в генеральной совокупности.

5.2. Ошибки выборки

Расхождение между характеристиками выборки и генеральной совокупности составляет **ошибку выборки**. Она зависит от вариации изучаемого признака, численности выборки, методов отбора единиц в выборку, принятого уровня достоверности результата.

Количество отобранных в выборочную совокупность единиц определяют исходя из принятой **доли выборки $К_В$**

$$\mathbf{К_В = Доля\ выборки =} \frac{\text{объмвыборки}}{\text{объемгенеральнойсовокупности}} = К_В = \frac{n}{N}.$$

Например, при $К_В$ = 5% выборки из партии N = 1000 единиц объем выборки n = 50 единиц, а при $К_В$ = 10% – n = 100 единиц.

Различают два вида обобщающих показателей:

- **относительную величину альтернативного признака**, т.е. долю p (удельный вес) единиц совокупности, обладающих данным значением признака (например, доля нестандартных изделий в партии товара, удельный вес продукции собственного производства в товарообороте общепита, удельный вес продавцов в общей численности);
- **среднюю величину количественного признака**.

Выборочная доля ($w = \frac{m}{n}$ **)**, или частость, определяется отношением числа единиц m, обладающих изучаемым признаком, к общему числу выборочной совокупности n.

Например, если из 100 деталей (n = 100) 95 оказались стандартными (m = 95), то выборочная доля $w = \frac{95}{100} = 0{,}95$.

Для характеристики выборочных показателей используют понятие ошибки выборки.

Ошибка выборки ε (репрезентативности) представляет разность выборочных и генеральных характеристик:

- для средней количественного признака $e_{\tilde{x}} = |\tilde{x} - \bar{x}|$;
- для доли (альтернативного признака) $e_w = |w - p|$.

При случайном повторном отборе средние ошибки рассчитываются по формулам:

- для среднего количественного признака $m_{\bar{x}} = \sqrt{\frac{s^2}{n}}$,
- для доли (альтернативного признака) $\mu_p = \sqrt{\frac{p(1-p)}{n}}$.

Так как s^2, дисперсия признака в генеральной совокупности точно неизвестна, на практике пользуются значением дисперсии S^2, рассчитанным для выборочной совокупности на основании закона больших чисел (выборочная совокупность при достаточно большом объеме выборки достаточно точно воспроизводит характеристики генеральной совокупности). Генеральная дисперсия выражается через выборочную $s^2 = S^2 \frac{n}{n-1}$. При достаточно больших n можно принять $s^2 \approx S^2$. Среднее и дисперсия количественного признака в выборке определяется по формулам

$$\tilde{x} = \frac{\sum x_i}{n} \quad или \quad \tilde{x} = \frac{\sum x_i f_i}{\sum f_i}; \qquad S^2 = \frac{\sum (x_i - \tilde{x})^2}{n} \quad или \quad S^2 = \frac{\sum (x_i - \tilde{x})^2 f_i}{\sum f_i}.$$

При случайном повторном отборе формулы средних ошибок примут вид

$$m_{\tilde{x}} = \sqrt{\frac{S^2}{n}} \quad \text{и} \quad m_w = \sqrt{\frac{w(1-w)}{n}}$$

При случайном бесповторном отборе численность генеральной совокупности в ходе выборки сокращается. Формулы для среднего значения количественного признака и для доли имеют вид:

$$m_{\tilde{x}} = \sqrt{\frac{S^2}{n}\left(1 - \frac{n}{N}\right)}, \qquad \mu_w = \sqrt{\frac{w(1-w)}{n}\left(1 - \frac{n}{N}\right)}.$$

Малые выборки. Под малой выборкой понимается несплошное статистическое обследование с небольшим числом единиц – от 4 до 30. Тогда:

1. Генеральная дисперсия выражается через выборочную по формуле: $s^{2} = S^2 \dfrac{n}{n-1}$.

2. Средняя ошибка малой выборки при повторном отборе имеет вид: $m_{\tilde{x}} = \sqrt{\dfrac{S^2}{n-1}}$.

Распространение выборочных результатов на генеральную совокупность производится с учетом закона больших чисел, который определяет с заданной вероятностью предел возможной ошибки различий средних

$$P[|\,\bar{x} - \tilde{x}\,| \le \Delta_{\tilde{x}}] = \Phi(t)\ , \quad \Delta_{\tilde{x}} = tm_x\,, .$$

где P – вероятность события, указанного в скобках, t – коэффициент доверия, Φ(t) – интеграл нормальной плотности распределения с нулевым средним и единичной дисперсией на отрезке [-t,t]. Функция Φ(t) задается таблицами. Для выборок при n>30 значения t и Φ(t) приведены в таблице.

t	1	1.960	2	2.58	3
Φ(t)	0.683	0.95	0.954	0.99	0.997

Таким образом при заданной вероятности Φ(t) и определяемом из таблицы значении t можно определить **предельные значения характеристик генеральной совокупности и их доверительные интервалы**:

- для средней величины количественного признака $\bar{x} = \tilde{x} \pm \Delta_{\tilde{x}}\,, \quad \Delta_{\tilde{x}} = tm_{\tilde{x}}$
- для доли альтернативного признака $p = w \pm \Delta_w\,, \quad \Delta w = tm_w$

Например, при P = 0,683 → t = 1. Следовательно в 68,3% случаев ошибка не выйдет за пределы ±1μ (одной средней ошибки выборки).

Распространение характеристик выборки на генеральную совокупность производится следующими способами.

Способ прямого расчета состоит в том, что показатели выборочной доли w, или средней $\tilde{x}$ распространяются на генеральную совокупность с учетом ошибки выборки (количество поступивших в партии товаров нестандартных изделий, …).

Способ поправочных коэффициентов применяется в случаях, когда целью выборочного метода является уточнение результатов сплошного учета (ежегодная перепись скота у населения, …). Для этого после обобщения данных сплошного учета практикуется 10% выборочное обследование с определением так называемого «процента недоучета».

Пример 1. При проверке качества хлебобулочных изделий проведено 5% выборочное обследование партии батонов. Из 100 отобранных в выборку батонов 90 штук соответствовали требованиям стандарта. Средний вес одного батона 500,5 г при среднем квадратичном отклонении ± 15,4 г.

Установить возможные значения доли стандартных изделий и среднего веса одного изделия для всей партии товара.

Решение. Значение альтернативного признака m = 90. Число единиц выборки n = 100. Значение альтернативной доли (частости) определим как $w = \dfrac{m}{n} = \dfrac{90}{100} = 0{,}9$.

Средний вес изделия в выборке известен и равен $\tilde{x} = 500{,}5$ г.

$$s_x = \pm 15{,}4 \text{ г}; \quad K_B = \frac{n}{N}; \quad \mathrm{N} = \frac{\mathrm{n}}{\mathrm{K_B}} = \frac{100}{0{,}05} = 2000.$$

Найдем возможные значения ошибки выборки:

– для среднего веса изделия

$$\mu_x = \sqrt{\frac{\sigma_x^{\ 2}}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{15{,}4^2}{100}\left(1 - \frac{100}{2000}\right)} = \pm 1{,}5 \text{ г};$$

– для показателя средней доли стандартных изделий

$$\mu_w = \sqrt{\frac{w(1-w)}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{0{,}9(1-0{,}9)}{100}\left(1 - \frac{100}{2000}\right)} = \pm 0{,}029;$$

или от 87,1 до 92,9% удельного веса стандартных изделий в партии.

Средний вес изделий во всей партии $\bar{x} = \tilde{x} \pm \mu_{\tilde{x}} = 500{,}5 \pm 1{,}5$ $\quad 499 \le \bar{x} \le 502$

Предельную ошибку выборки определим с вероятностью 0,99. При этом t = 2,6. Доля стандартных изделий во всей партии $\mathrm{p} = \mathrm{w} \pm \mathrm{t}\mu = 0{,}9 \pm 2{,}6 \times 0{,}029$ или от 0,825 до 0,975.

Это значит, что в 99 случаях из 100 удельный вес стандартных изделий во всей партии будет находится в пределах от 82,5 до 97,5%.

Средний вес изделия во всей партии $\bar{x} = \tilde{x} \pm \mathrm{t}\mu = 500{,}5 \pm 2{,}6 \times 1{,}5$, от 500,5 – 3,9 до 500,5 + 3,9.

С вероятностью равной 0,99 можно утверждать, что средний вес изделия во всей партии находится в пределах от 496,9 до 504,4 г.

5.3. Оптимальная численность выборки

Размер ошибки средних повторной выборки зависит от численности выборочной совокупности n

- для средней величины признака $\Delta_{\tilde{x}} = t\sqrt{\frac{s_x^{\ 2}}{n}} \;\rightarrow\; \Delta_{\tilde{x}}^2 = t^2 \frac{s_x^{\ 2}}{n} \;\rightarrow n = t^2 \frac{s_x^{\ 2}}{\Delta_{\tilde{x}}^2}$;
- для доли альтернативного признака $\Delta_w^2 = t^2 \frac{w(1-w)}{n} \;\rightarrow\; n = \frac{t^2 w(1-w)}{\Delta_w^2}$;

Для способа бесповторного отбора численность выборки определятся аналогично

- – для средней величины количественного признака $n_x = \dfrac{Nt^2 s_x^{\ 2}}{N\Delta_x^2 + t^2 s_x^{\ 2}}$
- – для доли альтернативного признака $n_w = \dfrac{Nt^2 w(1-w)}{N\Delta_{w^2}^2 + t^2 w(1-w)}$

Пример 2. Исходя из требований ГОСТа, необходимо установить оптимальный объем выборки из партии нарезных батонов (2000 шт.), чтобы с вероятностью 0,997 предельная ошибка не превышала 3% веса 500 г батона.

Решение. Относительную ошибку выборки выразим абсолютной величиной

$$\Delta_x = \frac{500(\pm 3)}{100} = \pm 15 \text{ г}$$

Оптимальная численность выборки для бесповторного отбора

$$n_x = \frac{Nt^2\sigma_x{}^2}{N\Delta_x^2 + t^2\sigma_x{}^2} = \frac{2000 \times 3^2 \times 15{,}4^2}{2000 \times 15^2 + 3^2 \times 15{,}4^2} = 10 \text{ шт.}$$

Из 2000 шт. достаточно проверить 10.

Задача 1. Для определения скорости расчетов с кредиторами предприятий корпорации в коммерческом банке была проведена случайная выборка n = 100 платежных документов, по которым средний срок перечисления денег оказался равным $\tilde{x} = 22$ дням со стандартным отклонением σ = 6 дней. С вероятностью p = 0,954 определить предельную ошибку выборочной средней и доверительные интервалы (пределы) средней продолжительности расчетов предприятий.

Решение. Предельную ошибку $\Delta = t\mu$ определяем по формуле повторного отбора, так как численность генеральной совокупности неизвестна. t = 2.

$$\Delta_{\tilde{x}} = t\sqrt{\frac{s^2}{n}} = 2\sqrt{\frac{36}{100}} = 2 \times 0{,}6 = 1{,}2.$$

Предельная относительная ошибка выборки, %

$$\Delta_{\%} = \frac{\Delta_{\tilde{x}}}{\tilde{x}} \times 100 = \frac{1{,}2}{22} \times 100 = 5{,}45.$$

Генеральная средняя $\bar{x} = \tilde{x} \pm \Delta_{\tilde{x}}$

$$22 - 1{,}2 \le \bar{x} \le 22 + 1{,}2 \quad 20{,}8 \le \bar{x} \le 23{,}2$$

С вероятностью 0,954 можно утверждать, что средняя продолжительность расчетов предприятий колеблется в пределах от 20,8 до 23,2 дней.

Задача 2. Для определения урожайности зерновых культур проведено выборочное обследование 100 хозяйств региона различных форм собственности, в результате которого получены сводные данные.
С вероятностью 0,954 определить предельную ошибку выборочной средней и доверительные пределы средней урожайности зерновых по всем хозяйствам региона.

Хозяйства по форме собственности	**Кол-во хозяйств, f_i**	**Средняя урожайн., ц/га x_i**	**дисперсия урожайн., $s_i{}^2$**
Коллективные	30	18	15
Акционерные сообщества	50	20	25
Крестьянские (фермерские)	20	28	40
Итого	100	–	–

Решение. Предельную ошибку определим для типичной выборки $\Delta_{\tilde{x}} = t\sqrt{\frac{\bar{s}_i{}^2}{n}}$.

Определим среднюю из внутригрупповых дисперсий

$$\bar{s}_i{}^2 = \frac{\sum s_i{}^2 f_i}{\sum f_i} = \frac{15 \times 30 + 25 \times 50 + 40 \times 20}{100} = \frac{2500}{100} = 25.$$

Предельная ошибка выборки $\Delta_{\tilde{x}} = 2\sqrt{\frac{25}{100}} = 1{,}0.$

Генеральная средняя $\bar{x} = \tilde{x} \pm \Delta_{\tilde{x}}.$

Выборочная средняя $\tilde{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{18 \times 30 + 20 \times 50 + 28 \times 20}{100} = \frac{2100}{100} = 21$.

Доверительные пределы $21 - 1 \le \bar{x} \le 21 + 1 \quad 20 \le \bar{x} \le 22$

С вероятностью 0,954 можно гарантировать, что средняя урожайность зерновых культур по региону будет не менее, чем 20 ц/га и не более 22 ц/га.

Задача 3. Для определения среднего возраста 1200 студентов необходимо провести выборочное обследование методом случайного бесповторного отбора. Предварительно установлено, что среднее квадратическое отклонение возраста студентов равно 10 годам.

Сколько студентов нужно обследовать, чтобы с вероятностью 0,954 средняя ошибка выборки не превышала 3 года?

Решение. Рассчитаем необходимую численность выборки по формуле бесповторного отбора, учитывая, что t =2 при p = 0,954

$$n = \frac{t^2 \sigma^2 N}{\Delta_x^2 N + t^2 \sigma^2} = \frac{1200 \times 2^2 \times 10^2}{3^2 \times 1200 + 2^2 10^2} = \frac{480000}{11200} \approx 43$$

Выборка численностью 43 человека обеспечит заданную точность при бесповторном отборе.

6. РЯДЫ ДИНАМИКИ

Рядами динамики называются статистические данные, отображающие развитие изучаемого явления во времени. В каждом ряду динамики имеются два основных элемента: **показатель времени** t; соответствующие им **уровни развития** изучаемого явления y.

Номер	i	0	1	2	3
Время	t_i	1991 (t_0)	1992 (t_1)	1993 (t_2)	1994 (t_3)
Уровень	y_i	788 (y_0)	810 (y_1)	867 (y_2)	1054 (y_3)

В качестве **показаний времени** в рядах динамики выступают либо определенные даты (моменты), либо отдельные периоды (годы, кварталы, …)

Уровни рядов динамики отображают количественную оценку (меру) развития во времени изучаемого явления. Они могут быть выражены **абсолютными, относительными или средними величинами.** Уровни ряда должны быть **сопоставимыми.**

Моментные ряды отображают состояние изучаемых явлений на определенные даты (моменты) времени (товарные запасы, количество оборудования и т.д.).

Интервальные ряды отображают итоги развития (функционирования) явлений за отдельные периоды (интервалы) времени (данные о товарообороте, поступление и реализация товаров, суммы издержек обращения и др.).

Графическое изображение ряда динамики в виде **линейных, столбиковых и секторных диаграмм** позволяет наглядно представить развитие явления во времени.

6.1. Статистические показатели динамики

Показатели ряда динамики вычисляют посредством сопоставления его уровней на основе **постоянной** или **переменной базы сравнения**:

1. **На постоянной базе** (базисный показатель)– сравнение с фиксированным **базисным** уровнем.
2. **На переменной базе** (цепной показатель)– сравнение текущего уровень ряда (y_i)с предыдущим (y_{i-1}).

Показатель	Цепной	Базисный	За весь период
Абсолютный прирост (в натуральных единицах)	$\Delta y_i^ц = y_i - y_{i-1}$	$\Delta y_i^б = y_i - y_0$	$\Delta y = y_n - y_0 = \Delta y_n^б = = \sum \Delta y_i^ц = \sum(y_i - y_{i-1})$
Коэффициент роста (долях)	$Kp_i^ц = \dfrac{y_i}{y_{i-1}}$	$Kp_i^б = \dfrac{y_i}{y_0}$	$Kp_i^б = \prod Kp_i^ц = = Kp_n^б = \dfrac{y_i}{y_0}$
Темп роста (%)	$Tp_i^ц = Kp_i^ц \cdot 100\% = = \dfrac{y_i}{y_{i-1}} \cdot 100\%$	$Tp_i^б = Kp_i^б \cdot 100\% = = \dfrac{y_i}{y_0} \cdot 100\%$	$Tp = Kp \cdot 100\% = = \dfrac{y_n}{y_0} \cdot 100\%$

Коэффициент прироста (долях)	$Кпр_i^ц = \frac{\Delta y_i^ц}{y_{i-1}} =$ $= Кр_i^ц - 1$	$Кпр_i^б = \frac{\Delta y_i^б}{y_0} =$ $= Кр_i^б - 1$	$Кпр = Кпр_n^б =$ $= Кр_i - 1$
Темп прироста (%)	$Тпр_i^ц = Кпр_i^ц \cdot 100$	$Тпр_i^б = Кпр_i^б \cdot 100$	$Тпр = Кпр \cdot 100$

Абсолютное значение одного 1% прироста показывает, какая абсолютная величина скрывается за относительным показателем – одним процентом прироста. Представляет собой отношение абсолютного прироста к темпу прироста, выраженному в %

$$A = \frac{\Delta y_i^ц}{Тпр_i^ц} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}}100} = \frac{\Delta y(y_{i-1})}{\Delta y \times 100} = \frac{y_{i-1}}{100} = 0{,}01 y_{i-1}.$$

Коэффициент наращивания – измеряет наращивание во времени экономического потенциала $$Тн_i = \frac{\Delta y_i^ц}{y_0} = Кр_i^б - Кр_{i-1}^б.$$

6.2. Средние показатели в рядах динамики

К **обобщающим показателям динамики** относятся: средний уровень, средний абсолютный прирост, средний темп роста и прироста и др.

Средний уровень ряда называется **хронологической средней** или **временной средней**.

Средний уровень интервального ряда вычисляется по формуле средней арифметической простой $$\bar{y} = \frac{y_0 + y_2 + \mathbf{K} + y_n}{n} = \frac{\sum y}{n}.$$

Если отдельные периоды интервального ряда имеют неодинаковую длину, то для определения среднего уровня следует воспользоваться средней арифметической взвешенной $\bar{y} = \frac{\sum t_i y_i}{\sum t_i}$, где t_i – длины интервалов.

Средний уровень моментного ряда $\bar{y} = \frac{y_0/2 + y_1 + \mathbf{K} + y_n/2}{n}$.

Для моментных рядов при неравных интервалах применяется взвешивание полусумм смежных уровней по продолжительности периода между ними

$$\bar{y} = \frac{(y_0 + y_1)t_0 + (y_1 + y_2)t_1 + \mathbf{K} + (y_{n-1} + y_n)t_{n-1}}{2(t_0 + t_1 + \mathbf{K} + t_{n-1})}.$$

Средний абсолютный прирост – средняя величина из абсолютных приростов (за равные промежутки времени одного периода) ряда динамики

$$\overline{\Delta y} = \frac{\sum \Delta y_i^ц}{n} = \frac{\Delta y_0^ц + \Delta y_1^ц + \mathbf{K} + \Delta y_{n-1}^ц}{n} = \frac{y_n - y_0}{n} = \frac{\Delta y}{n}$$

Средний коэффициент и темп роста исчисляется по формуле средней геометрической $\overline{Kp} = \sqrt[n]{Kp_0 \times Kp_1 \times \mathbf{K} \times Kp_n} = \sqrt[n]{\Pi\, Kp} = \sqrt[n]{\frac{y_n}{y_0}}$, $\overline{Tp} = \overline{Kp} \cdot 100 = \sqrt[n]{\frac{y_n}{y_0}} \cdot 100$.

Средний коэффициент и темп прироста на основе взаимосвязи между темпами роста и прироста $\overline{Knp} = \overline{Kp} - 1$, $\overline{T}_{np} = \overline{T}_p - 100$

6.3. Изучение тенденции развития

В исследовании закономерностей динамики выявляют общую тенденцию развития (тренд).

Метод укрупненных интервалов – для выявления тренда в рядах колеблющихся уровней. Первоначальный ряд преобразовывается в ряды более продолжительных периодов (месячные в квартальные, квартальные в годовые, …).

Метод скользящей средней (сглаживание). Получают текущее среднее для t_i, включая в него одинаковое число (m) лет до и после текущего уровня при нечетном числе (2m+1) сглаживаемых уровней (2m+1: 3, 5, 7 и т.д.). При четном числе: число членов скользящей средней 2m, срединный уровень $t_{i+1/2} = (t_i + t_{i+1})/2$, то есть сдвиг по времени вычисляемого среднего уровня, для осреднения берется по m членов справа и слева от срединного уровня.

Метод аналитического выравнивания. Для получения описания в виде плавной линии развития (тренда) используют аналитическое выравнивание, т.е. находят уравнение закономерности изменения явления как функции времени $\tilde{y}_t = f(t)$. Вид уравнения определяется характером динамики развития конкретного явления. Логический анализ при выборе вида уравнения может быть основан на рассчитанных показателях динамики:

- – если стабильны абсолютные приросты (первые разности уровней приблизительно равны), сглаживание может быть выполнено по прямой $\tilde{y} = a_0 + a_1 t$, где a_0, a_1 – переменные; t – время; $a_1 > 0$ – возрастание; $a_1 < 0$ – снижение.
- – если абсолютные приросты равномерно увеличиваются (вторые разности уровней приблизительно равны), можно принять параболу второго порядка $\tilde{y} = a_0 + a_1 t + a_2 t^2$, где a_2 –развитие в единицу времени; $a_2 > 0$ – ускорение развития; $a_2 < 0$ – замедление развития.

Для выравнивания применяются функции: показательная $\hat{y} = a_0 \times a_1^t$, гиперболическая $\hat{y} = a_0 \times a_1 \frac{1}{t}$ и другие. На практике выбор кривой может быть основан на анализе графического изображения ряда. Целесообразно предварительно провести сглаживание ряда. Если условия формирования ряда изменяются, то расчет параметров следует вести по периодам.

Определение параметров уравнений основано на методе наименьших квадратов: $\sum(\hat{y}_t - y_t)^2 = \min$. Для определения параметров уравнения принятого для выравнивания составляется система нормальных уравнений. Для прямой $\hat{y}_t = a_0 + a_1 t$ определяются параметры a_0 и a_1

$$\begin{cases} na_0 + a_1 \sum t = \sum y & a_0 = \dfrac{\sum y \sum t^2 - \sum ty \sum t}{n \sum t^2 - \sum t \sum t} \\ a_0 \sum t + a_1 \sum t^2 = \sum yt & a_1 = \dfrac{n \sum ty - \sum t \sum y}{n \sum t^2 - \sum t \sum t} \end{cases},$$

где y –уровни; n – число уровней, t – время ряда динамики.

Система упрощается, если t преобразовать $\tilde{t} = (t - \bar{t}) \cdot c$, т.е. так, чтобы их сумма равнялась 0. Когда t – целое, то начало отсчета переносится в середину периода. Тогда для преобразованных моментов t

$$\begin{cases} na_0 = \sum y; & \text{откуда } a_0 = \dfrac{\sum y}{n}; \ a_1 = \dfrac{\sum yt}{\sum t^2} \\ a_1 \sum t^2 = \sum yt; & \text{где } \sum t^2 = \dfrac{n(n^2-1)}{3} - \text{чет.} \quad \sum t^2 = \dfrac{n(n^2-1)}{12} - \text{нечет.} \end{cases}$$

Если число уровней четное, то c=2, $\bar{t} = (t_1 + t_n)/2$, а преобразованные значения t принимают вид $-7 -5 -3 -1 +1 +3 +5 +7$.

При нечетном числе членов ряда c=1, $\bar{t} = t_{n/2}$, отсчет ведется от середины ряда, взятой за 0. $-2 -1 \ 0 +1 +2$.

По полученной модели динамического ряда определяются уровни и ошибка аппроксимации (среднее квадратичное отклонение тренда) по формуле

$$s_{\hat{y}} = \sqrt{\frac{\sum(y - \hat{y}_t)^2}{n - m}},$$

где y и $\hat{y}$ – соответственно фактические и расчетные уровни ряда; n – число уровней ряда; m – число параметров в уравнении тренда. Чем меньше среднее квадратичное отклонение, тем точнее подобрана модель динамического ряда.

6.4. Прогнозирование в рядах динамики

Основой прогнозирования является предположение, что закономерность, действующая внутри анализируемого ряда динамики, сохраняется и в дальнейшем. Точность прогноза зависит того, насколько обоснованными окажутся предположения о сохранении на будущее действий факторов, сформировавших в ряду динамики его компоненты.

Установление сроков прогнозирования зависит от задачи исследования. Чем короче сроки упреждения, тем надежнее результаты прогноза.

При прогнозировании выбор методов зависит от характера изменений в базисном ряду динамики и поставленной задачи исследования.

1. При прогнозировании на базе ряда динамики с постоянными абсолютными приростами $\Delta y \cong const$ (**на базе абсолютных приростов**) применяем формулу

$$y_{n+\mathbf{l}} = y_n + \overline{\Delta y} \times \mathbf{l},$$

где $y_{n+\mathbf{l}}$ – экстраполируемый уровень; y_n – конечный уровень ряда; $\mathbf{l}$ – срок прогноза.

2. При прогнозировании на базе ряда со стабильными темпами роста $Tp \cong const$ (**на базе темпов роста**) применяется формула $y_{n+\mathbf{l}} = y_n \left(\overline{T}_p / 100\right)^{\mathbf{l}}$.

3. При прогнозировании тренда **на основе модели аналитического выравнивания** применяется трендовая модель $y_t = a_0 + a_1 t \quad t = \mathbf{l}$.

Прогнозируемые уровни задаются вероятностями (р) принадлежности интервалу. Для границ интервалов используется формула $\hat{y}_t \pm t_a^{n-m} s_e$, где $\hat{y}_t$ – точечный прогноз, рассчитанный по модели; t_a^{n-m} – коэффициент доверия распределения Стьюдента, при уровне значимости α=1-p и степенях свободы *n-m* (Например при $\alpha = 0{,}05$ (p=95%) и n-m=8 t_a^{n-m} = 2.306). Остаточное среднеквадратическое отклонение тренда, скорректированное по числу степеней свободы (n – m)

$$s_{\hat{y}} = s_e = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-m}},$$

где n – число уровней базисного ряда динамики: m – число параметров модели.

Пример 1.Имеются данные о продаже легковых автомобилей .

Т	1991	1992	1993	1994
У	788	810	867	1054

Определить показатели динамики продаж от года к году (относительные показатели) и средние за весь анализируемый период.

Решение. Расчет показателей динамики

Наименование показателя		Годы			
		1991	**1992**	**1993**	**1994**
1	**2**	**3**	**4**	**5**	**6**
1. Абсолют. прирост, Δ, тыс. шт.	с переменной базой $\Delta_y = y_t - y_{t-1}$	–	810-788 22	867-810 57	1054-867 187
	с постоянной базой $\Delta_y = y_t - y_0$	–	810-788 22	867-788 79	1054-788 266
2. Коэффициент роста, K_p	с переменной базой $K_p = \frac{y_t}{y_{t-1}}$	1	$\frac{810}{788} = 1{,}028$	$\frac{867}{810} = 1{,}07$	$\frac{1054}{867} = 1{,}212$
	с постоянной базой $K_p = \frac{y_t}{y_0}$	1	$\frac{810}{788} = 1{,}028$	$\frac{867}{788} = 1{,}1$	$\frac{1054}{788} = 1{,}334$

3. Темп роста T_p% $T_p = \frac{y_t}{y_{t-1}}100$	с переменной базой $T_{p_t} = K_p \times 100$	100	102,8	107	121,2
	с постоянной базой $T_{p_t} = K_p \times 100$	100	102,8	110	133,4
4. Темп прироста, $T_{пр}$% $T_{пр} = \frac{\Delta y}{y_{i-1}} \times 100$	с переменной базой $T_{пр_ц} = (T_{p_t} - 1) \times 100$	–	102,8-100 =2,8	7	21,2
	с постоянной базой $T_{пр_б} = (T_{p_t} - 1) \times 100$	–	2,8	10	33,4
5. Абсолютное значение 1% прироста, А, тыс. шт.	с переменной базой $A = \frac{\Delta y}{T_{пр\%}} = 0{,}01 y_{i-1}$	–	$\frac{22}{2{,}8} = 7{,}86$	$\frac{57}{27} = 8{,}14$	$\frac{187}{21{,}2} = 8{,}86$

Средний уровень интервального ряда динамики

$$\bar{y} = \frac{\sum y}{n} = \frac{788 + 810 + 867 + 1054}{4} = \frac{3516}{4} = 879 \text{ тыс. шт.}$$

Средний абсолютный прирост

$$\bar{\Delta} = \frac{\sum \Delta_t}{n-1} = \frac{22 + 57 + 184}{4-1} = \frac{263}{3} = 87{,}67 \text{ тыс. шт.,}$$

или $$\bar{\Delta} = \frac{y_n - y_1}{n-1} = \frac{1054 - 788}{4-1} = 87{,}67 \text{ тыс. шт.}$$

Средний коэффициент роста $\bar{K}_p = \sqrt[n-1]{K_{p_1} \times \mathbf{K} \times K_{p_{n-1}}} = \sqrt[3]{1{,}028 \times 1{,}07 \times 1{,}212} = \sqrt[3]{1{,}333} = 1{,}101$,

Или $\bar{K}_p = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[3]{\frac{1051}{788}} = 1{,}101$.

Средний темп роста $\bar{T}_p = \bar{K}_p \times 100 = 1{,}101 \times 100 = 110{,}1\%$.

Средний темп прироста $\bar{T}_{пр} = (\bar{K}_p - 1) \times 100 = (1{,}101 - 1) \times 100 = 10{,}1\%$,

Или $\bar{T}_{пр} = \bar{T}_p - 100 = 10{,}1 - 100 = 10{,}1\%$.

Средняя величина абсолютного значения 1% прироста

$$\bar{A} = \frac{\bar{\Delta}}{\bar{T}_{пр}} = \frac{87{,}67}{10{,}1} = 8{,}68 \text{ тыс. шт.}$$

Пример 2. По данным о розничном товарообороте нужно провести анализ основной тенденции развития товарооборота (млрд. руб.)

Год	**Объем розничн. товароб., y**	**Темп роста, %**	**Абсолютный прирост, млрд. руб.**	t_i	t_i^2	$t_i y_i$	$\hat{y}_i$
1980	11,18	–	–	1	1	11,18	11,183
1981	12,23	109,4	1,05	2	4	24,46	12,226
1982	13,28	108,6	1,05	3	9	39,84	13,269
1983	14,31	107,7	1,03	4	16	57,24	14,312
1984	15,36	107,3	1,05	5	25	76,80	15,355
1985	16,40	106,8	1,04	6	36	98,40	16,398
В сред.	14,32	107,9	1,04	21	91	307,92	82,743
	$\sum$ 82,76						

Развитие товарооборота происходило с затухающими темпами роста и относительно стабильными абсолютными приростами.

Для установления типа развития определяющим признаком является характер изменения абсолютных приростов, так как при среднем абсолютном приросте 1,04 млрд. руб. величина их изменений незначительна $\pm 0{,}01$. Ряд можно считать с равномерным развитием, поэтому можно применить функцию $\hat{y}_t = a_0 + a_1 t$. Расчеты сведем в таблицу и определим коэффициенты уравнения выравнивания.

$$a_0 = \frac{\sum y \sum t^2 - \sum ty \sum t}{n \sum t^2 - \sum t \sum t} = \frac{82{,}76 \times 91 - 307{,}92 \times 21}{6 \times 91 - 21 \times 21} = 10{,}14 \text{ млрд. руб.}$$

$$a_1 = \frac{n \sum ty - \sum t \sum y}{n \sum t^2 - \sum t \sum t} = \frac{6 \times 307{,}92 - 21 \times 82{,}76}{6 \times 91 - 21 \times 21} = 1{,}043 \text{ млрд. руб.}$$

Составим трендовую модель динамического ряда $\hat{y} = 10{,}14 + 1{,}043t$

Параметр $a_1 = 1{,}043$ показывает, что объем возрастал в среднем на 1,043 млрд. руб. в год.

На основе полученной модели определим теоретические уровни и среднее квадратическое отклонение $s_{\hat{y}} = \sqrt{\dfrac{\sum (y_i - \hat{y}_i)^2}{n - 1}}$

Пример 3. На основе следующих отчетных данных по грузовому автотранспорту рассчитать интервальный прогноз перевозок на 1998 г. с вероятностью 0,99

t	1990	1991	1992	1993	1994	1995	1996	1997
y	360	381	401	422	443	463	485	505

Решение. Все вспомогательные расчеты выполним в таблице

Год	**Объем, тыс. т y**	**Первые разн.**	**t**	t^2	**Yt**	**Теор. уров.** $\hat{y}$	$(y - \hat{y})$	$(y - \hat{y})^2$
1990	360	–	-7	49	-2520	359,7		0,09
1991	381	21	-5	25	-1905	380,5		0,25
1992	401	20	-3	9	-1203	401,3		0,09
1993	422	21	-1	1	-422	422,1		0.01
1994	443	21	1	1	443	442,9		0,01
1995	463	20	3	9	1389	463,7		0.49
1996	485	22	5	25	2425	484,5		0,25
1997	505	20	7	49	3535	505,3		0,09
Итого	3460		0	168	1742	3460		1,28

Первые разности $\Delta y_ц$ приблизительно равны между собой, поэтому для выравнивания берем линейную модель $\hat{y}_t = a_0 + a_1 t$

Для нахождения a_0 и a_1 используется система нормальных уравнений

$$\begin{cases} \sum y = na_0 + a_1 \sum t \\ \sum yt = a_0 \sum t + a_1 \sum t^2 \end{cases}$$

Для упрощения системы показатели времени t обозначим так, чтобы $\sum t = 0$ (способ отсчета от условного нуля), тогда $\begin{cases} \sum y = na_0 \\ \sum yt = a_1 \sum t^2 \end{cases}$

Вычислим параметры уравнения

$$a_0 = \frac{\sum y}{n} = \frac{3460}{8} = 432{,}5 \text{ тыс. т.,} \qquad a_1 = \frac{\sum yt}{\sum t^2} = \frac{1742}{168} = 10{,}4 \text{ тыс. т.}$$

Составим модель динамического ряда (тренда) $\hat{y}_t = 432{,}5 + 10{,}4t$. Вычислим теоретические уровни ряда по полученной модели и точечный прогноз для 1998 г.

$$\hat{y}_9 = 432{,}5 + 10{,}4 \times 9 = 526{,}1 \text{ тыс. т}$$

Определим среднее квадратическое отклонение для полученной модели динамического ряда

$$s_{\hat{y}} = \sqrt{\frac{\sum(y - \hat{y})^2}{n - m}} = \sqrt{\frac{1{,}28}{8 - 2}} = 0{,}462 \text{ тыс. т}$$

Коэффициент доверия $t_2 = 3{,}4$ при вероятности р = 0,99, уровню значимости $\alpha = 0{,}01$ и числе степеней свободы $n - 1 = 7$.

Интервальный прогноз для 1998 г. $\hat{y}_t \pm t_2 s_{\hat{y}}$

$$Y_9 = 526{,}1 \pm 3{,}4 \times 0{,}462. \qquad 524{,}53 \text{ тыс. т} \le y_9 \le 527{,}67 \text{ тыс. т}$$

Прогнозирование на основе **постоянного абсолютного прироста** $\Delta y \cong const$

$$y_{t+l} = y_n + \overline{\Delta}y\,l$$

Определим средний абсолютный прирост $\overline{\Delta}y = \frac{\sum \Delta y}{n - 1} = \frac{144}{7} = 20{,}57 \approx 20{,}6$ тыс. т

Для 1998 г. значение $l = 1$ Прогноз на 1998 г. $y_9 = y_8 + \overline{\Delta}y \times 1 = 505 + 20{,}6 = 525{,}6$ тыс. т

7. ИНДЕКСНЫЙ МЕТОД В СТАТИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ

7.1. Понятие статистических индексов и их роль в экономике

С помощью индексов характеризуется развитие национальной экономики и ее отдельных отраслей, анализируются результаты производственной деятельности предприятий и организаций, исследуется роль отдельных факторов в формировании экономических показателей, выявляются резервы производства, определяется уровень жизни и т.д.

Индекс – это относительный показатель, характеризующий изменение величины явления (простого и сложного, состоящего из соизмеримых или несоизмеримых элементов, непосредственно не подлежащих суммированию) во времени пространстве или сравнительно с эталоном (нормативом, планом, прогнозом). В качестве **соизмерителя разнородных, несоизмеримых продуктов** можно использовать цену, себестоимость, трудоемкость.

Индексируемая величина – это значение изучаемого признака совокупности. Различают **индивидуальные индексы и общие** (сводные). **Индивидуальные индексы** – характеризуют изменение только одного элемента совокупности (выпуск автомобилей одной марки). **Общие индексы** – отражают изменение всей совокупности сложного явления (объем продукции представлен в стоимостном или трудовом выражении).

Различают индексы **количественных** показателей (объемных) и индексы **качественных** показателей (цен, себестоимости). Способы расчета индексов это **цепной и базисный**. В зависимости от методологии расчета различают **агрегатные** и **средние** индексы.

7.2. Индивидуальные индексы

Индивидуальные индексы (однотоварные индексы) характеризуют изменение отдельных единиц совокупности.

Индивидуальные индексы физического объема реализации отдельной товарной группы,

$$i_q = \frac{q_1}{q_0},$$

где q_1, q_0 - количество продаж в текущем и базисном периодах в натуральных измерителях.

Индивидуальные индексы цен реализации отдельной товарной группы

$$i_p = \frac{p_1}{p_0},$$

где p_1, p_0 - цена за единицу товара в текущем и базисном периодах.

Аналогично вычисляются и другие индивидуальные индексы (индивидуальный индекс реализации одноименного товара). Результат расчета индивидуальных отношений может выражаться в коэффициентах или в процентах. Выбор базы сравнения зависит от цели исследования.

7.3. Агрегатные формы общих индексов и их свойства

Общие индексы обладают синтетическими и аналитическими свойствами. **Синтетические свойства** состоят в соединении (агрегировании) разнородных единиц. **Аналитические свойства** состоят в определении влияния факторов на изменение явления. На основе изучения состава и роли факторов, выявления силы их действия осуществляется квалифицированное управление процессами.

Агрегатные индексы качественных показателей. Индекс потребительских цен является общим измерителем инфляции. Рассмотрим способы их построения.

Агрегатная формула общего индекса цен (р**) Г. Пааше** $I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}$.

Соизмеритель – количество реализованных товаров в текущем периоде q_1 ,

$\sum p_1 q_1$ - стоимость продажи товаров текущего периода по ценам этого периода;

$\sum p_0 q_1$ - стоимость продажи товаров текущего периода по ценам базисного периода.

Прирост товарооборота текущего периода за счет изменения цен

$$\sum \Delta_{pq(p)} = \sum p_1 q_1 - \sum p_0 q_1$$

Индекс Г. Паше – это индекс изменения цены товаров текущего периода. Применяется для определения экономического эффекта от изменения цен.

Агрегатная формула общего индекса цен (р**) Э. Ласпейреса** $I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0}$.

Соизмеритель – количестве реализации товаров в базисном периоде q_0.

$\sum p_1 q_0$ - стоимость продажи в базисном периоде по ценам текущего периода;

$\sum p_0 q_0$ - стоимость продажи в базисном периоде по ценам базисного периода;

Прирост товарооборота базисного периода за счет изменения цен

$$\sum \Delta_{pq(p)} = \sum p_1 q_0 - \sum p_0 q_0 .$$

Индекс Э. Ласпейреса показывает изменение цены товаров базисного периода. Используется для прогнозирования объема товарооборота в связи с намеченным изменением цен.

Аналогично могут быть построены **другие индексы качественных показателей**:

– себестоимости $I_z = \frac{\sum z_1 q_1}{\sum z_0 q_1}$; $I_z = \frac{\sum z_1 q_0}{\sum z_0 q_0}$, z – себестоимость единицы продукции;

– производительности труда $I_t = \frac{\sum t_1 q_0}{\sum t_0 q_0}$, t – затраты времени на производство единицы.

Агрегатные индексы количественных показателей

Общий индекс товарооборота (qp) $I_{qp} = \frac{\sum q_1 p_1}{\sum q_0 p_0}$, где $\sum q_1 p_1$ – товарооборот текущего периода, $\sum q_0 p_0$ – товарооборот базисного периода. Определяется общий результат воздействия на объем товарооборота количества реализации товаров q и их цены p.

Прирост товарооборота в текущем периоде по сравнению с базисным периодом $\sum \Delta_{qp(qp)} = \sum q_1 p_1 - \sum q_0 p_0$

Агрегатный индекс физического объема (q) продукции $I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}$.

С**оизмеритель** – цены базисного периода p_0 (используют сопоставимые фиксированные цены базисного периода).

$\sum q_1 p_0$ – стоимость товарной массы текущего периода в базисных ценах;

$\sum q_0 p_0$ – стоимость товарной массы базисного периода в базисных ценах.

Прирост товарооборота в текущем периоде по сравнению с базисным периодом в сопоставимых базисных ценах $\sum \Delta_{qp(q)} = \sum q_1 p_0 - \sum q_0 p_0$.

Индекс объема товарной массы в сопоставимых ценах используется при разработке рядов динамики.

Агрегатный индекс физического объема (q) **товарной массы** $I_q = \frac{\sum q_1 p_1}{\sum q_0 p_1}$.

С**оизмеритель** – цены текущего периода p_1.

$\sum q_1 p_1$ – стоимость товарной массы текущего периода в текущих ценах;

$\sum q_0 p_1$ – стоимость товарной массы базисного периода в текущих ценах.

Прирост суммы товарооборота в текущем периоде по сравнению с базисным при текущих ценах $\sum \Delta_{qp(q)} = \sum q_1 p_1 - \sum q_0 p_1$.

7.4. Средние индексы

Для определения индексов цен и объема товарооборота в агрегатной форме необходимы данные о количестве отдельных товаров в натуральных измерителях. Обычно известными являются товарообороты и цены по видам продукции. Мы рассмотрим формы индексов на основе этой информации и индивидуальных индексов.

Общие индексы цен.

Агрегатная формула общего индекса цен (р) **Г. Пааше** на основе соотношения $p_0 = \frac{p_1}{i_p}$ может быть преобразована к **средней гармонической** индивидуальных индексов цен:

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} \quad \rightarrow \quad I_p = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{i_p}}.$$

Исходные данные: товарооборот по видам продукции текущего периода и индивидуальные индексы цен.

Прирост товарооборота текущего периода за счет изменения цен

$$\sum \Delta_{\text{qp(p)}} = \sum q_1 p_1 - \sum \frac{q_1 p_1}{i_p}.$$

Агрегатная формула общего индекса цен (р) **Э. Ласпейреса** на основе соотношения $p_1 = p_0 i_p$ может быть преобразована к **средней арифметическая** индивидуальных индексов цен:

$$I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0} \quad \rightarrow \quad I_p = \frac{\sum i_p q_0 p_0}{\sum q_0 p_0},$$

Исходные данные: товарооборот по видам продукции базисного периода и индивидуальные индексы цен.

Прирост товарооборота базисного периода за счет изменения цен

$$\sum \Delta_{\text{qp(p)}} = \sum i_p q_0 p_0 - \sum q_0 p_0.$$

Эта формула применяется **при прогнозе** изменения товарооборота при изменении цен.

Общий индекс физического объема.

Так как учет реализации товаров ведется в стоимостном выражении и данные о количестве товаров (в натуральных измерителях) отсутствуют, то применение агрегатных индексов физического объема без преобразований невозможно.

Пусть соизмеритель – это цены базисного периода p_0 и известны **индивидуальные индексы цен** и **стоимости товарной массы** по видам продукции в текущем

$(q_1 p_1)$ и базисном $(q_0 p_0)$ периодах. Тогда, используя формулу $p_0 = \frac{p_1}{i_p}$, **индекс физического объема** можно представить в виде: $I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} \rightarrow I_q = \frac{\sum \frac{q_1 p_1}{i_p}}{\sum q_0 p_0}$,

где $\sum \frac{q_1 p_1}{i_p} -$ стоимость товарной массы текущего периода в базисных ценах;

$\sum q_0 p_0 -$ стоимость товарной массы базисного периода в базисных ценах.

Прирост товарооборота в результате изменения физического объема продажи товаров в базисных ценах: $\sum \Delta_{qp(q)} = \sum \frac{q_1 p_1}{i_p} - \sum q_0 p_0$.

Пусть соизмеритель – это цены текущего периода p_0 и известны **индивидуальные индексы физического объема** и **стоимости товаров** по видам продукции в базисном $(q_0 p_0)$ периоде. Тогда, используя формулу $i_q = \frac{q_1}{q_0}$, **индекс физического объема** товаров можно представить в виде **средней арифметической взвешенной**:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} \rightarrow I_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0},$$

где $\sum \frac{q_1 p_1}{i_p} -$ стоимость товарной массы текущего периода в базисных ценах; $\sum q_0 p_0 -$ стоимость товарной массы базисного периода в базисных ценах; $q_0 p_0 -$ веса осредняемых признаков i_q.

Прирост товарооборота в результате изменения физического объема продажи товаров в базисных ценах:

$$\sum \Delta_{qp(q)} = \sum i_q q_0 p_0 - \sum q_0 p_0.$$

Пусть соизмеритель – это цены текущего периода p_1 и известны индивидуальные индексы физического объема i_q и фактическая стоимость товаров в текущем периоде $(q_1 p_1)$. Тогда **общий индекс физического объема** определяется по формуле средней гармонической

$$I_q = \frac{\sum q_1 p_1}{\sum q_0 p_1} \rightarrow I_q = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{i_q}},$$

где $q_1 p_1$ – вес осредняемой величины i_q.

Сумма прироста стоимости продукции вследствие изменения физического объема в текущих ценах

$$\sum \Delta_{qp(q)} = \sum q_1 p_1 - \sum \frac{q_1 p_1}{i_q}.$$

7.5. Индексы с постоянными и переменными весами

При изучении динамики коммерческой деятельности приходиться производить индексные сопоставления более чем за два периода. Поэтому индексные величины могут определяться как **на постоянной**, так и **на переменной** базах сравнения, т.е. **цепные и базисные**.

Индивидуальные цепные индексы физического объема

$$i_{q1/0} = \frac{q_1}{q_0};\ i_{q2/1} = \frac{q_2}{q_1};\ i_{q3/2} = \frac{q_3}{q_2} \text{ и т.д.}$$

Индивидуальные базисные индексы физического объема

$$i_{q1/0} = \frac{q_1}{q_0};\ i_{q2/0} = \frac{q_2}{q_0};\ i_{q3/0} = \frac{q_3}{q_0} \text{ и т.д.}$$

Индивидуальные цепные индексы цен

$$i_{p1/0} = \frac{p_1}{p_0};\ i_{p2/1} = \frac{p_2}{p_1};\ i_{p3/2} = \frac{p_3}{p_2} \text{ и т.д.}$$

Индивидуальные базисные индексы цен

$$i_{p1/0} = \frac{p_1}{p_0};\ i_{p2/0} = \frac{p_2}{p_0};\ i_{p3/0} = \frac{p_3}{p_0} \text{ и т.д.}$$

Произведение цепных индивидуальных индексов равно последнему базисному индексу

$$i_{p2/0} = i_{p1/0} \times i_{p2/1};\ i_{p3/0} = i_{p1/0} \times i_{p2/1} \times i_{p3/2}.$$

Общие индексы вычисляются с переменными и постоянными весами – соизмерителями.

Цепные агрегатные индексы физического объема продукции

$$I_{q1/0} = \frac{\sum q_1 p_0}{\sum q_0 p_0};\ I_{q2/1} = \frac{\sum q_2 p_1}{\sum q_1 p_1};\ I_{q3/2} = \frac{\sum q_3 p_2}{\sum q_2 p_2} \text{ и т.д.}$$

Базисные агрегатные индексы физического объема продукции

$$I_{q1/0} = \frac{\sum q_1 p_0}{\sum q_0 p_0};\ I_{q2/0} = \frac{\sum q_2 p_0}{\sum q_0 p_0};\ I_{q3/0} = \frac{\sum q_3 p_0}{\sum q_0 p_0} \text{ и т.д.}$$

Базисный агрегатный индекс физического объема продукции может быть получен как произведение цепных агрегатных индексов при постоянных соизмерителях

$$I_{q3/0} = I_{q1/0} \times I_{q2/1} \times I_{q3/2} = \frac{\sum q_1 p_0}{\sum q_0 p_0} \times \frac{\sum q_2 p_0}{\sum q_1 p_0} \times \frac{\sum q_3 p_0}{\sum q_2 p_0}$$

Цепные агрегатные индексы цен

$$I_{p1/0} = \frac{\sum p_1 q_1}{\sum p_0 q_1};\ I_{p2/1} = \frac{\sum p_2 q_2}{\sum p_1 q_2};\ I_{p3/2} = \frac{\sum p_3 q_3}{\sum p_2 q_3} \text{ и т.д.}$$

Базисные агрегатные индексы цен

$$I_{p1/0} = \frac{\sum p_1 q_1}{\sum p_0 q_1};\ I_{p2/0} = \frac{\sum p_2 q_2}{\sum p_0 q_2};\ I_{p3/0} = \frac{\sum p_3 q_3}{\sum p_0 q_3} \text{ и т.д.}$$

7.6. Факторный анализ индексных моделей

Связь между изменениями объема товарооборота, количеством продажи товаров и уровнем их цен выражается в системе взаимосвязанных индексов товарооборота. Индекс товарооборота в фактических ценах вычисляется по следующей формуле

$$I_{qp} = I_q \times I_p$$

На основе этой формулы выявляется влияние отдельных факторов на изменение товарооборота. Зная изменение товарооборота I_{qp} и цен I_p можно определить изменение товарооборота в неизменных (сопоставимых) ценах

$$I_q = \frac{I_{qp}}{I_p}.$$

По известным индексам товарооборота в фактических ценах I_{qp} и товарооборота в сопоставимых ценах I_q определяется индекс цен

$$I_p = \frac{I_{qp}}{I_q}.$$

7.7. Индексы среднего уровня (переменного состава)

Качественные индексируемые показатели часто отображаются средними величинами (средняя цена по области, и т.д.). Общая средняя величина качественного показателя – это взвешенная средняя из частных средних которая зависит от уровня цен на отдельные товары и от удельного веса каждого товара в общем его вкладе в формировании цены. При анализе динамики среднего уровня возникает вопрос в какой мере изменение среднего уровня обусловлено действием каждого фактора в отдельности.

Индексы, отражающие изменение средних уровней за счет двух факторов: изменения данных уровней и изменения удельных весов (структуры) совокупности, называются индексами **среднего уровня**, или индексами **переменного состава**. Он состоит из двух сомножителей. Первый показывает, как изменяется средний уровень под влиянием изменения качественного показателя – **индекс фиксированного состава**. Второй показывает влияние изменения структуры и называется **индексом структурных сдвигов**.

Индекс переменного состава $I_{\overline{p}} = I_p \times I_{стр}$, где $I_{\overline{p}} = \frac{\overline{p}_1}{\overline{p}_0}$ – индекс средних цен (переменного состава); $\overline{p}_1$ и $\overline{p}_0$ – средние взвешенные цены по количеству реализованных товаров:

$$\overline{p}_1 = \frac{\sum p_1 q_1}{\sum q_1};\quad \overline{p}_0 = \frac{\sum p_0 q_0}{\sum q_0}$$

Влияние структуры реализации товаров на среднюю цену показывает **индекс структурных сдвигов**

$$I_{стр} = \frac{\Sigma p_0 q_1}{\Sigma q_1} : \frac{\Sigma p_0 q_0}{\Sigma q_0} = \frac{\Sigma p_0 q_1}{\Sigma p_0 q_0} : \frac{\Sigma q_1}{\Sigma q_0},$$

где $\frac{\Sigma p_0 q_1}{\Sigma q_1}$ – расчетная средняя цена текущего периода; $\frac{\Sigma p_0 q_0}{\Sigma q_0}$ – средняя цена базисного периода.

В абсолютном выражении это абсолютный прирост средней цены (переплата) за каждый килограмм $\Sigma\Delta_{стр} = \frac{\Sigma p_0 q_1}{\Sigma q_1} - \frac{\Sigma p_0 q_0}{\Sigma q_0}$.

Влияние изменения отдельных цен на среднюю величину цены показывает **индекс фиксированного состава** $I_p = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}$.

Пример 1. Известно количество продаж и цены в магазинах Определить:
индексы переменного состава, постоянного состава и структурных сдвигов;
величины абсолютных приростов за счет действия каждого фактора

Магазин	**Базисный период**		**Текущий Период**		**i_p**	**Удельный вес**	
	p_0	**q_0**	**p_1**	**q_1**		**базисный период**	**текущий период**
1 2 3	50 35 40	200 400 400	48 34 38	800 600 600	0,96 0,97 0,95	20 40 40	40 30 30
Итого	–	1000	–	2000	–	100	100

Определим средние цены в отчетном и базисном периодах

$$\bar{p}_1 = \frac{\Sigma p_1 q_1}{\Sigma q_1} = \frac{48 \times 800 + 34 \times 600 + 38 \times 600}{800 + 600 + 600} = \frac{81600}{2000} = 40{,}8 \text{ руб.}$$

$$\bar{p}_0 = \frac{\Sigma p_0 q_0}{\Sigma q_0} = \frac{50 \times 200 + 35 \times 400 + 40 \times 400}{200 + 400 + 400} = \frac{40000}{1000} = 40 \text{ руб.}$$

Индекс средних цен или переменного состава

$$I_{\bar{p}} = \frac{\bar{p}_1}{\bar{p}_0} = \frac{40{,}8}{40} = 1{,}02.$$

Средняя цена реализации возросла на 2%. Прирост средней цены реализации

$$\Delta_{\bar{p}} = 40{,}8 - 40 = 0{,}8 \text{ руб.}$$

Население при покупке каждого килограмма продукта переплачивало по 0,8 руб.

Индекс структурных сдвигов

$$I_{стр} = \frac{\Sigma p_0 q_1}{\Sigma q_1} : \frac{\Sigma p_0 q_0}{\Sigma q_0} = \frac{50 \times 800 + 35 \times 600 + 40 \times 600}{800 + 600 + 600} : 40 = \frac{85000}{2000} : 40 = \frac{42{,}5}{40} = 1{,}0625 \cdot$$

Структурные сдвиги (изменение долей продаж) в реализации объема продукции вызвали повышение средней цены на 6,25%.

$$\Delta_{стр} = 42{,}5 - 40 = 2{,}5 \text{ руб.}$$

Переплата населением на каждый килограмм продукции составила 2,5 руб.
Индекс постоянного состава

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{81600}{85000} = 0{,}962$$

Индекс показывает снижение цены на 3,8%. За счет снижения цен произошла экономия в общей сумме

$$\Delta_p = 81600 - 85000 = -3400 \text{ руб.}$$

Экономия на каждый килограмм составила $\frac{3400}{2000} = 1{,}7$ руб.

Рассчитаем индекс средних цен другим способом (мультипликативная модель)

$$I_{\bar{p}} = I_{стр} \times I_p = 1{,}0625 \times 0{,}962 = 1{,}02$$

Величина изменения средней цены продаж составила

$$\Delta_{\bar{p}} = \Delta_{стр} - \Delta_p = 2{,}5 - 1{,}7 = 0{,}8 \text{ руб.}$$

Пример 2. Имеются данные о реализации товара. Определить индивидуальные и общие индексы; абсолютный прирост за счет действия отдельных факторов.

Товар	**Ед. изм.**	I период		**II период**		**Индивидуальный индекс**	
		цена за 1 ед., p_0	**кол-во, q_0**	**цена за 1 ед., p_1**	**кол-во, q_1**	$i_p = \frac{p_1}{p_0}$	$i_q = \frac{q_1}{q_0}$
А	т	20	7500	25	9500	1,25	1,27
Б	м	30	2000	30	2500	1,0	1,25
В	шт.	15	1000	10	1500	0,67	1,5

Решение. Результаты расчета индивидуальных индексов цен и физического объема находятся в таблице.

Общий индекс цен

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{25 \times 9500 + 30 \times 2500 + 10 \times 1500}{20 \times 9500 + 30 \times 2500 + 15 \times 1500} = \frac{327500}{287500} = 1{,}139 \text{ или } 113{,}9\%$$

По данному ассортименту товаров в целом цены повысились на 13,9%.

Абсолютный прирост товарооборота за счет фактора изменения цен

$$\sum \Delta_{qp(p)} = \sum p_1 q_1 - \sum p_0 q_1 = 327500 - 287500 = 40000 \text{ руб.}$$

Повышение цен на 13,9% обусловило увеличение объема товарооборота в текущем периоде на 40 тыс. руб.

Если (-40), то перерасход денежных средств населением при покупке товаров по ценам повышенным на 13,9%

Общий индекс цен расчетный

$$I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0} = \frac{25 \times 7500 + 30 \times 2000 + 10 \times 1000}{20 \times 7500 + 30 \times 2000 + 15 \times 1000} = \frac{257500}{225000} = 1{,}144 \text{ или } 114{,}4\%$$

По ассортименту в целом повышение цены составило в среднем 14,4%.

Сумма прироста

$$\sum \Delta_{qp(p)} = \sum p_1 q_0 - \sum p_0 q_0 = 257500 - 225000 = 32500 \text{ руб.}$$

Повышение цен в текущем периоде в среднем на 14,4% обуславливает увеличение объема товарооборота на 32,5 тыс. руб.

Общий индекс физического объема

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{9500 \times 20 + 2500 \times 30 + 1500 \times 15}{7500 \times 20 + 2000 \times 30 + 1000 \times 15} = \frac{287500}{225000} = 1{,}278 \text{ или } 127{,}8\%.$$

Прирост физического объема реализации в текущем периоде составил в среднем 27,8%.

Сумма прироста товарооборота

$$\sum \Delta_{qp(q)} = 287500 - 225000 = 62500 \text{ руб.}$$

В результате изменения физического объема реализации товаров в текущем периоде получен прирост объема товарооборота в сопоставимых ценах на 62,5 тыс. руб.

$$I_q = \frac{\sum q_1 p_1}{\sum q_0 p_1} = \frac{9500\times 25 + 2500\times 30 + 1500\times 10}{7500\times 25 + 2000\times 30 + 1000\times 10} = \frac{327500}{257500} = 1{,}272 \text{ или } 127{,}2\%$$

По данному ассортименту реализованных в текущем периоде товаров прирост физического объема товарооборота составил 27,2%.

Абсолютный прирост суммы товарооборота в результате изменения физического объема продажи товаров составил

$$\sum \Delta_{qp(q)} = 327500 - 257500 = 70000 \text{ руб.}$$

При этом за счет роста физического объема продажи товаров на 27,8% этот прирост составил 62,5 тыс. руб., а повышение цен в среднем на 13,9% увеличило объем товарооборота на 40 тыс. руб.

Общий индекс товарооборота в текущих ценах вырос на 45,5%.

$$I_{qp} = \frac{327500}{225000} = 1{,}455 \text{ или } 145{,}5\%$$

Прирост фактического объема товарооборота в текущем периоде

$$\sum \Delta_{qp(q)} = 327500 - 225000 = 102500 \text{ руб.}$$

Пример 3. Данные о продаже товаров в магазине. Определить общий индекс цен.

Товар	**Продажа в ценах соответствующего периода**		**Изменение цен в текущем периоде по сравнению с базовым, %**	Расчет	
	базовый период q_0p_0	**текущий период q_1p_1**		$i_p = \frac{p_1}{p_0}$	$\frac{q_1p_1}{i_p}$
А	153,5	185,0	-4	0,96	192,71
Б	245,0	260,6	+10	1,1	236,91
В	21,5	29,4	без изм.	1,0	29,4
Итого	420,0	475,0	–	–	459,02

Индивидуальные (однотоварные) индексы цен $i_p = \frac{p_1}{p_0}$

$$i_p^А = \frac{100-4}{100} = 0{,}96; \quad i_p^Б = \frac{100+10}{100} = 1{,}1$$

По каждому товару определим стоимость продажи товара в текущем периоде по ценам базисного

$$p_0 q_1 = \frac{q_1 p_1}{i_p}, \qquad А = \frac{185{,}0}{0{,}96} = 192{,}71; \quad Б = \frac{260{,}6}{1{,}1} = 236{,}91; \quad В = \frac{29{,}4}{1{,}0} = 29{,}4 .$$

Общий индекс цен

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}} = \frac{475{,}0}{459{,}02} = 1{,}035 \text{ или } 103{,}5\%,$$

то есть по данному ассортименту цены повышены на 3,5%.

Прирост товарооборота за счет изменения цен

$$\sum \Delta_{qp(p)} = \sum q_1 p_1 - \sum \frac{q_1 p_1}{i_p} = 475{,}0 - 459{,}02 = 15{,}98 \text{ тыс. руб.}$$

Определим общий индекс физического объема товарооборота в сопоставимых (базисных) ценах

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{\sum \frac{q_1 p_1}{i_p}}{\sum q_0 p_0} = \frac{459{,}02}{420} = 1{,}093 \text{ или } 109{,}3\%,$$

то есть физический объем продажи товаров увеличился в текущем периоде на 9,3%. Прирост суммы товарооборота в текущем периоде в результате изменения физического объема продажи товаров

$$\sum \Delta_{qp(q)} = \sum \frac{q_1 p_1}{i_p} - \sum q_0 p_0 = 459{,}02 - 420 = 39{,}02 \text{ тыс. руб.}$$

Общий прирост товарооборота в текущем периоде

$$\sum \Delta_{qp(qp)} = \sum q_1 p_1 - \sum q_0 p_0 = 475 - 420 = 55 \text{ тыс. руб.}$$

Задача 1. Известна выработка продукции на предприятии Рассчитать индивидуальные индексы физического объема; агрегатные индексы физического объема продукции и стоимости; абсолютные приросты.

Продукция, ед. изм.	**Выработка продукции, тыс.**		**Цена за единицу, руб.**		$i_q = \frac{q_1}{q_0}$
	Q_0	**q_1**	**p_0**	**p_1**	
А, кг	500	500	150	140	1,0
Б, м	200	240	100	110	1,2
В, шт.	600	420	250	300	0,7

Решение. Результат расчета индивидуальных индексов физического объема записаны в таблицу. Общий индекс физического объема продукции

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{500\times150 + 240\times100 + 420\times250}{500\times150 + 200\times100 + 600\times250} = \frac{204000}{245000} = 0{,}833 \text{ или } 83{,}3\%$$

Физический объем всей продукции в отчетном периоде составляет 83,3% от его уровня в базисном периоде, он снизился на 16,7%.

Абсолютный прирост (снижение) в неизменных ценах

$$\sum \Delta_{pq(q)} = \sum p_1 q_0 - \sum p_0 q_0 = 204000 - 245000 = -41000$$

В отчетном периоде стоимость продукции уменьшилась на 41 млн. руб. (только за счет снижения на 16,7% физического объема производства продукции).

Агрегатный индекс стоимости продукции или товарооборота

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{500\times140 + 240\times110 + 420\times300}{245000} = \frac{222400}{245000} = 0{,}908 \text{ или } 90{,}8\%$$

Общий выпуск продукции (стоимость) в фактических ценах в текущем периоде составил 90,8% ее выпуска в базисном периоде или с учетом изменения цен снизился на 9,2%, то есть выпуск продукции уменьшился в абсолютном выражении на

$$\sum p_1 q_1 - \sum p_0 q_0 = 222400 - 245000 = -22600 = -22{,}6 \text{ тыс. руб.}$$

Задача 2. Имеются данные о средней заработной плате работников и число работников организаций по трем отраслям.

№ п/п	**Отрасль экономики**	**Заработная плата, руб.**		**Число работников, чел.**	
		X_0	**X_1**	**T_0**	**T_1**
1	Здравоохранение	600	700	2400	1600
2	Образование	550	620	2100	2000
3	Культура и искусство	510	590	1500	1400

Определить индекс заработной платы переменного состава, постоянного состава и структурных сдвигов.

Решение. 1. Определим среднюю заработную плату работников

базисный период

$$\overline{X}_0 = \frac{\sum X_0 T_0}{\sum T_0} = \frac{600\times2400 + 500\times2100 + 510\times1500}{6000} = \frac{3360000}{6000} = 560 \text{ руб.}$$ отчетный период

$$\overline{X}_1 = \frac{\sum X_1T_1}{\sum T_1} = \frac{700\times1600 + 620\times2000 + 590\times1400}{5000} = \frac{3186000}{5000} = 637{,}2 \text{ руб.}$$

2. Индекс заработной платы переменного состава

$$I_{\overline{X}} = \frac{\overline{X}_1}{\overline{X}_0} = \frac{\sum X_1T_1}{\sum T_1} : \frac{\sum X_0T_0}{\sum T_0} = \frac{637{,}2}{560} = 1{,}138 \text{ или } 113{,}8\%$$

Заработная плата выросла на 13,8%. Абсолютный прирост составил
637,2 – 560 = 77,2 руб.
Изменение средней заработной платы происходило под влиянием двух факторов: уровня заработной платы и числа работников.

3. Индекс заработной платы постоянного состава

$$I_X = \frac{\sum X_1T_1}{\sum T_1} : \frac{\sum X_0T_1}{\sum T_1} = 637{,}2 : \frac{600\times1600 + 550\times2000 + 510\times1400}{5000} = = \frac{637{,}2}{554{,}8} = 1{,}149 \text{ или } 114{,}9\%$$

Средняя заработная плата работников увеличилась на 14,9% за счет увеличения заработной платы. Абсолютный прирост средней заработной платы составил

637,2 – 554,8 = 82,4 руб.

4. Индекс структурных сдвигов

$$I_{стр} = \frac{\sum X_0T_1}{\sum T_1} : \frac{\sum X_0T_0}{\sum T_0} = \frac{2774000}{5000} : \frac{3366000}{6000} = \frac{554{,}8}{560} = 0{,}9907 \text{ или } 99{,}07\%$$

Увеличение доли работников с низкой заработной платой в общей численности привело к снижению средней заработной платы на 0,03%. Абсолютное снижение составило 554,8 – 560 = –5,2 руб.

Задача 3. Имеются данные выпуска продукции по заводу строительных пластмасс.

Вид продукции	**Выпуск продукции в I квартале, млн. руб.**	**Изменение объема производства в натуральном выражении во II квартале**	**Индивидуальные индексы**
Пленка	30	+10	1,1
пеноплен	25	-10	0,9
линолеум	40	-25	0,75

Определим сводную оценку изменения объема производства продукции (в натуральном выражении)

Решение. Индекс физического объема продукции

$$I_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0} = \frac{1{,}1\times30 + 0{,}9\times25 + 0{,}75\times40}{30 + 25 + 40} = \frac{85{,}5}{95{,}0} = 0{,}9$$

Объем производства в натуральном выражении во 2-ом квартале по сравнению с 1-ым уменьшился на 10%, что составило 9,5тыс.руб.

Задача 4. Имеются данные о продаже товаров в магазине

Товар, ед. изм.	**Продано в отчетном периоде p_1q_1, тыс. руб.**	**Изменение цен на товары, %**	**Индивидуальные индексы цен i_p**
Туфли муж., пары	186	+3	1,03
Костюмы, шт.	214	+6	1,06
Итого	400	–	–

Определить индивидуальные и общий индексы цен.

Решение. Результаты расчета индивидуальных индексов представлены в виде коэффициентов и находятся в таблице. Среднегармонический индекс цен

$$I_p = \frac{\sum p_1q_1}{\sum \frac{p_1q_1}{i_p}} = \frac{186 + 214}{\frac{186}{1{,}03} + \frac{214}{1{,}06}} = \frac{400}{382{,}47} = 1{,}046 \text{ или } 104{,}6\%.$$

Объем реализации за счет изменения цены повысились в среднем на 4,6%.

8. СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ СВЯЗИ МЕЖДУ ЯВЛЕНИЯМИ

Важной задачей статистики является изучение статистических закономерностей, знание которых дает основу для предсказания и управления социально экономическими процессами. Перечислим некоторые виды связи.

Балансовая связь – характеризует зависимость между источниками формирования ресурсов и их использованием

$$О_н + П = В + О_к ,$$

где: $О_н$, $О_к$ – остатки на начало и конец; П, В – поступление и выбытие.

Компонентные связи: изменение статистического показателя определяется изменением компонентов, входящих в этот показатель как сомножители (мультипликативная модель), используются в индексном методе

$$I_{pq} = I_p \times I_q$$

Причинная форма связи – это порождение одного явления другим. Признак, характеризующий следствие, называется результативным, а причину – факторным. Выделяют два вида причинно следственных связей: функциональную и статистическую.

Функциональную связь можно представить уравнением: $y = f(x)$, где y – результативный признак, x – факторный, $f(x)$ – известная функция.

Статистическую связь можно представить в виде: $y = f(x)+\varepsilon$, где $f(x)$ – известная функция, а ε – часть результативного признака, определяемая неучтенными и неконтролируемыми признаками.

8.1. Параметрические методы изучения связи

Корреляционно-регрессионный анализ позволяет выбрать вид модели, оценить ее параметры, измерить тесноту связи, определить наиболее влияющие факторы на результативный признак.

Линейная форма связи и оценка ее параметров. При линейной форме связи зависимость результативного признака y от факторного показателя x определяется уравнением регрессии $y = a_0 + a_1x$. Оценивание неизвестных параметров (a_0, a_1) производится методом наименьших квадратов (МНК) по исходным данным (y_i, x_i, i=1,2,…n). МНК дает систему нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum x = \sum y \\ a_0 \sum x + a_1 \sum x^2 = \sum xy \end{cases},$$

решая которые находятся неизвестные парметры

$$a_0 = \frac{\sum y \sum x^2 - \sum xy \sum x}{n \sum x^2 - \sum x \sum x}; \quad a_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - \sum x \sum x}.$$

Подставляя в общее уравнение найденные параметры, получим **уравнение регрессии**

$$\hat{y} = a_0 + a_1 x .$$

Проверка адекватности регрессионной модели.

Введем обозначения

Среднеквадратическое отклонение результативного признака y_i от выровненных $\hat{y}_i$. $s^2_{ост}$ – остаточная дисперсия	$s_{ост} = \sqrt{\dfrac{\sum(\hat{y}_i - y_i)^2}{n}}$
Среднеквадратическое отклонение факторного признака x_i от средней $\bar{x}$	$\sigma_x = \sqrt{\dfrac{\sum(x_i - \bar{x})^2}{n}}$
Общая дисперсия	$s^2_y = \dfrac{\sum(y_i - \bar{y})^2}{n}$
Среднеквадратическое отклонение модельных значений от средней.	$s_{\hat{y}} = \sqrt{\dfrac{\sum(\hat{y}_i - \bar{y})^2}{n}}$
Межгрупповая дисперсия измеряет вариацию групповых средних $\bar{x}_i$ вокруг общей средней. Она измеряет вариацию, обусловленную признаком, положенным в основу группировки.	$d^2 = \dfrac{\sum(\bar{x}_i - \bar{x})^2 f_i}{\sum f_i}$

Для проверки з**начимости коэффициентов** линейной регрессии $\hat{y} = a_0 + a_1 x$ при n<30 используют t-критерий Стьюдента. Для этого вычисляют расчетные значения t – критерия

$$\text{для параметра } a_0: \; t_{a_0} = |a_0| \frac{\sqrt{n-2}}{s_{ост}}; \qquad \text{для параметра } a_1: \; t_{a_1} = |a_1| \frac{s_x \sqrt{n-2}}{s_{ост}}.$$

Полученные значения t_{a_0}, t_{a_1} сравнивают с критическими $t_{кр}$, которые определяют по таблице Стьюдента при заданном уровне значимости α (обычно α=0.05) и числом степеней свободы ν=n-m=n-2. Параметр признается значимым, если $t_{расч} > t_{кр}$.

Теснота корреляционной связи между x и y может быть измерена **империческим корреляционным отношением** $h_э = \sqrt{d^2 / s^2_y}$ ($0 \le h_э \le 1$). Чем ближе оно к 1, тем теснее связь. При $h_э$=0 связи нет.

Теснота корреляционной связи между x и y для при заданной зависимости определяется **индексом корреляции**

$$R = \sqrt{\frac{s^2_{\hat{y}}}{s^2_y}} = \sqrt{1 - \frac{s^2_{ост}}{s^2_y}} = \sqrt{1 - \frac{\sum(y - \hat{y})^2}{\sum(y - \bar{y})^2}} \quad (0 \le R \le 1).$$

Чем ближе R к 1, тем теснее связь. При R=0 связи нет. Величину R^2 называют **коэффициентом детерминации**. Коэффициент детерминации характеризует какая часть общей вариации y объясняется изучаемым фактором x.

Показателем тесноты линейной связи является **линейный коэффициент корреляции**

$$r=\frac{\overline{xy}-\bar{x}\,\bar{y}}{s_x s_y}=\frac{\sum(x-\bar{x})(y-\bar{y})}{n s_x s_y}=\frac{\left(\sum xy-\frac{\sum x\sum y}{n}\right)}{\sqrt{\left[\sum x^2-\frac{(\sum x)^2}{n}\right]\left[\sum y^2-\frac{(\sum y)^2}{n}\right]}}\quad (-1\le r\le 1).$$

Величину r2 называют линейным коэффициентом детерминации. Для оценки **значимости коэффициента корреляции** r используют t-критерий Стьюдента. Для этого вычисляют расчетные значения t – критерия

$$t_{расч}=|r|\sqrt{\frac{n-2}{1-r^2}},$$

где n-2 – число степеней свободы при заданном уровне значимости α (обычно α=0.05). Расчетное значение $t_{расч}$ сравнивают с табличными значениями t-критерия $t_{кр}$, которые определяют по таблице Стьюдента при заданном уровне значимости α (обычно α=0.05) и числом степеней свободы ν=n-m=n-2. Параметр r признается значимым, если $t_{расч} > t_{кр}$.

8.2. Непараметрические методы оценки корреляции связи

1) При исследовании степени **тесноты связи между качественными признаками**, каждый из которых представлен в виде альтернативных признаков, применяют **коэффициент ассоциации Д. Юла (К$_а$)**, и **коэффициент контингенции К. Пирсона (К$_к$).**

Расчетная таблица

Признак (работа)	**Да (муж.)**	**Нет (жен.)**
Да	A	B
Нет	C	D

$$K_a=\frac{ad-bc}{ad+bc},$$

$$K_к=\frac{ad-bc}{\sqrt{(a+b)(b+d)(a+c)(c+d)}}$$

где: a, b, c, d – числа таблицы.

Значения коэффициентов лежат в интервале $-1 < К_к < 1$ (если показатель отсутствует, то заменяют его единицей для $К_к$)

2) Для определения **тесноты связи как между количественными**, так и между качественными признаками (если их можно проранжировать или упорядочить) можно использовать **коэффициент ранговой корреляции Спирмена**

$$P=1-\frac{6\sum d_i^2}{N\left(N^2-1\right)}$$

где: $d_i^2=(R_x^i\text{-}R_y^i)^2$ – квадрат разности рангов связанных величин x и y; R_x^i (R_y^i) – ранги (номера) величин x (y) в упорядоченном ряду по возрастанию x (y); N – число наблюдений (число пар рангов). Сила связи определяется по шкале Чеддока.

Знач. Р	0.1-0.3	0.3-0.5	0.5-0.7	0.7-0.9	0.9-1
Теснота связи	слабая	умеренная	заметная	высокая	Весьма высокая

Пример 1. Распределение работников торговли по полу и оценка содержания работы. Определить тесноту связи отношения к работе мужчин и женщин.

Работа	Мужчины	Женщины	Всего
Интересная	300 (a)	201 (b)	501 (a+b)
Неинтересная	130 (c)	252 (d)	382 (c+d)
Итого	430 (a+c)	453 (b+d)	883 (a+b+c+d)

Коэффициент ассоциации $K_a = \frac{ad - bc}{ad + bc} = \frac{300 \times 252 - 201 \times 130}{300 \times 252 + 201 \times 130} = 0{,}486$.

Средний размер связи по критерию шкалы Чеддока.

Пример 2. Определить связи между обеспеченностью товарной продукцией ряда предприятий и накладными расходами по реализации

Решение. а) Упорядочиваем ряды x, y по возрастанию.

б) Присваиваем порядковые номера (ранги) ранжированным рядам x, y.

в) Сравниваем x, y с ранжированным и записываем их ранговые номера R_x, R_y.

Обесп. тов. прод., млн. руб., x	Накл. Расх. по реал., y	Ранжирование				Сравнивание		Разность рангов d_i	d_i^2
		x	ранг $R_x x$	y	ранг $R_y y$	R_x	R_y		
12,0	462	11,0	1	462	1	2	1	1	1
18,8	939	12,0	2	506	2	5	6	-1	1
11,0	506	15,4	3	765	3	1	2	-1	1
29,0	1108	17,5	4	804	4	9	9	0	0
17,5	872	18,8	5	872	5	4	5	-1	1
23,4	765	20,7	6	939	6	7	3	4	16
35,6	1368	23,4	7	998	7	10	10	0	0
15,4	1002	26,1	8	1002	8	3	8	-5	25
26,1	998	29,0	9	1108	9	8	7	1	1
20,7	804	35,0	10	1368	10	6	4	2	4
									50

г) Коэффициент корреляции Спирмена $P = 1 - \frac{6\sum d_i^2}{N\left(N^2 - 1\right)} = 1 - \frac{6 \times 50}{10(100 - 1)} = 0{,}7$

Связь заметная по шкале Чеддока.

СПИСОК ЛИТЕРАТУРЫ

1. М.Р.Ефимова, Е.В.Петрова, В.Н.Румянцев. Общая теория статистики. 2-е издание. М.:ИНФРА-М, 2000.
2. Статистика. Учебное пособие. Под редакцией М.Р.Ефимовой. М.:ИНФРА-М, 2000.
3. В.М.Гусаров. Статистика. М.:ЮНИТИ, 2001.
4. Теория статистики. Под редакцией проф.Р.А.Шмойловой. М.:Финансы и статистика, 2000.
5. Практикум по теории статистики. Под редакцией проф.Р.А.Шмойловой. М.:Финансы и статистика, 2000.

ОГЛАВЛЕНИЕ