КЫРГЫЗСКО-РОССИЙСКИЙ СЛАВЯНСКИЙ УНИВЕРСИТЕТ

В.И. Марук

ТЕОРИЯ СТАТИСТИКИ

Конспекты лекций и решение типовых задач

Бишкек 2008

Марук В.И.

ТЕОРИЯ СТАТИСТИКИ: Конспекты лекций и решение типовых задач. – Бишкек: КРСУ, 2008. – 109 с.

Задача предлагаемого учебно-методического пособия – ознакомить студентов с основными методами теории статистики и оказать им практическую помощь при выполнении контрольной работы по этой дисциплине.

Предназначено для студентов-заочников экономических специальностей.

Рекомендовано к изданию
кафедрой экономической теории КРСУ

Глава 1

МЕТОД ГРУППИРОВОК

1.1. Статистическая сводка и группировка

Статистическая сводка – это систематизация первичных материалов наблюдения, позволяющая охарактеризовать совокупность единиц обобщающими показателями. Задача сводки заключается в том, чтобы подытожить и систематизировать результаты наблюдения отдельных единиц совокупности и на этой основе выявить характерные черты и закономерности изучаемого явления. Сводка позволяет перейти от характеристики индивидуального значения признака отдельных единиц совокупности к характеристике общего его проявления в массе единиц.

В зависимости от способа разработки материала сводка бывает простой и сложной. Простая сводка – это операция по определению общего количества единиц совокупности или общей величины изучаемого показателя. Например, для определения общей численности студентов факультета вуза следует сложить данные о численности студентов на каждом курсе факультета.

Сложная сводка – это распределение единиц совокупности на группы и определение числа единиц или величины изучаемого показателя в каждой группе и по всем группам вместе. Например, распределить всех студентов факультета по успеваемости на четыре группы и подсчитать численность студентов в каждой группе.

Распределение единиц совокупности на однородные группы по существенным признакам называется статистической группировкой. Метод группировок широко применяется в статистике, так как является основным способом научной обработки данных статистического наблюдения. С помощью этого метода решаются следующие задачи статистического исследования: 1) выделение социально-экономических типов явлений; 2) изучение структуры явления и происходящих в нем структурных сдвигов; 3) выявление связей и зависимостей между признаками внутри совокупности.

Для решения любой из этих задач необходимо выбрать группировочный признак, по которому производится распределение единиц совокупности на группы. Группировочный признак может быть количественным, т.е. иметь числовое выражение (возраст, стаж, зарплата), и атрибутивным, т.е. характеризовать качество явления и не иметь числового выражения (пол, национальность, партийность). Количественные признаки подразделяются на дискретные, которые выражаются только целыми числами, и непрерывные, которые могут быть выражены как целыми, так и дробными числами.

Выбор группировочного признака зависит от сущности изучаемого явления и от цели статистического исследования.

1.2. Виды группировок

В зависимости от задач, которые решаются на основе группировок, статистические группировки подразделяются на три вида: типологические, структурные и аналитические.

Типологической называется группировка, на основе которой из качественно разнородной совокупности выделяются однородные группы единиц, представляющие социально-экономические типы явлений. Примером типологической группировки может служить распределение населения по социальным группам. Наглядное представление о типологической группировке дает табл. 1.2.1.

Таблица 1.2.1

Группировка хозяйствующих субъектов КР
по формам собственности, 2006 г.

Группы хозяйствующих субъектов	Количество субъектов, % к итогу
Государственной собственности	0,8
Муниципальной собственности	1,5
Частной собственности	97,7
Всего	100,0

Структурной называется группировка, на основе которой характеризуется распределение единиц однородной совокупности по какому-либо варьирующему признаку с целью изучения внутреннего строения явления. Примерами структурной группировки являются распределение населения страны по полу или возрасту; распределение работников предприятия по производственному стажу или размеру заработной платы. К структурным относятся различные территориальные группировки (табл. 1.2.2).

Таблица 1.2.2

Территориальное деление КР, 2006 г.

Область	Территория, тыс. км^2
Баткенская	17,0
Жалалабатская	33,7
Ыссыккольская	43,1
Нарынская	45,2
Ошская	29,2
Таласская	11,4
Чуйская	20,3
Итого	199,9

Сравнение структурных группировок какого-либо явления, относящихся к разным периодам или моментам времени, позволяет выявить произошедшие структурные изменения, которые называются структурными сдвигами.

Аналитической называется группировка, на основе которой выявляется взаимосвязь между признаками. Взаимосвязанные признаки делятся на факторные и результативные. Факторным называется признак, изменение которого является причиной изменения другого признака. Признак, изменяющийся под влиянием факторного признака, называется результативным. Если с ростом факторного признака результативный признак систематически увеличивается или уменьшается, то, следовательно, между ними имеется зависимость. Пример аналитической группировки представлен в табл. 1.2.3.

Таблица 1.2.3

Влияние производственного стажа рабочих
на степень выполнения ими нормы выработки

Стаж работы, лет	Численность рабочих, чел.	Выполнение нормы выработки, %
До 1	11	80
1–5	27	102
5–10	64	108
10 и выше	18	114
Итого	120	–

Из таблицы видно, что между признаками существует прямая зависимость: с ростом стажа работы степень выполнения рабочими нормы выработки увеличивается.

В зависимости от вида группировочного признака различаются группировки атрибутивные (например, по признаку пола, национальности, партийности) и количественные (например, по признаку возраста, стажа работы, заработной платы). Особая разновидность атрибутивных группировок, которая содержит подробную номенклатуру групп и подгрупп, утвержденную в качестве стандарта, называется классификацией (например, классификации товаров, профессий, отраслей экономики).

В зависимости от количества группировочных признаков группировки подразделяются на простые и сложные. Простые группировки построены по одному, а сложные – по нескольким признакам. Группировка, построенная по двум или более признакам, взятым в сочетании, называется комбинационной. Для построения комбинационной группировки группы единиц совокупности, образованные по одному признаку, делятся затем на подгруппы по другому признаку (например, каждая группа распределения рабочих предприятия по заработной плате делится на подгруппы по полу).

В тех случаях, когда приходится пользоваться существующей группировкой, которая не соответствует требованиям исследователя, применяется метод вторичной группировки. Вторичной группировкой называется операция по образованию новых групп на основании уже имеющейся группировки.

Вторичную группировку производят для решения следующих задач: 1) выявление характера распределения единиц совокупности; 2) приведение нескольких группировок с различными интервалами к сопоставимому виду; 3) образование на основе группировок по количественному признаку качественно однородных типичных групп.

Образование новых групп на основе имеющихся можно осуществить двумя способами: 1) изменением (обычно укрупнением) первоначальных интервалов; 2) закреплением за каждой группой определенной доли единиц совокупности. (Методика построения вторичной группировки приведена в конце главы.)

1.3. Число групп, размеры и границы интервалов

Число групп, которые необходимо образовать в процессе построения группировки, зависит от задачи исследования, вида группировочного признака, степени его вариации и количества единиц совокупности.

При группировках по атрибутивному признаку число групп определяется, как правило, количеством разновидностей этого признака. Так, при группировке работников предприятия по полу можно образовать только две группы, а при группировке по национальности – столько групп, сколько национальностей работников.

При группировках по количественному дискретному признаку с незначительной вариацией число групп определяется числом вариантов значений признака (например, группировка рабочих по разрядам).

Если группировочный признак непрерывный или если дискретный признак варьирует в широких пределах, то число групп зависит от количества единиц исследуемой совокупности и от вариации признака. Теория статистики дает следующие общие рекомендации. Число групп не должно быть слишком большим, чтобы сгруппированный материал был легко обозримым, чтобы в каждую группу попало достаточное количество единиц и тем самым было исключено получение случайных характеристик отдельных групп. С другой стороны, число групп не должно быть слишком малым, так как в этом случае не удается выделить наиболее типичные группы и охарактеризовать зависимость результативного признака от группировочного. Чем больше вариация группировочного признака, тем больше должно быть групп. Определить рекомендуемое число групп в зависимости от количества единиц совокупности можно по формуле Стерджесса:

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

где n – число групп; N – число единиц совокупности.

Если число единиц совокупности велико и у большинства из них разная величина группировочного признака, то для построения группировки с оптимальным числом групп, приходится объединять близкие по значению признака единицы в одну группу. Для каждой группы устанавливаются максимальное и минимальное значение признака, т.е. устанавливаются интервалы. Например, при группировке рабочих предприятия по возрасту установлены следующие интервалы, лет:

до 20; 20–30; 30–40; 40–50; 50–60; 60 и более.

Интервалы бывают открытыми и закрытыми. Закрытыми называются интервалы, которые имеют верхнюю и нижнюю границы. Открытыми называются интервалы, имеющие только верхнюю или только нижнюю границу. Для обозначения открытых интервалов применяются слова “до”, “менее”, “свыше”, “более”. В приведенном выше примере первая и последняя группы с открытыми интервалами, а остальные – с закрытыми.

Разность между наибольшим и наименьшим значениями признака в каждой группе называется величиной интервала. Интервалы групп бывают равными и неравными. Равные интервалы применяются, если признак изменяется в сравнительно узких границах и распределение единиц более или менее равномерно. Величина равного интервала определяется по формуле:

$$h = \frac{x_{\max} - x_{\min}}{n},$$

где $x_{\max}$ и $x_{\min}$ – максимальное и минимальное значения признака в совокупности единиц; n – число групп.

Рассмотрим определение величины и границ интервалов на примере.

Требуется произвести группировку рабочих предприятия по производственному стажу, образовав 5 групп с равными интервалами. Стаж рабочих предприятия колеблется от 3 до 43 лет.

Величина интервала в этом случае будет равна:

$$h = \frac{43 - 3}{5} = 8 \text{ лет}.$$

Прибавив к минимальному значению признака исчисленную величину интервала, получим верхнюю границу первой группы: 3 + 8 = 11 лет. Примем нижнюю границу второй группы равной верхней границе первой группы. Прибавив затем величину интервала к нижней границе второй группы, найдем ее верхнюю границу: 11 + 8 = 19 лет. Аналогично определяются границы остальных групп. Окончательно получим следующие группы рабочих по стажу, лет:

3–11; 11–19; 19–27; 27–35; 35–43.

При распределении рабочих по этим группам возникает вопрос, куда отнести рабочего со стажем, который указан в верхней и нижней границах двух смежных групп. Для устранения этой неопределенности, необходимо условиться, что верхняя граница каждой группы считается “включительно” или, наоборот, “исключительно”. Другой вариант решения этого вопроса заключается в том, чтобы нижнюю границу каждой группы, начиная со второй, увеличивать по сравнению с верхней границей предыдущей группы на единицу, если группировочный признак дискретный, или на десятую (сотую или тысячную) долю единицы, если признак непрерывный.

Когда вариация группировочного признака велика и тенденции изменения признака в низших и высших группах разные, применяются неравные интервалы. Неравные интервалы бывают прогрессивно возрастающие и прогрессивно убывающие. Примером применения неравных интервалов может быть группировка промышленных предприятий страны по численности работников, чел.: до 100, 100–300; 300–700; 700–1500; 1500 и более.

Если задача группировки заключается в том, чтобы отобразить качественное своеобразие отдельных групп, то применяются специализированные интервалы. Границы специализированных интервалов устанавливаются в зависимости от того, где происходит переход от одного качества к другому.

1.4. Ряды распределения

Рядом распределения называется упорядоченное распределение единиц совокупности на группы по какому-либо одному признаку. Ряд распределения позволяет судить о составе изучаемого явления, о закономерности распределения единиц совокупности и ее однородности. На основе ряда распределения исчисляются различные обобщающие показатели (средняя, мода, медиана, дисперсия и др.).

Ряды распределения подразделяются на атрибутивные, образованные по атрибутивному признаку, и вариационные, образованные по количественному признаку. Примером атрибутивного ряда является распределение работников предприятия по полу, национальности, образованию. Распределение этих же работников по возрасту, стажу работы или производительности труда – это примеры вариационных рядов. Наглядным примером вариационного ряда может служить распределение студентов по результатам экзамена (табл. 1.4.1).

Таблица 1.4.1

Экзаменационный балл	2	3	4	5
Количество студентов, чел.	8	12	3	2

Вариационный ряд распределения состоит из двух элементов: вариантов и частот. Вариантами называются отдельные значения изменяющегося признака. В табл. 1.4.1 вариантами являются оценки 2, 3, 4 и 5.

Частотами называются числа, показывающие, как часто встречаются те или иные варианты в ряду распределения. В табл. 1.4.1 частотами являются числа 8, 12, 3 и 2. Частоты, выраженные в долях единицы или в процентах к итогу, называются частостями.

Вариационные ряды в зависимости от характера вариации признака делятся на дискретные и интервальные. Дискретным называется вариационный ряд, образованный на основании дискретного признака. Примером дискретного ряда является распределение студентов по результатам экзамена (табл. 1.4.1). Интервальным называется вариационный ряд, у которого варианты представляют собой интервалы. Пример интервального ряда приведен в табл. 1.4.2, которая характеризует распределение рабочих предприятия по возрасту.

Таблица 1.4.2

Возраст рабочих, лет	20–30	30–40	40–50	50–60
Количество рабочих, чел.	7	28	15	5

Для наглядности и облегчения анализа вариационные ряды изображаются графически. Дискретные ряды изображаются при помощи полигона, а интервальные – при помощи гистограммы.

Решение типовых задач

Задача 1.4.1

Проверка содержания жира в 50 партиях сыра дала следующие результаты, %:

48	47	45	45	45	46	45	46	48	43
46	46	48	47	46	45	48	47	45	45
46	45	47	46	46	47	44	46	46	45
45	48	42	45	44	45	45	45	44	46
45	43	45	45	45	45	47	45	45	45

Постройте дискретный вариационный ряд.

Решение

Выпишем в порядке возрастания все значения изменяющегося признака (процент жирности сыра) и подсчитаем число партий сыра в каждой группе:

Процент жирности сыра	42	43	44	45	46	47	48	Всего
Число партий сыра, шт.	1	2	3	22	11	6	5	50

В этом вариационном ряду вариантами являются значения жирности сыра, а частотами – число партий. Результаты структурной группировки показывают, что чаще других встречаются (имеют наибольшую частоту) партии сыра с жирностью 45%.

Задача 1.4.2

По следующим данным о 25 рабочих цеха, изготовляющих одинаковую продукцию, постройте ряд распределения рабочих по стажу, образовав пять групп с равными интервалами:

№ п/п	Стаж работы, лет	Месячная выработка рабочего, шт.
1	5,0	244
2	3,0	205
3	9,2	298
4	4,4	250
5	6,9	280
6	2,5	230
7	16,0	310
8	13,2	284
9	14,0	320
10	11,0	295
11	12,0	279
12	4,5	222
13	10,5	276
14	1,0	234
15	9,0	270
16	6,5	252
17	5,0	241
18	6,0	256
19	10,1	262
20	5,5	245
21	2,5	240
22	5,3	252
23	7,5	253
24	7,0	252
25	8,0	262

Решение

Определим величину интервала группировочного признака (стажа работы):

$$h = \frac{x_{\max} - x_{\min}}{n} = \frac{16-1}{5} = 3 \text{ года.}$$

Следовательно, стаж рабочих первой группы от 1 до 4 лет, второй группы – от 4 до 7 лет и т.д. Подсчитаем число рабочих в каждой группе (частоту) и представим результаты в таблице:

№ группы	Группы рабочих по стажу, лет	Число рабочих	
		человек	% к итогу
I	1–4	4	16
II	4–7	9	36
III	7–10	5	20
IV	10–13	4	16
V	13–16	3	12
Итого		25	100

Как видно из таблицы, наибольшей по числу рабочих оказалась вторая группа, ее доля в общей численности рабочих составила 36%. Более половины всех рабочих (56%) имеют стаж работы от 4 до 10 лет.

Задача 1.4.3

Для изучения зависимости между стажем работы и месячной выработкой рабочих произведите аналитическую группировку данных задачи 1.4.2.

Решение

При построении аналитической группировки в качестве группировочного принимается факторный признак, каковым в решаемой задаче является стаж работы. Распределим всех рабочих по стажу на пять групп с равными интервалами. (Методика образования групп при построении интервального вариационного ряда рассмотрена в решении предыдущей задачи.) Охарактеризуем каждую группу числом рабочих, средним стажем работы и средней месячной выработкой рабочего. Для этого предварительно составим рабочую таблицу (интервалы групп приняты те же, что и в решении задачи 1.4.2):

№ п/п	Группы рабочих по стажу, лет	Номер рабочего	Стаж работы, лет	Месячная выработка, шт.
А	Б	1	2	3
I	2–4	2	3,0	205
		6	2,5	230
		14	1,0	234
		21	2,5	240
Итого		4	9,0	909
II	4–7	1	5,0	244
		4	4,4	250
		5	6,9	280
		12	4,5	222
А	Б	1	2	3
II	4–7	16	6,5	252
		17	5,0	241
		18	6,0	256
		20	5,5	245
		22	5,3	252
Итого		9	49,1	2242
III	7–10	3	9,2	298
		15	9,0	270
		23	7,5	253
		24	7,0	252
		25	8,0	262
Итого		5	40,7	1335
IV	10–13	10	11,0	295
		11	12,0	279
		13	10,5	276
		19	10,1	262
Итого		4	43,6	1112
V	13–16	7	16,0	310
		8	13,2	284
		9	14,0	320
Итого		3	43,2	914
Всего		**25**	**185,6**	**6512**

На основе групповых показателей рабочей таблицы исчислим средний стаж работы и среднюю месячную выработку рабочего для каждой группы. Полученные результаты представим в виде сводной аналитической таблицы:

№ группы	Группы рабочих по стажу, лет	Число рабочих, чел.	Средний стаж рабочего, лет	Средняя месячная выработка рабочего, шт.
I	1–4	4	2,3	227
II	4–7	9	5,5	249
III	7–10	5	8,1	267
IV	10–13	4	10,9	278
V	13–16	3	14,4	305
Итого		25	7,4	260

Из таблицы видно, что с ростом среднего стажа работы средняя месячная выработка рабочего увеличивается. Следовательно, между этими признаками существует прямая зависимость.

Задача 1.4.4

Имеются следующие данные о распределении предприятий отрасли экономики по численности работников:

№ группы	Группы предприятий по численности работников, чел.	Удельный вес группы предприятий, %
I	До 100	2
II	100–200	4
III	200–350	24
IV	350–450	38
V	450–600	22
VI	Свыше 600	10
Итого		100

Произведите перегруппировку данных, образовав новые группы с интервалами, чел.: до 200; 200–300; 300–400; свыше 400.

Решение

В первую группу вторичной группировки полностью войдут две первые группы предприятий из первичной группировки, сумма частостей которых равна 6% (2 + 4). Во вторую новую группу с численностью работников от 200 до 300 чел. войдет часть предприятий третьей группы, величина интервала которой составляет 150 чел. Примем условно, что количество предприятий внутри интервала распределяется равномерно. Тогда из третьей группы необходимо взять в новую вторую группу 2/3 (100 : 150) от количества ее предприятий (или от удельного веса), т.е. $2/3 \times 24 = 16\%$. В третью новую группу войдет оставшаяся часть предприятий третьей группы первичной группировки $(1/3 \times 24 = 8\%)$ и половина (50/100) предприятий четвертой группы, т.е. $1/2 \times 38 = 19\%$. В последнюю новую группу попадет вторая половина предприятий четвертой группы и все предприятия пятой и шестой групп.

Результаты вторичной группировки представлены в следующей таблице:

№ группы	Группы предприятий по численности работников, чел.	Удельный вес группы предприятий, %	Расчет удельного веса группы предприятий, %
I	До 200	6	2 + 4 = 6
II	200–300	16	$\frac{100}{150}24 = 16$
II	300–400	27	$\frac{50}{150}24 + \frac{50}{100}38 = 27$
IV	Свыше 400	51	$\frac{50}{100}38 + 22 + 10 = 51$
Итого		100	

Из таблицы следует, что каждое из 49% всех предприятий отрасли имеет не более 400 работников, причем на каждом предприятии первой группы, удельный вес которой составляет 6%, работает не более 200 человек.

Глава 2

ОБОБЩАЮЩИЕ СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ

2.1. Классификация обобщающих статистических показателей

Обобщающий статистический показатель – это количественно-качественная характеристика какого-либо свойства группы единиц в конкретных условиях места и времени. Качественная составляющая статистического показателя – это его социально-экономическое содержание безотносительно к размеру изучаемого явления, а количественная составляющая – это величина показателя, его числовое значение.

Обобщающие статистические показатели необходимы при планировании и для контроля выполнения планов, они применяются при анализе социально-экономических явлений и для управления ими, для характеристики состояния экономики и т.д. Обобщающие статистические показатели – важнейшая категория статистической науки.

Поскольку общественные явления и процессы взаимосвязаны, то характеризующие их статистические показатели также связаны между собой, т.е. образуют систему показателей. Например, взаимосвязаны такие показатели промышленного предприятия, как объем выпускаемой продукции, численность работников и производительность их труда. Система статистических показателей позволяет охарактеризовать явление или процесс более объемно, чем набор отдельных показателей, так как дополнительно дает информацию о взаимозависимости показателей.

Различают объемные и качественные обобщающие статистические показатели. Объемные показатели характеризуют численность или величину признака совокупности единиц (например, количество работников фирмы, стоимость произведенной на предприятии продукции). Качественные показатели характеризуют количественные соотношения явлений, уровень их развития (например, цена единицы товара, себестоимость единицы продукции, производительность труда работника).

В зависимости от степени агрегирования явлений статистические показатели подразделяются на индивидуальные и сводные. Индивидуальные показатели характеризуют одну единицу совокупности, а сводные – группу единиц или всю совокупность в целом.

По признаку времени различают моментные и интервальные статистические показатели. Моментные показатели характеризуют состояние явления на определенный момент времени (например, численность населения на начало года, величина товарных запасов на конец месяца). Интервальные показатели характеризуют результаты развития явлений за определенный период времени (например, производство продукции за месяц, величина товарооборота за квартал).

В зависимости от формы выражения статистические показатели подразделяются на абсолютные, относительные и средние. Сущность, виды и способы расчета абсолютных и относительных показателей рассматриваются в следующих параграфах.

2.2. Абсолютные показатели

Абсолютный статистический показатель – это характеристика размера общественного явления в конкретных условиях места и времени. Абсолютные показатели служат исходными данными для расчета относительных и средних показателей. Однако абсолютные показатели и сами имеют большое практическое и познавательное значение, они широко используются в статистико-экономическом анализе. В абсолютных показателях выражаются многие экономические и социальные явления: выпуск продукции, объем капитальных вложений, фонд заработной платы, количество рабочих.

Примером абсолютных статистических показателей, характеризующих Кыргызскую Республику, могут служить следующие данные. В 2000 г. стоимость основных фондов страны составила 116,5 млрд. сом., в 2003 г. в республике имелось 22 города, численность населения Кыргызстана к началу 2006 г. достигла 5,2 млн. чел. Эти показатели характеризуют разные явления, относятся к разным годам и различаются по другим признакам. Однако все эти показатели, как и любой другой показатель, обладают следующими атрибутами: 1) качественной определенностью; 2) количественным размером; 3) пространственной характеристикой; 4) временной характеристикой; 5) единицей измерения.

В соответствии с классификацией, рассмотренной в предыдущем параграфе, абсолютные статистические показатели подразделяются на индивидуальные и суммарные (сводные). Индивидуальный абсолютный показатель характеризует количественный признак отдельного объекта или отдельной единицы совокупности, он получается в процессе статистического наблюдения (например, выработка рабочего за смену, заработная плата отдельного работника). Суммарный абсолютный показатель характеризует количественный признак совокупности единиц, он получа-

ется в результате сводки и группировки данных статистического наблюдения или расчетным путем (например, поголовье овец в стране, выпуск продукции предприятиями города, валовой внутренний продукт).

Индивидуальные и суммарные абсолютные статистические показатели обязательно имеют единицы измерения. В зависимости от сущности явления, его физических свойств в статистике применяются натуральные, стоимостные и трудовые единицы измерения.

Натуральные единицы измерения выражают величину явлений в физических мерах (тонны, метры, квадратные метры, литры, штуки и т.д.). Чтобы измерить общий объем нескольких разновидностей какого-то продукта, их количество пересчитывается в условных единицах измерения на основе специальных коэффициентов, которые определяются как отношение потребительских свойств отдельных разновидностей продукта к эталону. Так, сыр, творог, сметану и сливочное масло пересчитывают в молоко; различные виды топлива пересчитывают в условное топливо. В некоторых случаях применяются комбинированные единицы измерения, которые представляют собой произведение двух единиц (тонно-километры, киловатт-часы и т.д.).

Стоимостные единицы измерения позволяют дать денежную оценку (в сомах, рублях, долларах и др.) социально-экономических явлений и процессов. В стоимостных единицах выражаются, например, общий объем разнородной продукции промышленности, основные фонды предприятия, валовой национальный доход и др.

Трудовые единицы измерения (человеко-часы, человеко-дни) используют для учета затрат труда при производстве продукции и услуг, для определения показателей производительности труда.

Рассмотрим пример применения условно-натуральных единиц измерения.

Задача 2.2.1

За отчетный период фабрика выпустила тетрадей: 12-листных – 50 тыс. шт., 24-листных – 20 тыс. шт., 60-листных – 10 тыс. шт., 96-листных – 5 тыс. шт.

Определите общий выпуск тетрадей в пересчете на 12-листные.

Решение

Для ответа на вопрос задачи необходимо прежде определить коэффициенты пересчета тетрадей в условные. Если принять за условную единицу 12-листную тетрадь, то коэффициенты пересчета будут равны:

для 24-листных тетрадей: 24 : 12 = 2;

для 60-листных тетрадей: 60 : 12 = 5;

для 96-листных тетрадей: 96 : 12 = 8.

Общий выпуск условных тетрадей в пересчете на 12-листные составил:

$50+20\times2+10\times5+5\times8=180$ тыс. шт.

2.3. Относительные показатели

Характеристика интенсивности социально-экономического явления, его изменения во времени и пространстве, изучение структуры явления, определение степени выполнения планового задания, сравнение различных показателей между собой – это важнейшие задачи статистического анализа. Для решения этих задач используются относительные показатели.

Относительными называются обобщающие показатели, получаемые в результате сравнения (деления) двух статистических величин. Величина, с которой сравнивают другую величину, называется основанием, или базой сравнения. Сравниваемая величина называется текущей, или отчетной. Относительные показатели могут выражаться в коэффициентах (база сравнения принимается за 1), в процентах (база сравнения принимается за 100), в промилле (база сравнения принимается за 1000). Форма выражения относительных показателей выбирается в зависимости от их назначения и соразмерности сравниваемых величин.

Относительные показатели различаются не только по форме выражения, но и по сущности характеризуемых соотношений. По этому признаку относительные показатели делятся на следующие виды: динамики, планового задания, выполнения плана, структуры, координации, сравнения, интенсивности.

Относительный показатель динамики (*ОПД*) характеризует изменение явления во времени. Он исчисляется делением величины явления в текущем (отчетном) периоде или моменте времени на величину этого явления в прошлом периоде или моменте времени:

$$ОПД=\frac{Ф_1}{Ф_0},$$

где $Ф_1$ – фактическая величина явления в текущем периоде; $Ф_0$ – фактическая величина явления в базисном периоде.

Выражается относительный показатель динамики в коэффициентах или процентах (в этом случае правая часть формулы умножается на 100).

При наличии данных за три и более последовательных периода или момента времени относительные показатели динамики можно исчислить двумя разными методами. Если данные каждого периода сравниваются с данными предыдущего периода, то такой метод расчета называется цепным. Если данные каждого периода сравниваются с данными какого-либо одного периода, принятого за базу, то такой метод расчета называется базисным.

Относительный показатель планового задания (*ОППЗ*) характеризует планируемое изменение изучаемого явления по сравнению с базисным периодом. Рассчитывается этот показатель делением планового задания на предстоящий (текущий) период на фактическую величину изучаемого явления за предшествующий (базисный) период:

$$ОППЗ=\frac{П_1}{Ф_0},$$

где $П_1$ – плановое задание на текущий период; $Ф_0$ – фактическая величина явления за базисный период.

Обычно ОППЗ выражается в процентах.

Относительный показатель выполнения плана (*ОПВП*) характеризует степень выполнения планового задания за определенный период времени. Вычисляется ОПВП делением фактически достигнутого в данном периоде уровня явления на запланированный уровень:

$$ОПВП=\frac{Ф_1}{П_1},$$

где $Ф_1$ – фактическая величина явления в текущем периоде; $П_1$ – плановое задание на текущий период.

Обычно ОПВП выражается в процентах.

Чтобы определить, на сколько процентов перевыполнен или недовыполнен план, необходимо отнять от исчисленного показателя 100%, если планировалось повышение уровня, и, наоборот, отнять от 100% исчисленный показатель, если планировалось снижение уровня.

Относительные показатели динамики, планового задания и выполнения плана связаны между собой:

$$ОПД=ОППЗ\times ОПВП.$$

Относительные показатели структуры (*ОПСт*) характеризуют доли отдельных частей изучаемой совокупности во всем ее объеме. Относительный показатель структуры исчисляется делением количества единиц или объема признака части совокупности на общее количество единиц или объема признака всей совокупности:

$$ОПСт = \frac{ЧС}{ВС},$$

где $ЧС$ – часть совокупности; $ВС$ – вся совокупность.

Выражается относительный показатель структуры в долях единицы или в процентах. Если этот показатель выражен в процентах, то его называют удельным весом. Сумма долей всех частей совокупности равна единице, а сумма всех удельных весов равна 100%.

Примерами относительных показателей структуры являются: удельный вес мужчин или женщин в общей численности населения, удельный вес отдельных групп основных фондов в их общей стоимости, доля отдельных статей затрат в общей сумме издержек производства продукции.

Относительный показатель координации (*ОПК*) характеризует соотношение отдельных частей изучаемой совокупности. Исчисляется этот показатель делением части совокупности на другую часть этой же совокупности, принятой за базу сравнения:

$$ОПК = \frac{ЧС_1}{ЧС_2},$$

где $ЧС_1$ – сравниваемая часть совокупности; $ЧС_2$ – часть совокупности, принятая за базу сравнения.

Относительный показатель координации обнаруживает, во сколько раз сравниваемая часть больше базы или сколько единиц сравниваемой части приходится на 10, 100, 1000 единиц базисной части совокупности.

Примерами относительных показателей координации являются: соотношение численности мужчин и женщин, работающих на предприятии, соотношение территорий двух областей республики, количество служащих, приходящееся на 100 рабочих завода.

Относительный показатель сравнения (*ОПСр*) характеризует соотношение одноименных статистических величин разных объектов за один и тот же период или момент времени:

$$ОПСр = \frac{\text{Показатель объекта А}}{\text{Показатель объекта Б}}.$$

Выражается этот показатель в процентах или кратных отношениях, показывающих, во сколько раз одна величина больше другой.

Примерами относительных показателей сравнения являются: соотношение рыночной цены на какой-либо товар у разных продавцов, сравнение производительности труда двух рабочих, сравнение фактической величины явления с нормативом.

Относительный показатель интенсивности (*ОПИ*) характеризует степень распространения или развития изучаемого явления в определенной среде. Исчисляется относительный показатель интенсивности как отношение двух разноименных абсолютных величин, связанных между собой:

$$ОПИ = \frac{\text{Показатель, характеризующий явление А}}{\text{Показатель, характеризующий среду распространения явления А}}.$$

Примерами относительных показателей интенсивности являются: плотность населения (чел./км^2), производительность труда (шт./ч), фондовооруженность труда (тыс. сом./чел.).

В большинстве случаев этот показатель выражается в единицах измерения тех абсолютных величин, отношение которых он характеризует. Однако некоторые относительные показатели интенсивности выражаются в процентах или промилле (например, рождаемость, смертность, брачность и др.).

Решение типовых задач

Задача 2.3.1

Товарооборот магазина в I квартале составил 813 тыс. сом. Во II квартале магазин планировал увеличить товарооборот до 870 тыс. сом. Фактический товарооборот второго квартала составил 838 тыс. сом.

Определите относительные показатели планового задания и выполнения плана.

Решение

$$ОППЗ = \frac{870}{813}100 = 107\%.$$

Следовательно, во II квартале планировалось увеличить товарооборот магазина по сравнению с первым кварталом на 7%.

$$ОПВП = \frac{838}{870}100 = 96{,}3\%.$$

План по товарообороту магазином выполнен на 96,3%, т.е. недовыполнение плана составило 3,7%.

Задача 2.3.2

Фирма планировала выпустить продукцию с уровнем себестоимости 35 сом. Фактически удалось снизить уровень себестоимости до 33 сом.

Определите относительный показатель выполнения плана по себестоимости продукции.

Решение

$$ОПВП = \frac{33}{35}100 = 94{,}3\%.$$

Следовательно, фактическая себестоимость продукции составила 94,3% от плановой, т.е. план по снижению себестоимости перевыполнен на 5,7%.

Задача 2.3.3

По данным о количестве квартир, построенных в Кыргызстане в 2002–2005 гг., рассчитайте относительные показатели динамики:

Год	2002	2003	2004	2005
Количество построенных квартир, тыс.	4,7	4,9	5,3	5,2

Решение

Цепные относительные показатели динамики:

2003 г.	2004 г.	2005 г.
$\frac{4{,}9}{4{,}7}100 = 104{,}3\%;$	$\frac{5{,}3}{4{,}9}100 = 108{,}2\%;$	$\frac{5{,}2}{5{,}3}100 = 98{,}1\%.$

Базисные относительные показатели динамики:

2003 г.	2004 г.	2005 г.
$\frac{4{,}9}{4{,}7}100 = 104{,}3\%;$	$\frac{5{,}3}{4{,}7}100 = 112{,}8\%;$	$\frac{5{,}2}{4{,}7}100 = 110{,}6\%.$

Следовательно, по сравнению с предыдущим годом, в 2003 и 2004 гг. количество построенных квартир увеличилось соответственно на 4,3 и 8,2%, а в 2005 г. уменьшилось на 1,9%. По сравнению с базисным 2002 г. количество построенных квартир в 2004 г. увеличилось на 12,8%, а в 2005 г. – на 10,6%.

Задача 2.3.4

Завод планировал увеличить выпуск продукции в отчетном году по сравнению с предыдущим годом на 4%. Фактически в отчетном году производство продукции увеличилось на 6%. Определите относительный показатель выполнения плана по выпуску продукции.

Решение

Примем предыдущий год за базисный, т.е. за 100%. Тогда:

$$ОППЗ = 100 + 4 = 104\%.$$

$$ОПД = 100 + 6 = 106\%.$$

На основании зависимости $ОПД = ОППЗ \times ОПВП$ рассчитаем искомый показатель:

$$ОПВП = \frac{ОПД}{ОППЗ} = \frac{106}{104} = 1{,}019, \text{ или } 101{,}9\%.$$

Следовательно, план по выпуску продукции перевыполнен на 1,9%.

Задача 2.3.5

Известна численность городского и сельского населения области:

Группы населения	Численность населения, млн. чел.	*ОПСт*, % к итогу
Городское	1,7	44,7
Сельское	2,1	55,3
Всего	3,8	100,0

Определите:
1) относительные показатели структуры населения;
2) относительный показатель координации.

Решение

1. В последней графе таблицы приведены относительные показатели структуры, исчисленные следующим способом:

удельный вес городского населения $\frac{1{,}7}{3{,}8}100 = 44{,}7\%,$

удельный вес сельского населения $\frac{2{,}1}{3{,}8}100 = 55{,}3\%.$

2. Относительный показатель координации: $\frac{2,1}{1,7} = 1,24.$

Следовательно, численность сельского населения превышает численность городского населения в 1,24 раза.

Задача 2.3.6

Среднегодовая численность населения Кыргызстана в 2005 г. достигла 5115,8 тыс. человек. Валовой сбор зерна в целом по стране в 2005 г. составил 1667,4 тыс. т. Определите, сколько зерна произведено в расчете на душу населения.

Решение

Исчислим относительный показатель интенсивности:

$$ОПИ = \frac{1667,4}{5115,8} = 0,326\ \frac{\text{т}}{\text{чел.}}$$

Следовательно, в 2005 г. в Кыргызстане произведено на душу населения 326 кг зерна.

Задача 2.3.7

Имеются следующие данные о цене 1 кг картофеля на двух рынках города: 15 и 18 сом. Сравните цены картофеля на рынках города между собой.

Решение

Исчислим относительный показатель сравнения:

$$ОПСр = 18:15 = 1,2, \text{ или } 120\%.$$

Следовательно, на втором рынке цена 1 кг картофеля выше цены на первом рынке на 20%.

2.4. Условия правильного применения относительных показателей

При анализе относительных показателей необходимо учитывать, на основе каких абсолютных величин они исчислены, так как иначе

можно сделать неправильные выводы. Рассмотрим следующий пример. Двое рабочих изготовляют одинаковые детали. В январе первый рабочий изготовил 5, а второй – 9 деталей. В течение года производительность труда обоих рабочих возросла. В декабре первый рабочий изготовил 10 деталей, а второй – 15. Таким образом, производительность труда у первого рабочего выросла в 2 раза (10 : 5), а у второго – в 1,7 раза (15 : 9). Если не учитывать абсолютные показатели выработки рабочих, то можно подумать, что первый рабочий достиг лучших результатов. Однако, если сравнить декабрьскую выработку рабочих, то оказывается, что у второго рабочего она в 1,5 раза (15 : 10) выше, чем у первого. Следовательно, производительность труда второго рабочего и в январе, и в декабре была выше.

Правильно объяснить исчисленный относительный показатель можно лишь с учетом того, какая из двух сравниваемых величин принята за базу сравнения. Рассмотрим результаты работы двух фирм. В течение месяца первая фирма произвела продукции на 500 тыс. сом., а вторая – на 800 тыс. сом. Сравним эти величины, приняв за базу сравнения объем продукции, произведенной первой фирмой:

$$\frac{800}{500} = 1,6, \quad \text{или } 160\%.$$

Следовательно, объем продукции второй фирмы превышает объем продукции первой в 1,6 раза, или на 60%. Казалось бы, что и, наоборот, объем продукции первой фирмы меньше объема продукции второй на 60%. Однако это неверно. Сравним исходные данные, приняв за базу сравнения объем продукции, произведенной второй фирмой:

$$\frac{500}{800}100 = 62,5\%.$$

Следовательно, объем продукции первой фирмы меньше объема продукции второй на 37,5% (62,5 – 100).

При вычислении относительных показателей необходимо учитывать требование сопоставимости сравниваемых статистических величин по территории, времени, единицам измерения, методики расчета и по другим признакам.

Глава 3

СРЕДНИЕ ВЕЛИЧИНЫ И ПОКАЗАТЕЛИ ВАРИАЦИИ

3.1. Сущность средних величин и показателей вариации

Статистическая совокупность состоит из отдельных единиц, обладающих индивидуальными особенностями и поэтому отличающихся друг от друга величиной количественного признака. Чтобы одним числом охарактеризовать уровень признака у всех единиц совокупности, когда размер признака у отдельных единиц колеблется (варьирует), исчисляется средняя величина.

Средней величиной в статистике называется обобщающий показатель, характеризующий размер признака, отнесенный к единице совокупности. В большинстве случаев средняя величина исчисляется делением общего объема варьирующего признака совокупности единиц на общую численность этих единиц. В соответствии с законом больших чисел в средней величине взаимопогашаются случайные индивидуальные различия между единицами и отражается то общее, что имеется в каждой единице однородной совокупности. Средние являются именованными числами, они выражаются в тех же единицах измерения, в каких измеряется признак.

Средние показатели подразделяются на типические и системные. Средняя величина, исчисленная для совокупности единиц, качественно однородной по изучаемому признаку, называется типической. Например, типической величиной является средний возраст пенсионеров страны. Средняя величина, исчисленная для единой, но неоднородной совокупности, называется системной. Например, системной величиной является средний возраст всех жителей страны, при расчете которого учитывается возраст младенцев, молодежи, пенсионеров и других возрастных групп. И типические, и системные средние величины широко применяются в статистике, так как специфика предмета вызывает необходимость постоянно оперировать обобщающими показателями.

Необходимо отметить, что средняя величина не позволяет достаточно полно охарактеризовать совокупность единиц, так как по ней

нельзя судить о том, как отклоняются от средней индивидуальные значения признака, т.е. какова вариация. Однако размер вариации признака позволяет судить о типичности и надежности средней величины. Чем меньше вариация, тем однороднее совокупность и, следовательно, тем типичнее средняя величина. При статистическом анализе социально-экономических явлений бывает, что средняя величина признака в двух совокупностях единиц одинакова, а вариация признака совершенно разная, и поэтому практическое значение этих средних тоже разное.

Для количественной характеристики колеблемости признака в совокупности единиц применяются показатели вариации.

3.2. Виды средних и способы их вычисления

Наиболее часто встречающиеся в статистике виды средних (арифметическая, гармоническая, квадратическая и геометрическая) относятся к классу степенных средних. Выбор вида средней зависит от сущности решаемой задачи. В зависимости от состояния исходных данных, по которым рассчитывается средний показатель, каждая из этих четырех средних может быть простой и взвешенной. Средние простые рассчитываются по несгруппированным единицам совокупности, а средние взвешенные – по сгруппированным, т.е. для рядов распределения.

Из всех степенных средних в статистике наиболее часто применяется *средняя арифметическая*. Она применяется, когда общий объем варьирующего признака для всей совокупности образуется как сумма значений признака отдельных единиц этой совокупности.

Формула средней арифметической имеет вид:

простая

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\Sigma x}{n},$$

взвешенная

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

где $\bar{x}$ – средняя величина; x – индивидуальные значения признака (варианты); n – численность единиц совокупности; f – частоты (веса), показывающие, сколько раз встречаются значения признака в совокупности единиц ($\Sigma f = n$).

Рассмотрим применение этих формул на примерах.

Задача 3.2.1

Бригада рабочих состоит из пяти человек. Месячная заработная плата рабочих составляет: 2400, 2500, 2800, 3000 и 3200 сом. Определите среднюю заработную плату рабочего бригады.

Решение

Поскольку имеются несгруппированные исходные данные, то для расчета средней месячной заработной платы воспользуемся формулой средней арифметической простой:

$$\overline{x} = \frac{\Sigma x}{n} = \frac{2400 + 2500 + 2800 + 3000 + 3200}{5} = 2780 \text{ сом.}$$

Задача 3.2.2

Определите среднюю заработную плату рабочих бригады, если известен месячный заработок каждого из них:

Месячная заработная плата рабочего, сом. (*x*)	Количество рабочих, чел. (*f*)	Произведение вариантов и частот (*xf*)
2400	3	7200
2600	5	13000
2800	10	28000
3000	2	6000
Итого	20	54200

Решение

По условию задачи имеется вариационный дискретный ряд, т.е. исходные данные сгруппированы. Поэтому для расчета средней заработной платы воспользуемся формулой средней арифметической взвешенной (сумма из произведений вариантов и частот вычислена в последней графе таблицы):

$$\overline{x} = \frac{\Sigma xf}{\Sigma f} = \frac{54200}{20} = 2710 \text{ сом.}$$

Задача 3.2.3

Определите среднюю заработную плату рабочих бригады по следующим данным:

Месячная заработная плата рабочего, сом.	Количество рабочих, чел. (*f*)	Середина интервала (*x*)	Произведение вариантов и частот (*xf*)
До 2600	3	2500	7500
2600–2800	7	2700	18900
2800–3000	5	2900	14500
3000–3200	3	3100	9300
Более 3200	2	3300	6600
Итого	20	–	56800

Решение

По условию задачи имеется интервальный вариационный ряд, т.е. исходные данные сгруппированы. Поэтому, как и при решении задачи 3.2.2, воспользуемся формулой средней арифметической взвешенной. С этой целью преобразуем интервальный вариационный ряд в дискретный, определив для каждого интервала его середину. Для закрытого интервала (вторая, третья и четвертая группы) среднее значение равно полусумме его верхней и нижней границ. Если интервальный ряд имеет открытые границы, то условно величина интервала первой группы приравнивается к величине интервала второй группы, а величина интервала последней группы принимается равной величине интервала предпоследней группы.

Все вспомогательные расчеты выполнены в таблице. Средняя заработная плата рабочих бригады составит:

$$\overline{x} = \frac{\Sigma xf}{\Sigma f} = \frac{56800}{20} = 2840 \text{ сом.}$$

Следует отметить, что исчисленная средняя является неточной, так как при расчете середины интервала предполагалось, что внутри интервалов варианты распределены равномерно, а это не всегда соответствует действительности. Кроме того, то обстоятельство, что границы первого и последнего интервалов вариационного ряда открыты и поэтому их середина найдена условно, делает среднюю еще менее точной.

Средняя гармоническая – это величина, обратная средней арифметической из обратных значений признака. Она применяется, когда общий объем варьирующего признака образуется как сумма обратных значений признака отдельных вариантов.

Формула средней гармонической имеет вид:

простая

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

взвешенная

$$\bar{x}_h = \frac{M_1 + M_2 + \ldots + M_n}{\frac{M_1}{x_1} + \frac{M_2}{x_2} + \ldots + \frac{M_n}{x_n}} = \frac{\Sigma M}{\sum \frac{M}{x}},$$

где M – объем изучаемого признака по группам ($M = xf$).

Из сравнения формул средней арифметической взвешенной и средней гармонической взвешенной следует, что последняя применяется, когда неизвестны частоты и приходится взвешивать варианты по объемам признака в каждой группе.

Рассмотрим применение формулы средней гармонической взвешенной на примере[1].

Задача 3.2.4

Определите среднюю заработную плату рабочих бригады по следующим данным:

Месячная заработная плата рабочего, сом. *(x)*	Месячный заработок группы рабочих, сом. *(M)*	Расчетная численность рабочих, чел. $\left(\frac{M}{x}\right)$
2400	7200	3
2600	13000	5
2800	28000	10
3000	6000	2
Итого	54200	20

Решение

По условию задачи имеются варианты и суммарные объемы признака (месячный заработок по группам рабочих), а информация о частотах (численности рабочих по группам) отсутствует. Поэтому определим среднюю месячную заработную плату рабочих бригады по формуле

[1] Средняя гармоническая простая в статистике применяется очень редко.

средней гармонической взвешенной (сумма из отношений M/x вычислена в последней графе таблицы):

$$\bar{x}_h = \frac{\Sigma M}{\sum \frac{M}{x}} = \frac{54200}{20} = 2710 \text{ сом.}$$

Средняя квадратическая исчисляется, когда общий объем варьирующего признака образуется как сумма квадратов отдельных вариантов. Формула средней квадратической имеет вид:

простая

$$\bar{x}_q = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}} = \sqrt{\frac{\Sigma x^2}{n}},$$

взвешенная

$$\bar{x}_q = \sqrt{\frac{x_1^2 f_1 + x_2^2 f_2 + \ldots + x_n^2 f_n}{f_1 + f_2 + \ldots + f_n}} = \sqrt{\frac{\Sigma x^2 f}{\Sigma f}}.$$

Средняя квадратическая применяется в статистике при исчислении показателей вариации.

Средняя геометрическая исчисляется, когда общий объем варьирующего признака образуется как произведение отдельных вариантов. Формула средней геометрической имеет вид:

простая

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \ldots x_n} = \sqrt[n]{\prod x},$$

взвешенная

$$\bar{x}_g = \sqrt[\Sigma f]{x_1^{f_1} \cdot x_2^{f_2} \ldots x_n^{f_n}} = \sqrt[\Sigma f]{\prod x^f}.$$

Средняя геометрическая применяется в статистике при исчислении средних темпов роста.

3.3. Мода и медиана

Кроме степенных средних, в статистике применяются структурные средние, к которым относятся мода и медиана. В отличие от степенной средней, являющейся абстрактной величиной, мода и медиана в дискретном вариационном ряду – это конкретные значения отдельных вариантов.

Модой называется величина признака, которая наиболее часто встречается в данной совокупности единиц. Мода применяется, например, для характеристики наиболее распространенного размера заработной платы на предприятии; модальной является цена, по которой на рынке было продано наибольшее количество данного товара.

Медианой называется величина признака у той единицы совокупности, которая находится в середине ранжированного ряда. Медиана делит вариационный ряд на две части, в одной из них все единицы имеют значение признака меньшее, чем у медианы, а в другой – большее. Медиану целесообразно применять вместо средней арифметической при наличии в вариационном ряду открытых интервалов. Из-за условно установленных границ открытых интервалов ряда средняя арифметическая является неточной, а на точность расчета медианы это не влияет. Медиана применяется при статистических методах контроля качества продукции. Совместное использование средней арифметической, моды и медианы при анализе вариационных рядов позволяет охарактеризовать тип распределения совокупности единиц.

В дискретном вариационном ряду модой является вариант, которому соответствует наибольшая частота. Для нахождения медианы в дискретном вариационном ряду необходимо определить накопленные частоты. Порядковый номер медианы ряда с нечетным числом вариантов (n) равен $\frac{n+1}{2}$. Если число вариантов четное, то медиана равна средней из двух вариантов, находящихся в середине ряда, т.е. имеющих порядковые номера $\frac{n}{2}$ и $\frac{n+2}{2}$.

Мода и медиана интервального вариационного ряда могут быть исчислены лишь приближенно. Сначала необходимо определить модальный и медианный интервалы. Модальным называется интервал, которому соответствует наибольшая частота. Медианным называется первый интервал, накопленная частота которого превышает половину общей суммы частот.

Модальная величина признака для интервального вариационного ряда определяется по формуле:

$$Mo = x_0 + h\frac{f_2 - f_1}{(f_2 - f_1)+(f_2 - f_3)},$$

где Mo – мода; x_0 – нижняя граница модального интервала; h – величина модального интервала; f_2 – частота модального интервала; f_1 – частота интервала, предшествующего модальному; f_3 – частота интервала, следующего за модальным.

Медиана для интервального вариационного ряда рассчитывается по формуле:

$$Me = x_0 + h\frac{0,5\Sigma f - S_{m-1}}{f_m},$$

где Me – медиана; x_0 – нижняя граница медианного интервала; h – величина медианного интервала; Σf – сумма частот ряда; S_{m-1} – сумма частот интервалов, предшествующих медианному; f_m – частота медианного интервала.

Решение типовых задач

Задача 3.3.1

Имеются следующие данные о распределении студентов по результатам сдачи экзамена (табл.). Определите моду и медиану.

Оценка	2	3	4	5
Количество студентов, чел.	28	32	23	22
Накопленные частоты	28	60	83	105

Решение

В этой таблице модальной оценкой является тройка, так как большинство студентов (32 человека) получило именно эту оценку. В середине ранжированного ряда находится студент с порядковым номером 53 $\left(\frac{105+1}{2}\right)$. Студент с этим порядковым номером относится ко второй группе, поскольку в нее попали студенты с порядковыми номерами от 29 до 60. Следовательно, медиана равна трем.

Задача 3.3.2

Имеются следующие данные о заработной плате работников фирмы (табл.). Определите моду и медиану заработной платы работников.

Заработная плата работника, сом.	2800–2900	2900–3000	3000–3100	3100–3200	3200–3300
Количество работников, чел.	12	22	28	44	10
Накопленные частоты	12	34	62	106	116

Решение

Модальным является интервал 3100–3200, так как ему соответствует наибольшая частота (44 человека). Подставим данные задачи в формулу моды:

$$Mo = 3100 + 100\frac{44-28}{(44-28)+(44-10)} = 3132 \text{ сом.}$$

Следовательно, у работников фирмы наиболее часто встречается заработная плата в размере 3132 сом.

Чтобы определить медианный интервал, в последней строке таблицы подсчитаны накопленные частоты. Первым интервалом, накопленная частота которого превышает половину общей суммы частот (58), является интервал третьей группы. Поэтому медианным будет интервал 3000–3100. Подставив соответствующие данные в формулу медианы, получим:

$$Me = 3000 + 100\frac{0,5\times116-34}{28} \approx 3086 \text{ сом.}$$

Следовательно, у половины работников фирмы заработная плата меньше, а у другой половины больше 3086 сом.

3.4. Показатели вариации признака

Для характеристики вариации количественного признака в совокупности единиц применяются следующие показатели: размах вариации, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение, коэффициент вариации и другие. Наиболее простым из перечисленных показателей является размах вариации (*R*), который представляет собой разность между максимальным и минимальным значениями признака в изучаемой совокупности:

$$R = x_{\max} - x_{\min}.$$

Чем больше значение *R*, тем больше вариация. Недостаток этого показателя заключается в том, что его величина зависит только от двух крайних значений признака и не учитывает, как часто встречается тот или иной вариант в данной совокупности единиц, т.е. не учитывает частоты отдельных вариантов. Следующие показатели позволяют более объективно охарактеризовать вариацию признака.

Среднее линейное отклонение представляет собой среднюю арифметическую из абсолютных значений отклонений отдельных вариантов от их средней арифметической. Среднее линейное отклонение $(\bar{d})$ вычисляется по формулам:

для несгруппированных данных

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

для сгруппированных данных

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

Чем больше значение $\bar{d}$, тем больше вариация.

Наибольшее применение среди показателей вариации получили дисперсия и среднее квадратическое отклонение. Значительное распространение этих показателей в качестве меры вариации признака объясняется их математическими свойствами. Дисперсией называется средний квадрат отклонений вариантов от их средней арифметической. Дисперсия признака (σ^2) вычисляется по формулам:

для несгруппированных данных

$$\sigma^2 = \frac{\sum(x-\bar{x})^2}{n};$$

для сгруппированных данных

$$\sigma^2 = \frac{\sum(x-\bar{x})^2 f}{\sum f}.$$

Ввиду важности этого показателя рассмотрим порядок его расчета по сгруппированным данным:

1) определяется отклонение каждого варианта от средней арифметической: $(x-\bar{x})$;

2) каждое отклонение возводится в квадрат: $(x-\bar{x})^2$;

3) каждый квадрат отклонения умножается на соответствующую частоту: $(x-\bar{x})^2 f$;

4) все найденные произведения суммируются: $\sum(x-\bar{x})^2 f$;

5) сумма произведений делится на сумму частот: $\frac{\sum(x-\bar{x})^2 f}{\sum f}$.

Среднее квадратическое отклонение (σ) представляет собой корень квадратный из дисперсии:

для несгруппированных данных

$$\sigma = \sqrt{\frac{\sum (x-\bar{x})^2}{n}};$$

для сгруппированных данных

$$\sigma = \sqrt{\frac{\sum (x-\bar{x})^2 f}{\sum f}}.$$

Среднее квадратическое отклонение, как и среднее линейное отклонение, показывает, насколько в среднем отклоняются индивидуальные варианты от их средней величины. Среднее квадратическое отклонение и среднее линейное отклонение выражаются в тех же единицах измерения, что и варианты.

Для сравнения вариации одного и того же признака в разных совокупностях или для сравнения вариации различных признаков в одной совокупности используется относительный показатель колеблемости – коэффициент вариации.

Коэффициент вариации (V) представляет собой процентное отношение среднего квадратического отклонения к средней арифметической величине изучаемого признака:

$$V = \frac{\sigma}{\bar{x}}100 \cdot$$

При измерении вариации альтернативного признака дисперсия определяется по формуле:

$$\sigma^2 = pq,$$

где p – доля единиц совокупности, обладающих изучаемым признаком; q – доля единиц совокупности, не обладающих изучаемым признаком.

Поскольку $p+q=1$, то формулу дисперсии альтернативного признака можно представить в виде:

$$\sigma^2 = p(1-p).$$

Решение типовых задач

Задача 3.4.1

По следующим данным определите показатели вариации возраста рабочих бригады:

Возраст рабочих, лет (x)	17	18	21	24	26
Количество рабочих, чел. (f)	2	2	6	3	1

Решение

Прежде всего необходимо определить средний возраст рабочих бригады. Для этой цели воспользуемся формулой средней арифметической взвешенной, так как по условию задачи известны варианты и частоты. Вспомогательные расчеты, необходимые для вычисления средней и показателей вариации, выполнены в таблице:

x	f	xf	$\lvert x-\bar{x}\rvert$	$\lvert x-\bar{x}\rvert f$	$\lvert x-\bar{x}\rvert^2 f$
17	2	34	4	8	32
18	2	36	3	6	18
21	6	126	0	0	0
24	3	72	3	9	27
26	1	26	5	5	25
Итого	14	294	–	28	102

1. Средний возраст рабочих бригады:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{294}{14} = 21 \text{ год}.$$

2. Среднее линейное отклонение:

$$\bar{d} = \frac{\sum \lvert x-\bar{x}\rvert f}{\sum f} = \frac{28}{14} = 2 \text{ года}.$$

3. Дисперсия

$$\sigma^2 = \frac{\sum (x-\bar{x})^2 f}{\sum f} = \frac{102}{14} = 7{,}3.$$

4. Среднее квадратическое отклонение:

$$\sigma = \sqrt{\frac{\sum (x-\bar{x})^2 f}{\sum f}} = \sqrt{7{,}3} = 2{,}7 \text{ года}.$$

5. Коэффициент вариации:

$$V = \frac{\sigma}{\bar{x}}100 = \frac{2{,}7}{21}100 = 12{,}9\%.$$

Задача 3.4.2

При обследовании 850 единиц изготовленной продукции оказалось, что 102 единицы имели различные дефекты и были забракованы. Определите дисперсию доли бракованной продукции.

Решение

Определим долю бракованной продукции среди обследованных единиц:

$$p = \frac{102}{850} = 0{,}12.$$

Дисперсия доли бракованной продукции:

$$\sigma^2 = p(1-p) = 0{,}12(1-0{,}12) = 0{,}106.$$

Глава 4

СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ДИНАМИКИ

4.1. Ряды динамики и их виды

Рядом динамики называется ряд числовых значений статистического показателя, расположенных в хронологической последовательности. Построение и анализ рядов динамики позволяет выявить тенденции и закономерности развития явления во времени, провести сравнительный анализ динамики нескольких явлений, обнаружить взаимосвязь между явлениями.

Ряд динамики состоит из двух элементов: 1) числовых значений статистического показателя, которые называются уровнями ряда; 2) периодов или моментов времени, к которым относятся уровни.

В зависимости от того, каким обобщающим показателем представлены уровни ряда, различают ряды динамики абсолютных, относительных и средних величин. По времени, которому соответствуют уровни ряда, различают моментные и интервальные ряды динамики.

Уровни моментного ряда динамики характеризуют состояние общественного явления на определенные моменты времени (на начало или другую дату месяца, квартала, года и т.п.). Примером моментного ряда динамики могут служить следующие данные о сумме вкладов населения в коммерческих банках Кыргызской Республики (на конец года, млрд. сом.):

2000 г.	2001 г.	2002 г.	2003 г.	2004 г.	2005 г.
1,5	1,6	1,5	1,9	2,2	2,8

Уровни интервального ряда динамики характеризуют итоги развития общественного явления за некоторый интервал времени (месяц, квартал, год и т.п.). Примером интервального ряда динамики могут служить следующие данные о величине годового выпуска продукции заводом, млн. сом.:

2000 г.	2001 г.	2002 г.	2003 г.	2004 г.	2005 г.
8,7	8,3	8,1	8,4	8,8	9,4

Уровень интервального ряда динамики представляет собой сумму уровней за более короткие интервалы времени. Так, годовой выпуск продукции завода – это сумма продукции, выпущенной заводом в течение 12 месяцев. Следовательно, значение уровня интервального ряда зависит от величины интервала времени. Уровни моментного ряда динамики не суммируются, поскольку это приводит к повторному счету, так как в каждом последующем уровне содержится полностью или частично значение предыдущего уровня.

В зависимости от продолжительности времени между уровнями ряды динамики подразделяются на два вида: 1) с равноотстоящими уровнями; 2) с неравноотстоящими уровнями. В рядах динамики с равноотстоящими уровнями периоды или даты времени следуют друг за другом с равными интервалами, а в рядах с неравноотстоящими уровнями эти интервалы неодинаковы. Оба приведенных выше примера – это ряды с равноотстоящими уровнями, а ниже следует пример ряда с неравноотстоящими уровнями:

Год	1980	1982	1985	1990	2000
Урожайность пшеницы, ц/га	18	20	21	28	26

При построении рядов динамики необходимо соблюдать требование сопоставимости уровней ряда между собой по единицам измерения, методике расчета, времени регистрации, территории и другим признакам. Если сопоставимость уровней ряда между собой нарушена, ряд динамики распадается на несколько отдельных частей; анализ тенденции динамики явления за весь изучаемый период в этом случае становится невозможным. Преобразование уровней ряда с целью устранения их несопоставимости называется смыканием рядов динамики. Смыкание позволяет получить единый ряд для исследования динамики явления за длительный период времени.

Рассмотрим прием смыкания рядов динамики на примере.

Задача 4.1.1

Имеются данные о розничном товарообороте района, млн. сом.:

Год	2000	2001	2002	2003	2004	2005
В старых границах района	520	540	600			
В новых границах района			750	792	810	935

Приведите ряды динамики к сопоставимому виду (сомкните ряды).

Решение

Показатели за 2003–2005 гг. не сопоставимы с показателями за 2000–2002 гг., так как они относятся к разным территориям (в 2002 г. территория района увеличилась). Необходимо исчислить данные за 2000–2001 гг. в новых границах.

Определим для 2002 г., в котором произошло изменение границ района, коэффициент соотношения уровней двух рядов:

$$k = 750 : 600 = 1{,}25.$$

Принимаем допущение, что в предыдущие годы коэффициент соотношения уровней двух рядов был приблизительно таким же. Умножив на этот коэффициент уровни ряда в старых границах за 2000–2001 гг., получим:

520 × 1,25 = 650 млн. сом.;

540 × 1,25 = 675 млн. сом.

Сомкнутый сопоставимый ряд динамики розничного товарооборота района (в новых границах) имеет вид, млн. сом.:

2000 г.	2001 г.	2002 г.	2003 г.	2004 г.	2005 г.
650	675	750	792	810	835

4.2. Показатели анализа ряда динамики

Для анализа скорости и интенсивности изменения отдельных уровней ряда динамики применяются показатели: 1) абсолютный прирост; 2) темп роста; 3) темп прироста; 4) абсолютное значение 1% прироста. Все эти показатели могут быть исчислены цепным или базисным методами. Если каждый уровень ряда динамики сравнивается с предыдущим, то такой метод расчета показателей называется цепным. Если все уровни ряда динамики сравниваются с каким-то одним уровнем, принятым за базу сравнения, то такой метод расчета называется базисным.

При любом методе расчета абсолютный прирост характеризует изменение явления во времени в абсолютном выражении, т.е. характеризует скорость изменения уровня. Этот показатель представляет собой разность между сравниваемыми уровнями ряда и выражается в тех же единицах измерения, что и сами уровни.

При цепном методе расчета абсолютный прирост исчисляется по формуле:

$$\Delta y_ц = y_i - y_{i-1},$$

где y_i – сравниваемый уровень ряда динамики; y_{i-1} – предыдущий уровень ряда.

При базисном методе расчета абсолютный прирост исчисляется по формуле:

$$\Delta y_б = y_i - y_б,$$

где $y_б$ – базисный уровень ряда.

Абсолютный прирост может быть положительным, если уровни ряда возрастают, и отрицательным, если уровни ряда убывают. Сумма последовательных цепных абсолютных приростов, начиная с первого, равна базисному абсолютному приросту последнего периода:

$$\sum \Delta y_ц = \Delta y_б = y_n - y_1,$$

где y_n – последний уровень ряда; y_1 – первый уровень ряда.

Темп роста характеризует относительное изменение уровня ряда динамики, т.е. характеризует интенсивность изменения явления за период времени. Этот показатель представляет собой отношение одного уровня ряда к другому и выражается в коэффициентах или в процентах.

При цепном методе расчета темп роста исчисляется по формуле:

$$Тр_ц = \frac{y_i}{y_{i-1}}.$$

При базисном методе расчета формула темпа роста имеет вид:

$$Тр_б = \frac{y_i}{y_б}.$$

Для того, чтобы выразить темп роста в процентах, правая часть этих формул умножается на 100. Выбор формы выражения (коэффициенты или проценты) зависит от масштабности динамики изучаемого явления[1]. Темп роста показывает, во сколько раз сравниваемый уровень больше уровня базисного периода (если $Тр > 1$), или какую часть сравниваемый уровень составляет от базисного уровня (если $Тр < 1$). Темп роста всегда число положительное.

Между цепными и базисными темпами роста, выраженными в коэффициентах, существуют следующие зависимости:

1. Произведение последовательных цепных темпов роста равно базисному темпу роста последнего периода.

2. Отношение базисного темпа роста последнего периода к предшествующему базисному темпу роста равно цепному темпу роста последнего периода.

Темп прироста дает относительную оценку скорости изменения уровня ряда динамики. Этот показатель представляет собой отношение

[1] Темп роста, выраженный в коэффициентах, авторы некоторых учебников называют коэффициентом роста.

абсолютного прироста к уровню, принятому за базу сравнения. Поскольку при цепном методе расчета за базу сравнения для каждого уровня принимается предшествующий ему уровень, то формула цепного темпа прироста имеет вид:

$$Тп_ц = \frac{\Delta y_ц}{y_{i-1}}.$$

Базисный темп прироста исчисляется по формуле:

$$Тп_б = \frac{\Delta y_б}{y_б}.$$

Темп прироста выражается в коэффициентах или в процентах. Чтобы выразить темп прироста в процентах, правая часть этих формул умножается на 100. Темп прироста может быть положительным, отрицательным или равным нулю. Темп прироста, выраженный в процентах, показывает, на сколько процентов увеличился (положительное число) или уменьшился (отрицательное число) текущий уровень по сравнению с базисным, принятым за 100%.

Между темпом прироста и темпом роста существует зависимость:

$$Тп = Тр - 1, \text{ или } Тп(\%) = Тр(\%) - 100.$$

Абсолютное значение одного процента прироста характеризует абсолютное изменение текущего уровня ряда динамики при одном проценте прироста. Этот показатель представляет собой отношение абсолютного прироста к темпу прироста, выраженному в процентах:

$$A = \frac{\Delta y_ц}{Тп_ц(\%)} = \frac{\Delta y_ц}{\frac{\Delta y_ц}{y_{i-1}} 100} = \frac{y_{i-1}}{100}.$$

Абсолютное значение одного процента прироста рассчитывается только цепным методом, так как при базисном методе расчета показатель постоянен для всех периодов или моментов времени. Выражается этот показатель в тех же единицах измерения, что и уровни ряда.

Рассмотрим пример исчисления показателей анализа ряда динамики.

Задача 4.2.1

Имеются следующие данные о производстве продукции предприятием за 2000–2005 гг., млн. сом.:

2000 г.	2001 г.	2002 г.	2003 г.	2004 г.	2005 г.
8,7	8,3	8,1	8,4	8,8	9,4

Определите (по годам и к базисному 2000 г.): 1) абсолютные приросты; 2) темпы роста; 3) темпы прироста; 4) абсолютное значение 1% прироста.

Решение

1. Цепные абсолютные приросты, млн. сом.:

2001 г. 8,3 – 8,7 = -0,4;

2002 г. 8,1 – 8,3 = -0,2 и т.д.

Базисные абсолютные приросты, млн. сом.:

2001 г. 8,3 – 8,7 = -0,4;

2002 г. 8,1 – 8,7 = -0,6 и т.д.

2. Цепные темпы роста, %:

2001 г. $\frac{8,3}{8,7}100 = 95,4;$

2002 г. $\frac{8,1}{8,3}100 = 97,6$ и т.д.

Базисные темпы роста, %:

2001 г. $\frac{8,3}{8,7}100 = 95,4;$

2002 г. $\frac{8,1}{8,7}100 = 93,1$ и т.д.

3. Цепные темпы прироста, %:

2001 г. 95,4 – 100 = -4,6;

2002 г. 97,6 – 100 = -2,4 и т.д.

Базисные темпы прироста, %:

2001 г. 95,4 – 100 = -4,6;

2002 г. 93,1 – 100 = -6,9 и т.д.

4. Абсолютное значение 1% прироста, млн. сом.:

2001 г. $\frac{8,7}{100} = 0,087;$

2002 г. $\frac{8,3}{100} = 0,083$ и т.д.

Для наглядности и удобства анализа исходные данные и рассчитанные показатели представим в таблице. Как видно из этой таблицы, до 2002 г. происходило снижение производства продукции, причем в 2002 г. по сравнению с 2001 г. темп снижения уменьшился. Начиная с 2003 г. производство продукции ежегодно увеличивалось как абсолютно, так и относительно; росло абсолютное значение 1% прироста.

Динамика производства продукции предприятием за 2000–2005 гг.

Год	Продукция, млн. сом.	Абсолютные приросты, млн. сом.		Темпы роста, %		Темпы прироста, %		Абсолютное значение 1% прироста, млн. сом.
		цепные	базисные	цепные	базисные	цепные	базисные	
2000	8,7	–	–	–	100,0	–	–	–
2001	8,3	-0,4	-0,4	95,4	95,4	-4,6	-4,6	0,087
2002	8,1	-0,2	-0,6	97,6	93,1	-2,4	-6,9	0,083
2003	8,4	0,3	-0,3	103,7	96,6	3,7	-3,4	0,081
2004	8,8	0,4	0,1	104,8	101,1	4,8	1,1	0,084
2005	9,4	0,6	0,7	106,8	108,0	6,8	8,0	0,088

4.3. Средние характеристики ряда динамики

При анализе динамики явления за период в целом используются обобщающие показатели: 1) средний уровень ряда динамики; 2) средний абсолютный прирост; 3) средний темп роста; 4) средний темп прироста. Потребность в этих показателях возникает в тех случаях, когда аналитические показатели ряда динамики, исчисленные для индивидуальных уровней, попеременно возрастают и убывают. Так, в сельском хозяйстве урожайность растений, количество произведенной продукции, ее себестоимость и другие показатели могут существенно варьировать по годам, поскольку сильно зависят от погодных условий года. Следовательно, анализ динамики отдельных уровней таких рядов не всегда целесообразен. В этих случаях анализируют средние показатели ряда динамики по пятилетиям или за другие периоды. Средние характеристики ряда динамики требуются для расчета многих показателей социально-экономической статистики и при анализе их изменения во времени.

Средний уровень ряда динамики называется средней хронологической. Методы расчета средней хронологической зависят от вида ряда динамики.

Средний уровень интервального ряда динамики с равноотстоящими уровнями рассчитывается по формуле средней арифметической простой:

$$\overline{y} = \frac{\Sigma y}{n},$$

где n – число уровней ряда.

Средний уровень интервального ряда динамики с неравноотстоящими уровнями рассчитывается по формуле средней арифметической взвешенной:

$$\bar{y} = \frac{\Sigma yt}{\Sigma t},$$

где t – длительность интервала времени между смежными датами; y – уровни ряда динамики, не меняющиеся в течение времени t.

Средний уровень моментного ряда динамики с равноотстоящими уровнями определяется по формуле:

$$\bar{y} = \frac{\frac{1}{2}y_1 + y_2 + y_3 + \ldots + y_{n-1} + \frac{1}{2}y_n}{n-1}.$$

При расчете по этой формуле средней хронологической за квартал суммируются уровни ряда за 4 месяца, для года суммируются уровни ряда за 13 месяцев (с первого января отчетного года по первое января следующего года включительно).

Средний уровень моментного ряда динамики с неравноотстоящими уровнями определяется по формуле:

$$\bar{y} = \frac{(y_1 + y_2)t_1 + (y_2 + y_3)t_2 + \ldots + (y_{n-1} + y_n)t_{n-1}}{2(t_1 + t_2 + t_3 + \ldots + t_{n-1})}.$$

Средний абсолютный прирост рассчитывается по формуле средней арифметической простой из цепных абсолютных приростов за последовательные и равные интервалы времени:

$$\overline{\Delta y} = \frac{\Sigma \Delta y_ц}{m},$$

где m – число цепных абсолютных приростов.

Если цепные абсолютные приросты за последовательные равные интервалы времени не известны, то средний абсолютный прирост можно определить по формуле:

$$\overline{\Delta y} = \frac{y_n - y_1}{n-1},$$

где y_1 – начальный уровень ряда; y_n – конечный уровень ряда.

Средний темп роста представляет собой среднюю геометрическую из цепных темпов роста. Средний темп роста для ряда динамики с равноотстоящими уровнями исчисляется по формуле средней геометрической простой:

$$\overline{Tp} = \sqrt[m]{Tp_1 \times Tp_2 \times \ldots \times Tp_m},$$

где m – количество цепных темпов роста; Tp_1, Tp_2, …, Tp_m – цепные темпы роста за последовательные и равные интервалы времени.

Если цепные темпы роста не известны, то средний темп роста можно определить по одной из следующих формул:

$$\overline{Tp} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[n-1]{Tp_б},$$

где $Tp_б$ – базисный темп роста за рассматриваемый период.

Средний темп роста, выраженный в коэффициентах, показывает, во сколько раз изменяется уровень ряда динамики в среднем за единицу времени (например, ежемесячно, ежеквартально, ежегодно). Для определения средних темпов роста применяются специальные таблицы, которые позволяют без вычислений получить искомый результат в зависимости от величины подкоренного числа и степени корня.

Средний темп роста для ряда динамики с неравноотстоящими уровнями исчисляется по формуле средней геометрической взвешенной:

$$\overline{Tp} = \sqrt[\Sigma t]{Tp_1^{t1} \times Tp_2^{t2} \times \ldots \times Tp_m^{tm}}.$$

Средний темп прироста рассчитывается на основе среднего темпа роста:

$$\overline{Tn} = \overline{Tp} - 1, \text{ или } \overline{Tn}(\%) = \overline{Tp}(\%) - 100.$$

Средний темп прироста, выраженный в процентах, показывает, на сколько процентов изменялся уровень ряда в среднем за единицу времени (например, ежегодно). Средний темп прироста может быть положительной или отрицательной величиной. Отрицательный средний темп прироста характеризует снижение, а положительный – увеличение уровней ряда.

Решение типовых задач

Задача 4.3.1

По данным задачи 4.2.1 определите за период с 2000 по 2005 гг.: 1) средний уровень ряда; 2) среднегодовой темп роста; 3) среднегодовой темп прироста; 4) средний абсолютный прирост.

Решение

1. По условию задачи имеется интервальный ряд динамики с равноотстоящими уровнями, поэтому средний уровень ряда следует исчислить по формуле средней арифметической простой:

$$\bar{y} = \frac{\Sigma y}{n} = \frac{8,7 + 8,3 + 8,1 + 8,4 + 8,8 + 9,4}{6} = 8,6 \text{ млн. сом.}$$

2. Среднегодовой темп роста равен

$$\overline{T}p = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[6-1]{\frac{9,4}{8,7}} = \sqrt[5]{1,080} = 1,015 \text{ (101,5\%).}$$

3. Среднегодовой темп прироста составляет

$$\overline{T}n = \overline{T}p(\%) - 100\% = 101,5 - 100 = 1,5\%.$$

Следовательно, в течение пяти лет с 2001 г. по 2005 г. производство продукции на предприятии ежегодно увеличивалось в среднем на 1,5%.

4. Средний абсолютный прирост равен

$$\overline{\Delta y} = \frac{y_n - y_1}{n-1} = \frac{9,4 - 8,7}{6-1} = 0,14 \text{ млн. сом.}$$

Следовательно, с 2001 по 2005 гг. производство продукции на предприятии ежегодно увеличивалось в среднем на 0,14 млн. сом.

Задача 4.3.2

Имеются следующие данные о товарных запасах магазина на начало каждого месяца текущего года, тыс. сом.:

1/I	1/II	1/III	1/IV
158	169	133	142

Определите средние товарные запасы магазина за первый квартал.

Решение

По условию задачи имеется моментный ряд динамики с равноотстоящими уровнями, поэтому средний уровень ряда определяется по формуле:

$$\overline{y} = \frac{\frac{1}{2}y_1 + y_2 + y_3 + \ldots + y_{n-1} + \frac{1}{2}y_n}{n-1} = \frac{\frac{1}{2}158 + 169 + 133 + \frac{1}{2}142}{4-1} = 151 \text{ тыс. сом.}$$

Задача 4.3.3

В течение апреля произошли следующие изменения в списочном составе работников предприятия, чел.:

состояло по списку на 1/IV 122
выбыло с 6/IV 3
зачислено с 14/IV 2
зачислено с 23/IV 4

Определите среднедневную списочную численность работников предприятия в апреле.

Решение

Определим продолжительность каждого календарного периода с постоянной численностью работников.

Календарные периоды апреля	Численность работников, чел. y	Продолжительность периода, дни t
1–5	122	5
6–13	119	8
14–22	121	9
23–30	125	8

Имеется интервальный ряд динамики с неравноотстоящими уровнями, поэтому средний уровень ряда определяется по формуле:

$$\overline{y} = \frac{\Sigma yt}{\Sigma t} = \frac{122 \cdot 5 + 119 \cdot 8 + 121 \cdot 9 + 125 \cdot 8}{5+8+9+8} = 122 \text{ чел.}$$

Задача 4.3.4

Известны следующие данные об остатках сырья на складе предприятия, т:

1/I 2005	1/V 2005	1/VIII 2005	1/I 2006
59	53	48	67

Определите средний уровень остатков сырья на складе предприятия в 2005 г.

Решение

По условию задачи имеется моментный ряд динамики с неравноотстоящими уровнями, поэтому средний уровень остатков сырья определяется по формуле:

$$\overline{y}=\frac{(y_1+y_2)t_1+(y_2+y_3)t_2+\ldots+(y_{n-1}+y_n)t_{n-1}}{2(t_1+t_2+t_3+\ldots+t_{n-1})}=$$

$$=\frac{(59+53)\cdot 4+(53+48)\cdot 3+(48+67)\cdot 5}{2(4+3+5)}=55,3 \text{ т.}$$

Задача 4.3.5

Темп роста производительности труда на предприятии составил за пятилетие 2001–2005 гг. 131,3%, а в 2006 г. он был равен 106,2%. Определите среднегодовой темп роста производительности труда на предприятии за 2001–2006 гг.

Решение

Базисный темп роста производительности труда в 2006 г. по сравнению с 2000 г. равен произведению двух цепных темпов роста:

$$Tp_б=1,313\times 1,062=1,3944.$$

Среднегодовой темп роста производительности труда на предприятии за 2001–2006 гг., т.е. в 2006 г. по сравнению с 2000 г., определяется по формуле:

$$\overline{T}p=\sqrt[n-1]{Tp_б}=\sqrt[7-1]{1,3944}=1,057 \text{ (105,7\%)}.$$

Задача 4.3.6

За 2001–2005 гг. в магазине объем продаж продовольственных товаров увеличился в 1,28 раза, а непродовольственных товаров возрос на 37%. Определите среднегодовые темпы прироста объемов продаж продовольственных и непродовольственных товаров за 2001–2005 гг.

Решение

Продовольственные товары:
среднегодовой темп роста $\overline{T}p=\sqrt[n-1]{Tp_б}=\sqrt[5]{1,28}=1,051$ (105,1%);
среднегодовой темп прироста $\overline{T}n=\overline{T}p(\%)-100=105,1-100=5,1\%$.

Непродовольственные товары:
среднегодовой темп роста $\overline{T}p=\sqrt[5]{1,37}=1,065$ (106,5%);

среднегодовой темп прироста $\overline{T}n=106,5-100=6,5\%$.

Следовательно, в среднем за рассматриваемый период времени годовой объем продаж непродовольственных товаров увеличивался более высокими темпами, чем объем продаж продовольственных товаров.

4.4. Выявление общей тенденции развития ряда динамики

Под тенденцией развития ряда динамики понимается общее направление изменения его уровней, свободное от случайных колебаний. На развитие явления во времени могут оказывать влияние различные по своему характеру и силе воздействия факторы. Одни из них оказывают постоянное воздействие и формируют в ряду динамики определенную тенденцию развития (тренд). Воздействие других факторов может быть периодическим или разовым, что вызывает кратковременные или разовые колебания уровней ряда динамики, их отклонение от общей тенденции развития явления.

Влияние случайных факторов на отдельные уровни ряда делает менее очевидной сущность явления, мешает выявить тенденцию развития ряда динамики. Поэтому встречаются ряды динамики, тенденция развития которых не очевидна. Чтобы освободиться от влияния случайных факторов, мешающих выявлению основной тенденции ряда динамики, применяются следующие методы: 1) укрупнение интервалов; 2) сглаживание ряда динамики скользящей средней; 3) аналитическое выравнивание.

Укрупнение интервалов заключается в том, что уровни ряда динамики за короткие интервалы времени, подверженные случайным колебаниям, заменяют средним уровнем за более длительный период. Например, ряд с ежедневными уровнями явления преобразуют в ряд динамики с недельными или месячными средними уровнями этого явления. Укрупнение обычно начинают с объединения 2–3 интервалов, а если это не проясняет тенденцию, то объединяют большее число интервалов. Недостатком этого метода является то, что при его использовании остаются неисследованными процессы изменения уровней внутри укрупненного периода.

Сглаживание ряда динамики с помощью скользящей средней заключается в том, что исчисляется средняя для укрупненного периода, состав которого последовательно меняется отбрасыванием одного уровня в начале и добавлением одного уровня в конце периода. При выборе продолжительности периода, по которому рассчитывается скользящая средняя, обычно объединяют нечетное число уровней ряда динамики. Это позволяет отнести исчисленную среднюю к середине укрупненного периода. Сглаживание ряда динамики можно начать с исчисления средних для трех уровней ряда, а если это не прояснит тенденцию, то число уровней в укрупненном периоде увеличивают.

Сущность этого метода можно выразить при помощи формул. Если сглаживание проводится осреднением трех уровней укрупненного периода, то формулы имеют вид:

$$\overline{y}_1 = \frac{y_1 + y_2 + y_3}{3};\ \overline{y}_2 = \frac{y_2 + y_3 + y_4}{3};\ \overline{y}_3 = \frac{y_3 + y_4 + y_5}{3} \text{ и т.д.}$$

В результате получается новый ряд динамики, в значительной мере очищенный от случайных колебаний. Недостатком метода скользящей средней является уменьшение числа уровней в сглаженном ряду. Если сглаживание проводилось осреднением трех уровней ряда, то новый ряд будет короче фактического на один уровень в начале и на один уровень в конце. При замене средней величиной пяти уровней новый ряд будет короче фактического на два члена в начале и в конце ряда. Чем больше уровней ряда включается в укрупненный период, тем более он будет сглажен и тем отчетливее проявится тенденция динамики, но и тем короче будет сглаженный ряд.

Более совершенным, но и более сложным методом выявления тенденции развития ряда динамики является аналитическое выравнивание. Сущность этого метода заключается в замене фактических уровней ряда динамики уровнями, исчисленными на основе математической модели, отображающей основную тенденцию ряда динамики как функцию времени. Аналитическое выравнивание ряда динамики состоит из следующих этапов.

1. Выбирается математическая модель, отражающая основную тенденцию ряда динамики, как функцию времени, которая наилучшим образом характеризует динамику изучаемого явления. Вид функции (прямая, парабола, гипербола или иная) выбирается на основе сочетания теоретического анализа сущности явления с исследованием тенденции изменения фактических уровней ряда динамики. Для наглядности обычно выравниваемый ряд динамики изображается графически.

2. Вычисляются параметры выбранной функции. Параметры функции обычно определяются методом наименьших квадратов. В соответствии с этим методом неизвестные параметры уравнения подбираются таким образом, чтобы сумма квадратов отклонений фактических уровней ряда динамики от соответствующих им выравненных, т.е. исчисленных при помощи найденного уравнения, была бы минимальной:

$$\sum (y - y_t)^2 \to \min,$$

где y – фактические уровни ряда динамики; y_t – выравненные (расчетные) уровни.

3. Подставляя в найденное уравнение фактические значения времени, вычисляют уровни выравненного ряда динамики.

Наибольшее распространение на практике получило выравнивание ряда динамики по прямой:

$$y_t = a_0 + a_1 t,$$

где a_0, a_1 – параметры уравнения прямой; t – порядковый номер момента или периода времени.

Параметры уравнения прямой a_0 и a_1, соответствующие требованию метода наименьших квадратов, находятся решением системы нормальных уравнений:

$$\begin{cases} a_0 n + a_1 \Sigma t = \Sigma y; \\ a_0 \Sigma t + a_1 \Sigma t^2 = \Sigma ty, \end{cases}$$

где n – число уровней ряда.

Для выявления параметров иных функций используются другие системы уравнений, с которыми можно ознакомиться в учебниках теории статистики [1; 2; 6].

Рассмотрим применение методики выявления общей тенденции развития ряда динамики на примере следующей задачи.

Задача 4.4.1

Имеются данные о потреблении фруктов и ягод населением области за 1997–2005 гг. в расчете на одного члена домохозяйства в год, кг:

1997 г.	1998 г.	1999 г.	2000 г.	2001 г.	2002 г.	2003 г.	2004 г.	2005 г.
22	24	21	25	22	24	26	23	25

Требуется выявить основную тенденцию потребления фруктов и ягод за 1997–2005 гг.: 1) методом укрупнения интервалов; 2) методом скользящей средней; 3) методом аналитического выравнивания.

Решение

1. Метод укрупнения интервалов

Разделим исходные данные на три трехлетних интервала и исчислим средний уровень для каждого из них.

Период 1997–1999 гг.: $\frac{22+24+21}{3}=22,3$ кг.

Период 2000–2002 гг.: $\frac{25+22+24}{3}=23,7$ кг.

Период 2003–2005 гг.: $\frac{26+23+25}{3}=24,7$ кг.

Так как 22,3 < 23,7 < 24,7, то можно сделать вывод о том, что в 1997–2005 гг. наблюдалась общая тенденция к росту потребления фруктов и ягод населением области.

2. Метод скользящей средней.

Исчислим трехлетние скользящие средние:

$$\overline{y}_1=\frac{y_1+y_2+y_3}{3}=\frac{22+24+21}{3}=22,3 \text{ кг};$$

$$\overline{y}_2=\frac{y_2+y_3+y_4}{3}=\frac{24+21+25}{3}=23,3 \text{ кг};$$

$$\overline{y}_3=\frac{y_3+y_4+y_5}{3}=\frac{21+25+22}{3}=22,7 \text{ кг и т.д.}$$

Результаты расчета трехлетней скользящей средней представлены в таблице.

Год	Потребление фруктов и ягод на одного члена домохозяйства в год, кг (y)	Трехлетние скользящие суммы	Трехлетние скользящие средние $(\overline{y})$
1997	22	–	–
1998	24	67	22,3
1999	21	70	23,3
2000	25	68	22,7
2001	22	71	23,7
2002	24	72	24,0
2003	26	73	24,3
2004	23	74	24,7
2005	25	–	–

Уровни сглаженного ряда динамики показывают, что в 1997–2005 гг. наблюдалась общая тенденция роста потребления фруктов и ягод населением области.

3. Метод аналитического выравнивания ряда динамики по прямой.

Для того чтобы найти параметры a_0 и a_1 уравнения прямой, необходимо решить систему нормальных уравнений

$$\begin{cases} a_0 n + a_1 \Sigma t = \Sigma y; \\ a_0 \Sigma t + a_1 \Sigma t^2 = \Sigma t y. \end{cases}$$

Найдем значения Σy, Σt, Σt^2, $\Sigma t y$ при помощи таблицы.

Год	Фактические уровни ряда (y)	Условное обозначение времени (t)	t^2	ty	Выравненные уровни ряда (y_t)
1997	22	1	1	22	22,4
1998	24	2	4	48	22,7
1999	21	3	9	63	23,0
2000	25	4	16	100	23,3
2001	22	5	25	110	23,6
2002	24	6	36	144	23,9
2003	26	7	49	182	24,2
2004	23	8	64	184	24,5
2005	25	9	81	225	24,8
Итого	212	45	285	1078	212,4

Следовательно, для данной задачи система нормальных уравнений имеет вид:

$$\begin{cases} a_0 9 + a_1 45 = 212; \\ a_0 45 + a_1 285 = 1078. \end{cases}$$

Решив эту систему уравнений, найдем:

$$a_0 = 22,1; \quad a_1 = 0,3.$$

В результате получаем уравнение общей тенденции ряда динамики:

$$y_t = 22,1 + 0,3t.$$

Подставив в это уравнение значения t (1, 2, 3, 4 и т.д.), вычислим выравненные теоретические значения y_t (табл.).

Теоретически Σy_t должна быть равна Σy. Однако в нашем примере они незначительно отличаются из-за округлений результатов при расчете параметров a_0 и a_1.

Вычисление значений параметров a_0 и a_1 можно несколько упростить, если обозначить время (t) так, чтобы начало отсчета времени приходилось на середину рассматриваемого периода [1; 5; 6].

4.5. Сравнительный анализ нескольких рядов динамики

Задача сравнительного анализа рядов динамики заключается в сравнении характеристик направления и интенсивности развития явлений во времени.

Сравнительный анализ нескольких рядов динамики подразделяется на два вида: 1) сравнение развития одноименных явлений, относящихся к разным объектам; 2) сравнение развития разных, но взаимосвязанных явлений. Примером анализа первого вида может служить сравнение рядов динамики производительности труда на двух предприятиях, цен на какой-либо товар на разных рынках, производства какой-либо продукции в разных странах или в регионах одной страны. При сравнительном анализе первого вида используют рассмотренные выше аналитические показатели, а также сравнивают между собой абсолютные значения уровней рядов динамики за один и тот же период или момент времени. Сравним, например, данные о производстве сливочного масла по двум областям Кыргызской Республики, т:

Таблица 4.5.1

Год / Область	2000	2001	2002	2003	2004	2005
Таласская область	118	151	211	232	336	459
Ысыккольская область	163	214	262	323	456	518

Расчеты, выполненные по табличным данным, показывают, что в 2000 г. производство сливочного масла в Таласской области составляло 72,4% объема производства в Ысыккольской области. За пятилетие 2001–2005 гг. среднегодовые темпы прироста производства сливочного масла в Таласской области были выше, чем в Ысыккольской области (соответственно 31 и 26%), поэтому разрыв между областями сократился. В 2005 г. объем производства сливочного масла в Таласской области составил 88,6% уровня Ысыккольской области.

Примером сравнительного анализа второго вида может служить таблица 4.5.2, характеризующая динамику количества вузов и студентов в Кыргызстане.

Очевидно, что в этом случае, как и при любом сравнительном анализе второго вида, сравнивать между собой абсолютные значения уровней разных рядов динамики нельзя. Поэтому их заменяют относительными показателями – базисными темпами роста, которые затем

сравнивают между собой. Для того чтобы темпы роста рядов динамики разных, но взаимосвязанных явлений можно было сравнивать, их рассчитывают по отношению к одному и тому же году. Этот прием называется приведением рядов динамики к общему основанию (к общей базе сравнения).

Таблица 4.5.2

Год	2002	2003	2004	2005
Количество вузов, единиц	46	47	49	51
в % к 2002 г.	100,0	102,2	106,5	110,9
Количество студентов, тыс. чел.	199,1	203,0	218,3	231,1
в % к 2002 г.	100,0	101,9	109,6	116,1

В рассматриваемом примере за общую базу сравнения принят 2002 г. (табл. 4.5.2). Из таблицы видно, что в 2003 г. количество вузов росло быстрее числа студентов, но в последующие годы, наоборот, темпы роста численности студентов были значительно выше.

Для сравнения интенсивности развития явлений, отражаемых двумя рядами динамики, применяется коэффициент опережения. Коэффициентом опережения называется отношение базисного темпа роста одного ряда динамики к базисному темпу роста другого ряда за одинаковые периоды времени:

$$Ko = \frac{Tp_{б1}}{Tp_{б2}},$$

где $Tp_{б1}$, $Tp_{б2}$ – базисные темпы роста первого и второго рядов динамики.

Коэффициент опережения показывает, во сколько раз быстрее растут уровни одного ряда динамики по сравнению с уровнями другого.

Рассчитаем коэффициент опережения для рассматриваемого примера (табл. 4.5.2).

$$Ko = \frac{116{,}1}{110{,}9} = 1{,}05.$$

Следовательно, в 2002–2005 гг. в Кыргызской Республике численность студентов росла быстрее, чем количество вузов, в 1,05 раза.

Если уровни рядов динамики изменяются неравномерно (но в одном направлении!), то при исчислении коэффициента опережения сравниваются средние темпы роста этих рядов.

4.6. Интерполяция и экстраполяция рядов динамики

Вскрытая динамика развития явления позволяет определить неизвестные уровни внутри ряда и прогнозировать величину его будущих уровней.

Приблизительный расчет неизвестных уровней внутри ряда динамики называется интерполяцией. Приблизительный расчет уровней, находящихся за пределами известных значений членов ряда динамики, называется экстраполяцией. При интерполяции или экстраполяции исходят из предпосылки, что в том периоде, для которого рассчитывается неизвестный уровень, действует общая тенденция развития изучаемого ряда динамики.

Если известно уравнение, характеризующее тенденцию ряда динамики как функцию времени, то определить искомые уровни можно на основе этого уравнения. Так, на основе уравнения, характеризующего тенденцию потребления фруктов и ягод населением области в 1997–2005 гг. (задача 4.4.1), можно определить ожидаемый уровень потребления в 2007 г. С этой целью подставим в найденное выше уравнение t = 11 (условный порядковый номер 2007 г.):

$$y_t = 22{,}1 + 0{,}3t = 22{,}1 + 0{,}3 \cdot 11 = 25{,}4 \text{ кг.}$$

Если в ряду динамики цепные абсолютные приросты приблизительно одинаковы, то неизвестные уровни ряда можно определить по формуле:

$$y_t = y_1 + \overline{\Delta y}(t-1),$$

где y_t – искомый уровень ряда; y_1 – начальный уровень ряда; $\overline{\Delta y}$ – средний абсолютный прирост ряда; t – порядковый номер искомого уровня ряда.

Если в ряду динамики цепные темпы роста мало отличаются друг от друга, то неизвестные уровни ряда можно определить по формуле:

$$y_t = y_1 \times \overline{T}p^{t-1},$$

где $\overline{T}p$ – средний темп роста ряда динамики.

Следует отметить приблизительность результатов, полученных при помощи этих методов интерполяции и экстраполяции, поскольку в действительности тенденция развития и аналитические характеристики ряда не остаются неизменными. Наибольшую осторожность следует проявлять при экстраполяции.

Решение типовых задач

Задача 4.6.1

Исчислите значение неизвестных уровней следующего ряда динамики для 2003 и 2007 гг.:

Год	2000	2001	2002	2003	2004	2005	2006	2007
Годовой выпуск продукции предприятием, млн. сом.	3,8	4,2	4,6	–	5,4	5,8	6,1	
Цепной абсолютный прирост, млн. сом.	–	0,4	0,4	–	–	0,4	0,3	
Цепной темп роста, %	–	110,5	109,5	–	–	107,4	105,2	
t	1	2	3	4	5	6	7	8

Решение

Для того, чтобы выбрать какой-либо из рассмотренных методов, исчислим по исходным данным цепные абсолютные приросты и цепные темпы роста. Пронумеруем уровни ряда, начиная с уровня 2000 г. (Вся эта информация представлена в таблице.)

Как видно из таблицы, в течение всего рассматриваемого периода уровни ряда изменялись в одном направлении, а цепные абсолютные приросты мало отличались один от другого. Цепные темпы роста, наоборот, существенно варьировали. Поэтому неизвестные уровни ряда определим по формуле:

$$y_t = y_1 + \overline{\Delta y}(t-1).$$

Средний абсолютный прирост за 2001–2006 гг. составляет:

$$\overline{\Delta y} = \frac{y_n - y_1}{n-1} = \frac{6{,}1 - 3{,}8}{7-1} = 0{,}4 \text{ млн. сом.}$$

Найдем искомые уровни ряда:

y_4 = 3,8 + 0,4(4 – 1) = 5,0 млн. сом.;

y_8 = 3,8 + 0,4(8 – 1) = 6,6 млн. сом.

Задача 4.6.2

Среднегодовой темп роста численности работников на промышленных предприятиях города составил за 2001–2005 гг. 103,2%. Определите методом экстраполяции численность работников на промышленных предприятиях города в 2007 г., если численность работников на этих предприятиях в 2000 г. составила 84 тыс. чел.

Решение

Если пронумеровать годы ряда динамики численности работников промышленных предприятий города, начиная с 2000 г., то порядковый номер 2007 г. окажется равным 8 (t = 8). Следовательно, численность работников промышленных предприятий города в 2007 г. составит:

$$y_t = y_1 \times \overline{T}p^{t-1} = 84 \times 1,032^{8-1} = 104,7 \text{ тыс. чел.}$$

4.7. Сезонные колебания и методы их изучения

Сезонными колебаниями называются относительно устойчивые внутригодовые колебания в ряду динамики. В большинстве отраслей экономики сезонность проявляется в виде внутригодовых чередований подъемов и спадов выпуска продукции, ее себестоимости, прибыли, производительности труда и других показателей. Наиболее заметно сезонность проявляется в сельском хозяйстве и соответственно в работе промышленных предприятий, перерабатывающих сельскохозяйственное сырье. Наблюдаются сезонные колебания и в других отраслях экономики – в строительстве, транспорте, торговле и т.д.

Сезонность – отрицательное явление. Из-за сезонности в отдельные периоды года простаивает оборудование и недоиспользуются прочие основные фонды предприятий, неравномерно используются трудовые ресурсы, возникает потребность в создании резервных мощностей, повышается себестоимость продукции, растут издержки обращения.

Статистическое изучение сезонности имеет важное практическое значение. Выявление закономерности развития явления во внутригодовой динамике позволяет создавать условия для сглаживания, уменьшения сезонных колебаний, позволяет прогнозировать развитие явления на перспективу.

Для количественной характеристики сезонных колебаний исчисляются индексы сезонности, которые представляют собой процентное отношение фактического уровня явления за тот или иной период года к выравненному уровню за этот же период. Обычно индексы сезонности рассчитываются в среднем за несколько лет (как правило, не менее трех). Это позволяет получить устойчивые индексы, свободные от влияния случайных особенностей отдельных лет. Существуют различные

методы исчисления индексов сезонности. Рассмотрим два метода; применение того или иного из них зависит от характера динамики изучаемого явления.

Если в ряду динамики нет четко выраженной тенденции к росту или убыванию, то индекс сезонности исчисляется по формуле:

$$I_s = \frac{\overline{y}_i}{\overline{y}} 100,$$

где $\overline{y}_i$ – уровень i-го внутригодового периода, осредненный за три или более лет; $\overline{y}$ – средний уровень для всего ряда динамики за три или более лет.

Если в ряду динамики ясно выражена тенденция к росту или убыванию, то индекс сезонности исчисляется по формуле:

$$I_s = \frac{1}{n} \sum \frac{y_i}{y_{ti}} 100,$$

где y_i – фактический i-й уровень ряда; y_{ti} – выравненный (или сглаженный) i-й уровень ряда; n – число анализируемых лет.

Совокупность индексов сезонности, исчисленных для внутригодовых периодов (например, для всех месяцев года), называется сезонной волной. На основании индексов сезонной волны рассчитывается обобщающий показатель, который называется коэффициентом сезонности:

$$K_s = \sqrt{\frac{\Sigma \left(I_s - 100 \right)^2}{m}},$$

где m – число внутригодичных интервалов (например, месяцев).

Коэффициент сезонности характеризует колеблемость индексов сезонности внутри года.

Рассмотрим пример исчисления индексов сезонности.

Задача 4.7.1

Имеются следующие данные о продаже товара «А» по кварталам года, тыс. сом.:

Квартал	2003 г.	2004 г.	2005 г.	В среднем за три года	Индексы сезонности, %
I	52	48	49	49,7	95,9
II	55	50	53	52,7	101,8
III	60	52	58	56,7	109,5
IV	49	46	50	48,3	93,2
В среднем за квартал	54	49	52	51,8	100,0

Для анализа внутригодовой динамики реализации товара «А» определите индексы сезонности.

Решение

Среднеквартальные уровни продажи товара за каждый год свидетельствуют о том, что четко выраженной тенденции роста явления в 2003–2005 гг. не было. Поэтому для расчета индексов сезонности воспользуемся формулой:

$$I_s = \frac{\overline{y}_i}{\overline{y}}100.$$

Определим для каждого квартала среднюю величину продажи товара за три года ($\overline{y}_i$) и общую среднюю для всего ряда ($\overline{y}$). Рассчитаем индексы сезонности и представим их в последней графе таблицы:

$$\frac{49,7}{51,8}100 = 95,9\%; \ \frac{52,7}{51,8}100 = 101,8\% \text{ и т.д.}$$

Полученные результаты ясно указывают на сезонность продажи товара «А»: с начала года наблюдается ежеквартальный рост продажи, достигающий максимума в третьем квартале (109,5% среднего уровня), а затем происходит уменьшение продажи до 93,2% от среднего уровня.

Глава 5

ИНДЕКСНЫЙ МЕТОД

5.1. Индексы и их классификация

Индексный метод – один из важнейших методов статистики. Индексы также используются при экономическом анализе работы предприятий и организаций, при планировании, при международных сравнениях экономических показателей, в социологии, в демографии и пр.

Слово “индекс” (указатель, показатель) употребляется в разных значениях. В статистике индекс – это относительный показатель, характеризующий соотношение уровней экономического (социального) явления во времени (динамические индексы) или в пространстве (территориальные индексы). К динамическим можно также отнести индексы, характеризующие изменение явления по сравнению с эталоном (нормативом, прогнозом, планом).

Индексы классифицируются по разным признакам: по охвату элементов совокупности, по характеру изучаемых явлений, по методам расчета. Рассмотрим эти классификации.

По охвату элементов совокупности различают индивидуальные и общие индексы. Индивидуальным называется индекс, характеризующий соотношение уровней признака одного элемента совокупности. Примером индивидуального индекса является показатель, характеризующий изменение во времени цены или объема продажи какого-либо одного товара. Индивидуальными индексами являются относительные показатели динамики, сравнения, выполнения плана.

Общим называется индекс, характеризующий изменение сложной статистической совокупности, состоящей из элементов, непосредственно несоизмеримых. Примером статистической совокупности, состоящей из непосредственно несоизмеримых (несуммируемых) элементов, может служить физический объем продукции промышленности или сельского хозяйства. Если индексы охватывают не все элементы совокупности, а только какую-то их часть, то они называются групповыми или субиндексами. Групповым является индекс количества продукции отдельной отрасли промышленности, индекс цен по группе продовольственных или по группе непродовольственных товаров.

При помощи общих индексов решаются следующие задачи:

1. Характеристика общего изменения сложного экономического или социального явления. Например, количественная характеристика общего изменения физического объема товарооборота универмага или среднего изменения себестоимости нескольких видов продукции завода.

2. Измерение влияния изменения отдельных факторов на динамику сложного явления. Индексным методом можно определить, в какой мере производство продукции на предприятии увеличилось за счет роста численности работников и в какой мере – за счет повышения производительности труда.

3. Определение влияния структурных сдвигов на динамику средней величины. Индексным методом можно определить, в какой мере рост средней заработной платы работников предприятия обусловлен увеличением заработной платы у отдельных групп работников и в какой мере средняя заработная плата изменилась за счет изменения доли отдельных групп работников в общей их численности.

При решении первой задачи индекс является синтетическим показателем, а при решении второй и третьей задач он играет роль аналитического показателя, характеризующего изменение результативного признака под влиянием изменения факторного признака.

В зависимости от характера изучаемого явления различают индексы объемных и качественных показателей. К индексам объемных показателей относятся индексы явлений, размеры которых характеризуются абсолютными величинами (индекс физического объема розничного товарооборота, индекс физического объема сельскохозяйственной продукции и др.). К индексам качественных показателей относятся индексы явлений, уровень которых характеризуется средними или относительными величинами (индекс цен, индекс себестоимости, индекс производительности труда и др.).

В зависимости от методики расчета общие индексы подразделяются на агрегатные и средние из индивидуальных индексов. Агрегатные индексы рассчитываются сопоставлением общих уровней сложного явления во времени или в пространстве. Средние индексы рассчитываются осреднением индивидуальных индексов.

Для исчисления динамических индексов, которые получили наибольшее применение в статистической практике, требуется наличие исходных данных за два периода. Один из этих периодов, с которым производится сравнение, называется базисным, а другой – текущим или отчетным. Если анализируются данные за три или более периодов, то базисным считается один из них, а все остальные – текущие. В этом случае индексы могут быть исчислены двумя методами: цепным и базисным.

Введем условные обозначения индексов и индексируемых признаков, которыми будем пользоваться далее при изложении материала. Общий индекс обозначается прописной буквой I, а индивидуальный – строчной буквой i. Признаки принято обозначать следующим образом:

q – количество товара или продукции в натуральном выражении (физический объем);

p – цена единицы товара или продукции;

z – себестоимость единицы продукции.

Внизу справа от символа признака приводится нумерация, означающая период, к которому относятся данные. Данные отчетного периода обозначаются 1, а базисного 0. Например:

p_0 – цены базисного периода;

p_1 – цены отчетного периода.

Внизу, справа от знака индекса, приводится символ индексируемого показателя, изменение которого является объектом исследования. Например:

i_p – индивидуальный индекс цены;

I_q – общий индекс физического объема (количества) товара или продукции;

I_{pq} – индекс товарооборота в фактических ценах.

Индексы выражаются в коэффициентах и процентах. Если условие решаемой задачи не требует большой точности, то расчеты индексов в коэффициентах производят с точностью до 0,001, а в процентах – с точностью до 0,1.

5.2. Агрегатные индексы

Рассмотрим первую из указанных выше трех задач индексного метода, когда индекс играет роль синтетического показателя и характеризует общее изменение сложного явления, состоящего из элементов, непосредственно несоизмеримых.

Для решения этой задачи применяются агрегатные индексы. Агрегатным индекс называется потому, что его числитель и знаменатель представляют собой наборы разнородных элементов, которые суммируются[1]. Латинское слово *aggregatus* означает складывание, суммирование. Как же сложить объемы разных видов товаров? Как, например, сложить проданные в магазине метры ткани с количеством телевизоров или фотоаппаратов? Чтобы привести разные товары к суммируемому

[1] Любой индекс представляет собой простую дробь, т.е. у него есть числитель и знаменатель.

виду, их количество умножают на коэффициенты соизмерения, в качестве которых обычно выступают цены (иногда себестоимость или трудоемкость); эта операция называется взвешиванием.

Рассмотрим построение и исчисление агрегатных индексов на примере решения конкретной задачи.

Задача 5.2.1

Имеются следующие данные за два периода об объеме продажи товаров и о ценах на них на городском рынке (табл. 5.2.1).

Требуется определить изменение в отчетном периоде по сравнению с базисным выручки, количества проданных товаров и цен по всем товарам вместе.

Решение

Поскольку три товара имеют разные единицы измерения, то для определения их общего объема, необходимо предварительно взвесить физический объем каждого из них на соответствующую цену[1]. Результаты расчета общей выручки от продажи товаров (общего объема товарооборота) для базисного и отчетного периодов представлены в табл. 5.2.1.

Выполненные расчеты позволяют определить индекс стоимости товарооборота в фактических ценах:

$$I_{pq} = \frac{26500}{25200} = 1,052, \text{ или } 105,2\%.$$

Воспользуемся условными обозначениями, о которых было сказано выше, и запишем формулу индекса товарооборота в фактических ценах:

$$I_{pq} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}, \qquad (5.2.1)$$

где $\Sigma p_1 q_1$ – товарооборот отчетного периода; $\Sigma p_0 q_0$ – товарооборот базисного периода.

Индекс товарооборота в фактических ценах показывает, во сколько раз возросла (уменьшилась) стоимость товарооборота или сколько процентов составляет ее рост (снижение) в отчетном периоде по сравнению с базисным. Если индекс больше единицы (больше 100%), то это говорит о росте изучаемого явления; если индекс меньше единицы (меньше 100%), то – о снижении.

---

[1] Количество видов товаров может быть любым, но не менее двух.

Таблица 5.2.1

Товар	Единица измерения	Количество проданных товаров		Цена за единицу измерения, сом.		Выручка (товарооборот), сом.		Индивидуальные индексы	
		базисный период q_0	отчетный период q_1	базисный период p_0	отчетный период p_1	базисный период p_0q_0	отчетный период p_1q_1	i_q	i_p
А	кг	320	250	50	60	16000	15000	0,781	1,200
Б	шт.	25	19	80	100	2000	1900	0,760	1,250
В	л	40	60	180	160	7200	9600	1,500	0,889
Итого						25200	26500		

Выполненные расчеты позволяют охарактеризовать изменение товарооборота в отчетном периоде по сравнению с базисным не только в относительном, но и в абсолютном выражении. Разность между числителем и знаменателем формулы (5.2.1) характеризует абсолютный прирост товарооборота:

$$\Delta T = \Sigma p_1 q_1 - \Sigma p_0 q_0 \,. \quad (5.2.2)$$

По данным задачи: ΔT = 26500 – 25200 = 1300 сом.

Следовательно, в отчетном периоде товарооборот в фактических ценах увеличился по сравнению с базисным периодом на 5,2%, или в абсолютном выражении на 1300 сом.

Как видно из формулы (5.2.1), величина индекса стоимости товарооборота зависит от изменения двух факторов: цены и количества товара. Чтобы ответить на второй вопрос задачи, т.е. охарактеризовать общее изменение количества всех товаров, требуется индекс, в числителе и знаменателе которого были бы одни и те же неизменные цены. Поэтому индекс количества товаров (индекс физического объема товарооборота) представляет собой соотношение выручки от продажи товаров в отчетном и базисном периодах в ценах базисного периода, так как только они могут быть неизменными. Формула агрегатного индекса количества товаров имеет вид:

$$I_q = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}, \quad (5.2.3)$$

где $\Sigma q_1 p_0$ – товарооборот отчетного периода, выраженный в ценах базисного периода.

Индекс, исчисленный по этой формуле, показывает во сколько раз изменился физический объем товарооборота или сколько процентов составляет его рост (снижение) в отчетном периоде по сравнению с базисным.

Разность между числителем и знаменателем формулы (5.2.3) представляет собой абсолютный прирост количества проданных товаров в ценах базисного периода:

$$\Delta T_q = \Sigma q_1 p_0 - \Sigma q_0 p_0 \,. \quad (5.2.4)$$

Подставив в формулы (5.2.3) и (5.2.4) исходные данные задачи, получим:

$$I_q = \frac{250 \cdot 50 + 19 \cdot 80 + 60 \cdot 180}{320 \cdot 50 + 25 \cdot 80 + 40 \cdot 180} = \frac{24820}{25200} = 0{,}985, \text{ или } 98{,}5\%;$$

ΔT_q = 24820 – 25200 = –380 сом.

Следовательно, в отчетном периоде было продано товаров меньше, чем в базисном периоде на 1,5%, или в абсолютном выражении на 380 сом.

Перейдем к решению третьего вопроса задачи. Суммировать цены на разные товары бессмысленно, но выручку от продажи товаров по фактическим ценам складывать можно. Поэтому, чтобы охарактеризовать общее изменение цен по группе товаров, следует сравнить выручку от продажи этих товаров в отчетном и базисном периодах при одинаковом количестве товара каждого вида в числителе и знаменателе индекса цен.

Если при построении формулы агрегатного индекса цен принять для отчетного и базисного периодов в качестве весов количества товаров, проданных в отчетном периоде, то формула будет иметь вид:

$$I_p = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}. \quad (5.2.5)$$

Эта формула предложена немецким экономистом Пааше и носит его имя. Разность между числителем и знаменателем индекса цен Пааше характеризует абсолютный прирост выручки продавцов при повышении цен или сумму экономии, полученную покупателями при снижении цен:

$$\Delta T_p = \Sigma p_1 q_1 - \Sigma p_0 q_1 \,. \quad (5.2.6)$$

Подставив в формулы (5.2.5) и (5.2.6) исходные данные задачи, получим:

$$I_p = \frac{60 \cdot 250 + 100 \cdot 19 + 160 \cdot 60}{50 \cdot 250 + 80 \cdot 19 + 180 \cdot 60} = \frac{26500}{24820} = 1{,}068, \text{ или } 106{,}8\%;$$

ΔT_p = 26500 – 24820 = 1680 сом.

Следовательно, в отчетном периоде цены на товары повысились в среднем на 6,8%, что дополнительно принесло продавцам товаров 1680 сом. Необходимо обратить внимание на то, что цены повысились именно “в среднем”. Как видно из табл. 5.2.1, по товарам “А” и “Б” они действительно выросли, причем гораздо больше, чем на 6,8%, а по товару “В” цены, наоборот, снизились.

Агрегатный индекс цен может быть построен иначе. Если принять для отчетного и базисного периодов в качестве весов количества товаров, проданных в базисном периоде, то формула агрегатного индекса цен будет иметь вид:

$$I_p = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}, \qquad (5.2.7)$$

где $\Sigma p_1 q_0$ – товарооборот базисного периода, выраженный в ценах отчетного периода.

Эта формула предложена немецким экономистом Ласпейресом. Индексы цен Пааше и Ласпейреса, исчисленные по одним и тем же данным, не совпадают по величине, так как имеют разное экономическое содержание. Индекс Пааше показывает, во сколько раз товары в отчетном периоде стали дороже (или дешевле), чем в базисном периоде. Индекс Ласпейреса показывает, во сколько раз товары, проданные в базисном периоде, подорожали (или подешевели) бы, если бы в базисном периоде действовали цены отчетного периода. Индекс Пааше позволяет исчислить фактическую, а индекс Ласпейреса условную экономию (перерасход) покупателей от изменения цен.

У каждого из этих двух индексов есть свои достоинства и недостатки. Выбор того или иного из них зависит от целей и задач статистического исследования. Так, при расчете индексов потребительских цен применяется индекс Ласпейреса, а для характеристики изменения розничных цен используется индекс Пааше.

Американский экономист Фишер предложил исчислять индекс цен как среднюю геометрическую из произведения двух агрегатных индексов цен Ласпейреса и Пааше:

$$I_p = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}. \qquad (5.2.8)$$

Существенным недостатком этой формулы является то, что она не имеет конкретного экономического содержания. Индекс Фишера не получил широкого применения в статистической практике.

Кроме рассмотренных, в статистике применяются и другие агрегатные индексы. В большинстве случаев они построены аналогично формулам (5.2.3) и (5.2.5), т.е. индексы объемных показателей с весами базисного периода, а индексы качественных показателей с весами отчетного периода. Так, например, выглядит формула агрегатного индекса себестоимости продукции:

$$I_z = \frac{\Sigma z_1 q_1}{\Sigma z_0 q_1}, \qquad (5.2.9)$$

где $\Sigma z_1 q_1$ – фактические затраты на производство продукции отчетного периода; $\Sigma z_0 q_1$ – затраты на производство продукции отчетного периода по себестоимости базисного периода.

Этот индекс характеризует среднее изменение себестоимости сравнимой продукции отчетного периода по отношению к базисному периоду. Разность между числителем и знаменателем формулы (5.2.9) характеризует экономию или перерасход издержек производства от изменения себестоимости продукции.

Аналогично построен агрегатный индекс трудоемкости продукции и многие другие индексы качественных показателей.

5.3. Индивидуальные и средние индексы

Используя условные обозначения индексируемых показателей, о которых было сказано выше, запишем формулы индивидуальных индексов:

индивидуальный индекс цены:

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

индивидуальный индекс количества (физического объема) товара или продукции):

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

индивидуальный индекс товарооборота:

$$i_{pq} = \frac{p_1 q_1}{p_0 q_0}.$$

Аналогично рассчитываются индивидуальные индексы других показателей.

В табл. 5.2.1 представлены результаты расчета индивидуальных индексов цены и количества товара по данным задачи предыдущего параграфа. Как видно из таблицы, по каждому товару и цена, и физический объем изменились. Так, по товару “А” цена выросла на 20%, а реализованный физический объем уменьшился на 21,9%. Очевидно, что вычисление индивидуальных индексов, представляющих собой отношение двух индексируемых величин, и объяснение полученных результатов осуществляется очень просто.

Индивидуальные индексы широко применяются в статистике не только для характеристики изменения величины признака у отдельного элемента совокупности, но и для характеристики изменения признака по всей совокупности элементов, т.е. для исчисления общих индексов, когда расчет агрегатного индекса невозможен. Так, если неизвестны цена и физический объем товара, но известны индивидуальные индексы и товарообороты по каждому виду товара, то общие индексы исчисляются как средние из индивидуальных. Средневзвешенные индексы цен и физического объема товаров получены преобразованием соответствующих агрегатных индексов. Рассмотрим методику этих преобразований.

Из формулы индивидуального индекса цен следует, что $p_0 = p_1/i_p$. Подставив правую часть этой формулы в знаменатель агрегатного индекса цен Пааше (формула 5.2.5), получим средний гармонический индекс цен:

$$I_p = \frac{\Sigma p_1 q_1}{\sum \frac{p_1 q_1}{i_p}}. \qquad (5.3.1)$$

Весами индивидуальных индексов цен в этой формуле служат товарообороты отчетного периода в фактических ценах (p_1q_1). Аналогично может быть преобразован агрегатный индекс себестоимости продукции.

Если формулу индивидуального индекса цен записать в виде $p_1 = i_p p_0$ и затем подставить правую часть этого равенства в числитель агрегатного индекса цен Ласпейреса [формула (5.2.7)], то получим средний арифметический индекс цен:

$$I_p = \frac{\Sigma i_p p_0 q_0}{\Sigma p_0 q_0}. \qquad (5.3.2)$$

В этой формуле индивидуальные индексы цен взвешены по величине товарооборота базисного периода (p_0q_0).

Из формулы индивидуального индекса количества товара следует, что $q_1 = i_q q_0$. Подставив i_qq_0 в числитель агрегатного индекса количества товаров [формула (5.2.3)], получим средний арифметический индекс количества (физического объема) товаров:

$$I_q = \frac{\Sigma i_q q_0 p_0}{\Sigma q_0 p_0}. \qquad (5.3.2)$$

В этой формуле, как и в предыдущей, в качестве весов использованы товарообороты базисного периода (q_0p_0).

Решение типовых задач

Задача 5.3.1

По данным задачи предыдущего параграфа (табл. 5.2.1) определите:
1) средний гармонический индекс цен;
2) средний арифметический индекс количества товаров;
3) средний арифметический индекс цен.

Решение

1) $$I_p = \frac{\Sigma p_1 q_1}{\sum \frac{p_1 q_1}{i_p}} = \frac{15000+1900+9600}{\frac{15000}{1,200}+\frac{1900}{1,250}+\frac{9600}{0,889}} = 1,068;$$

2) $$I_q = \frac{\Sigma i_q q_0 p_0}{\Sigma q_0 p_0} = \frac{0,781\cdot 16000 + 0,760\cdot 2000 + 1,500\cdot 7200}{16000+2000+7200} = 0,985.$$

Получены те же самые результаты, что и в предыдущем параграфе, но при этом не потребовались данные ни о количестве, ни о цене отдельных товаров.

3) $$I_p = \frac{\Sigma i_p p_0 q_0}{\Sigma p_0 q_0} = \frac{1,200\cdot 16000 + 1,250\cdot 2000 + 0,889\cdot 7200}{16000+2000+7200} = 1,115.$$

Индекс Ласпейреса всегда больше индекса Пааше.

Задача 5.3.2

Имеются следующие данные о продаже сельскохозяйственных товаров:

Товары	Продано в 1-м квартале, тыс. сом.	Изменение количества проданных товаров во 2-м квартале по сравнению с 1-м кварталом, %
Мясо	40	– 4
Овощи	130	+ 18

Определите общий индекс количества проданных товаров.

Решение

Примем второй квартал за отчетный период, а первый квартал за базисный. Чтобы иметь возможность воспользоваться формулой среднего арифметического индекса физического объема товаров, необходимо сначала определить индивидуальные индексы количества товара. Так как во втором квартале было продано мяса на 4% меньше, чем в первом квартале, то

i_q = 100 – 4 = 96%, или 0,96.

Для овощей i_q = 100 + 18 = 118%, или 1,18.

Вычислим средний арифметический индекс количества товаров:

$$I_q = \frac{\Sigma i_q q_0 p_0}{\Sigma q_0 p_0} = \frac{96 \cdot 40 + 118 \cdot 130}{40 + 130} = 112{,}8\%.$$

Физический объем товарооборота во втором квартале увеличился по сравнению с первым на 12,8%.

Задача 5.3.3

Имеются следующие данные о затратах на производство обуви и об изменении ее себестоимости:

Продукция	Затраты на производство обуви во 2-м квартале, тыс. сом.	Изменение себестоимости единицы изделия во 2-м квартале по сравнению с 1-м кварталом, %
Сапоги женские	125	+ 3
Ботинки мужские	93	– 9
Туфли детские	68	Без изменения

Определите:

1) среднее изменение себестоимости обуви во втором квартале по сравнению с первым кварталом;

2) абсолютную сумму экономии, полученную от снижения себестоимости продукции.

Решение

Примем второй квартал за отчетный период, а первый квартал – за базисный. Определим индивидуальные индексы себестоимости.

Себестоимость сапог повысилась на 3%, значит

i_z = 100 + 3 = 103%, или 1,03.

Себестоимость ботинок снизилась на 9%, следовательно

i_z = 100 – 9 = 91%, или 0,91.

Себестоимость туфель не изменилась, т.е. i_z = 100%, или 1.

Вычислим средний гармонический индекс себестоимости:

$$I_z = \frac{\Sigma z_1 q_1}{\sum \frac{z_1 q_1}{i_z}} = \frac{125 + 93 + 68}{\frac{125}{1,03} + \frac{93}{0,91} + \frac{68}{1,00}} = \frac{286,0}{291,6} = 0,981, \text{ или } 98,1\%.$$

Себестоимость производства обуви снизилась во втором квартале по сравнению с первым в среднем на 1,9%.

Абсолютная экономия, полученная от снижения себестоимости продукции, составила:

$$Э = \sum z_1 q_1 - \sum \frac{z_1 q_1}{i_z} = 286,0 - 291,6 = -5,6 \text{ тыс. сом.}$$

5.4. Индексный метод анализа факторов динамики сложных явлений

В этом параграфе рассматривается вторая из указанных выше трех задач, решаемых при помощи индексного метода, – оценка роли факторов в изменении сложного явления. Решение этой задачи имеет большое практическое значение. Многие статистические показатели взаимосвязаны. Эта взаимосвязь часто проявляется в том, что результативный показатель представляет собой произведение двух и более факторных показателей. (Зависимости такого рода называются мультипликативными.) Например, товарооборот является произведением количества реализованного товара и цены единицы товара; выпуск продукции предприятием равен произведению производительности труда рабочего и численности рабочих. Естественно, что при возрастании или убывании величины факторов-сомножителей результативный показатель тоже изменяется. Индексный метод позволяет выяснить, за счет каких факторов и на сколько процентов изменилось исследуемое явление. Зная ответы на эти вопросы, можно попытаться увеличить или уменьшить величину того или иного фактора-сомножителя, чтобы изменить результативный показатель в нужном направлении.

Применение индексов для измерения влияния отдельных факторов на динамику сложного явления основывается на следующих принципах:

1. При мультипликативной зависимости результативного показателя от факторных между индексами этих показателей существует такая же связь, как и между самими показателями, т.е. эти индексы образуют систему индексов.

2. Чтобы индексы были взаимосвязаны, т.е. образовывали систему, при индексировании качественного показателя объемный показатель

фиксируется на уровне отчетного периода, а при индексировании объемного показателя качественный показатель фиксируется на уровне базисного периода.

Между тремя индексами, рассмотренными во втором параграфе, существует следующая связь:

$$I_{pq} = I_p \times I_q. \quad (5.4.1)$$

В справедливости этой зависимости легко убедиться, если подставить в равенство (5.4.1) формулы соответствующих агрегатных индексов.

Проверим, увязываются ли в индексную систему три индекса, исчисленные при решении задачи 5.2.1:

$$I_{pq} = I_p \times I_q = 1,068 \times 0,985 = 1,052,$$

т.е. получен тот же результат, что и при расчете индекса товарооборота по формуле (5.2.1).

На основании этой зависимости по результатам решения задачи 5.2.1 можно сделать общий вывод. Увеличение в отчетном периоде товарооборота на 5,2% произошло за счет среднего повышения цен на товары на 6,8% при уменьшении количества проданных товаров на 1,5%.

Таким образом, общие индексы количества товаров и цен обладают не только свойствами синтетического показателя, но и свойствами аналитического показателя. Так, индекс количества товаров характеризует еще и влияние изменения количества товаров на динамику товарооборота, а индекс Пааше характеризует зависимость товарооборота от изменения цен.

Аналогичные индексные системы можно построить и для других мультипликативных зависимостей результативного показателя от факторных. Например, если затраты предприятия на производство продукции (c) равны произведению себестоимости единицы продукции (z) на количество единиц продукции (q), то:

$$I_c = I_z \times I_q, \quad (5.4.2)$$

где $I_c = \frac{\Sigma z_1 q_1}{\Sigma z_0 q_0}$ – индекс затрат предприятия на производство продукции;

$I_z = \frac{\Sigma z_1 q_1}{\Sigma z_0 q_1}$ – индекс себестоимости единицы продукции;

$I_q = \frac{\Sigma q_1 z_0}{\Sigma q_0 z_0}$ – индекс количества продукции.

Индексные системы позволяют не только определить влияние отдельных факторов на динамику результативного показателя, но и по двум известным индексам определить третий.

Между абсолютными приростами товарооборота также существует связь:

$$\Delta T = \Delta T_q + \Delta T_p, \quad (5.4.3)$$

где ΔT – общий абсолютный прирост товарооборота; ΔT_q – абсолютный прирост товарооборота из-за изменения количества проданного товара; ΔT_p – абсолютный прирост товарооборота за счет изменения цен.

В справедливости этой зависимости можно убедиться, если подставить в равенство (5.4.3) правую часть формул (5.2.2), (5.2.4) и (5.2.6).

Аналогично взаимосвязаны общий абсолютный прирост затрат на производство продукции (Δc) с абсолютными приростами затрат за счет изменения количества продукции (Δc_q), а также ее себестоимости (Δc_z):

$$\Delta c = \Delta c_q + \Delta c_z, \quad (5.4.4)$$

Каждый член равенства (5.4.4) представляет собой разность между числителем и знаменателем соответствующего индекса.

Проверим взаимосвязь абсолютных приростов товарооборота, исчисленных при решении задачи 5.2.1:

$$\Delta T = \Delta T_q + \Delta T_p = -380 + 1680 = 1300 \text{ сом.},$$

т.е. получен тот же результат, что и при расчете общего абсолютного прироста товарооборота по формуле (5.2.2). Следовательно, за счет роста цен товарооборот увеличился на 1680 сом., а за счет сокращения количества проданных товаров он уменьшился на 380 сом. Совместное действие этих двух факторов обусловило общее увеличение товарооборота на 1300 сом.

Рассмотренные индексные системы называются двухфакторными, так как они построены на основе зависимостей, у которых результативный показатель связан с двумя факторными. В тех случаях, когда требуется построить многофакторную индексную систему (более двух факторов), необходимо учитывать несколько дополнительных правил построения индексных моделей [1], рассмотрение которых не входит в задачи данного пособия.

Решение типовых задач

Задача 5.4.1

Как изменился товарооборот в фактических ценах, если цены были снижены на 4%, а физический объем товарооборота увеличился на 12%?

Решение

Используем взаимосвязь индексов (5.4.1):

$$I_{pq} = I_p \times I_q = 0,96 \times 1,12 = 1,075, \text{ или } 107,5\%.$$

Товарооборот в фактических ценах увеличился на 7,5%.

Задача 5.4.2

Как изменилась себестоимость единицы продукции, если индекс физического объема продукции составил 1,26, а издержки производства увеличились на 13%?

Решение

Используя взаимосвязь индексов (5.4.2), рассчитаем неизвестный показатель I_z:

$$I_z = \frac{I_c}{I_q} = \frac{1,13}{1,26} = 0,897, \text{ или } 89,7\%.$$

Себестоимость единицы продукции снизилась на 10,3%.

5.5. Индексы средних величин

Рассмотрим третью задачу, которая решается с помощью общих индексов: измерение влияния структурных сдвигов на динамику средней величины.

Уровень качественного показателя совокупности единиц характеризуется средней величиной, которая исчисляется по формуле:

$$\bar{x} = \frac{\Sigma xf}{\Sigma f} = \sum x \frac{f}{\Sigma f} = \Sigma xd,$$

где x – величина изучаемого признака у отдельных единиц совокупности; f – вес (частота) отдельных единиц совокупности; d – относительный показатель структуры совокупности единиц (частость, доля).

Следовательно, средняя величина качественного показателя зависит не только от величины признака у отдельных единиц совокупности, но и от структуры этой совокупности. Индекс средней величины характеризует влияние на ее динамику двух факторов: 1) изменение значений осредняемого показателя; 2) изменение структуры совокупности (структурные сдвиги). Чтобы определить влияние каждого из этих двух факторов в отдельности, применяется система взаимосвязанных индексов:

$$I_{пс} = I_{фс} \times I_{сс},$$

где $I_{пс}$ – индекс переменного состава; $I_{фс}$ – индекс фиксированного (постоянного) состава; $I_{сс}$ – индекс структурных сдвигов.

Индекс переменного состава характеризует динамику средней величины показателя как за счет изменения величины признака у отдельных единиц совокупности, так и за счет изменения структуры совокупности. Для любого качественного показателя формула индекса переменного состава имеет вид:

$$I_{пс} = \bar{x}_1 : \bar{x}_0 = \frac{\Sigma x_1 f_1}{\Sigma f_1} : \frac{\Sigma x_0 f_0}{\Sigma f_0}.$$

Индекс фиксированного состава характеризует среднее изменение величины признака по всей совокупности единиц за счет изменения величины признака у отдельных единиц при постоянной (фиксированной) структуре совокупности. Индекс фиксированного состава исчисляется по формуле:

$$I_{фс} = \frac{\Sigma x_1 f_1}{\Sigma f_1} : \frac{\Sigma x_0 f_1}{\Sigma f_1} = \frac{\Sigma x_1 f_1}{\Sigma x_0 f_1},$$

т.е. средняя величина отчетного периода делится на условную среднюю величину, исчисленную по базисным значениям признака у единиц совокупности отчетного периода.

Индекс структурных сдвигов характеризует динамику средней величины показателя за счет изменения структуры совокупности единиц при неизменных значениях признака у отдельных единиц. Индекс структурных сдвигов исчисляется по формуле:

$$I_{сс} = \frac{\Sigma x_0 f_1}{\Sigma f_1} : \frac{\Sigma x_0 f_0}{\Sigma f_0}.$$

Если вместо частот (f) известны частости (d), то формулы для расчета индексов переменного, фиксированного состава и структурных сдвигов принимают вид:

$$I_{пс} = \frac{\Sigma x_1 d_1}{\Sigma x_0 d_0}; \quad I_{фс} = \frac{\Sigma x_1 d_1}{\Sigma x_0 d_1}; \quad I_{сс} = \frac{\Sigma x_0 d_1}{\Sigma x_0 d_0}.$$

Решение типовых задач

Задача 5.5.1

Имеются следующие данные о реализации товара “А” двумя продавцами на городском рынке:

Продавцы	Цена 1 л, сом.		Продано, л	
	базисный период p_0	отчетный период p_1	базисный период q_0	отчетный период q_1
№ 1	8	10	400	700
№ 2	10	13	200	250

Определите:
1) индекс цен переменного состава;
2) индекс цен фиксированного состава;
3) индекс структурных сдвигов.

Решение

С учетом ранее принятых обозначений формула индекса цен переменного состава имеет вид:

$$I_{nc} = \frac{\Sigma p_1 q_1}{\Sigma q_1} : \frac{\Sigma p_0 q_0}{\Sigma q_0}.$$

Подставим данные в эту формулу:

$$I_{nc} = \frac{10 \cdot 700 + 13 \cdot 250}{700 + 250} : \frac{8 \cdot 400 + 10 \cdot 200}{400 + 200} = 1,245, \text{ или } 124,5\%.$$

Средняя по двум продавцам цена повысилась в отчетном периоде по сравнению с базисным на 24,5%.

Индекс цен фиксированного состава равен:

$$I_{\phi c} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} = \frac{10 \cdot 700 + 13 \cdot 250}{8 \cdot 700 + 10 \cdot 250} = 1,265, \text{ или } 126,5\%.$$

Цены на товар “А”, реализованный двумя продавцами, выросли в отчетном периоде по сравнению с базисным в среднем на 26,5%.

Индекс структурных сдвигов определим по формуле:

$$I_{cc} = \frac{\Sigma p_0 q_1}{\Sigma q_1} : \frac{\Sigma p_0 q_0}{\Sigma q_0} = \frac{8 \cdot 700 + 10 \cdot 250}{700 + 250} : \frac{8 \cdot 400 + 10 \cdot 200}{400 + 200} = 0,984, \text{ или } 98,4\%.$$

За счет изменения структуры объема продажи (т.е. изменения доли продавцов в общем объеме продажи) средняя цена товара “А” снизилась на 1,6%.

Произведем проверку полученных результатов:

$$1,265 \times 0,984 = 1,245.$$

Если бы структура объема продажи не изменилась, то средняя цена товара “А” выросла бы на 26,5%. Однако увеличение доли товара, реализованного продавцом № 1, у которого цены были ниже, обусловило рост средней цены на 24,5%.

Задача 5.5.2

Имеются следующие данные о выпуске одноименной продукции и ее себестоимости по двум заводам:

Завод	Себестоимость 1 шт., сом.		Структура выпуска, % к итогу	
	базисный период	отчетный период	базисный период	отчетный период
№ 1	20	17	54,6	42,9
№ 2	18	16	45,4	57,1

Определите:
1) индекс себестоимости переменного состава;
2) индекс себестоимости фиксированного состава;
3) индекс структурных сдвигов.

Решение

Индекс себестоимости переменного состава:

$$I_{nc} = \frac{\Sigma z_1 d_1}{\Sigma z_0 d_0} = \frac{17 \cdot 42,9 + 16 \cdot 57,1}{20 \cdot 54,6 + 18 \cdot 45,4} = 0,861, \text{ или } 86,1\%.$$

Средняя по двум заводам себестоимость единицы продукции снизилась на 13,9%.

Индекс себестоимости фиксированного состава:

$$I_{\phi c} = \frac{\Sigma z_1 d_1}{\Sigma z_0 d_1} = \frac{17 \cdot 42,9 + 16 \cdot 57,1}{20 \cdot 42,9 + 18 \cdot 57,1} = 0,871, \text{ или } 87,1\%.$$

В среднем по двум заводам себестоимость единицы продукции снизилась на 12,9%.

Индекс структурных сдвигов:

$$I_{cc}=\frac{\Sigma z_0 d_1}{\Sigma z_0 d_0}=\frac{20\cdot 42,9+18\cdot 57,1}{20\cdot 54,6+18\cdot 45,4}=0,988, \text{ или } 98,8\%.$$

Индекс структурных сдвигов можно определить с помощью взаимосвязи индексов:

$$I_{cc}=\frac{I_{пс}}{I_{фс}}=\frac{0,861}{0,871}=0,988, \text{ или } 98,8\%.$$

Средняя по двум заводам себестоимость единицы продукции снизилась на 1,2% за счет изменения доли заводов в общем объеме выпуска продукции.

5.6. Цепные и базисные индексы

При изучении динамики явления за три или более последовательных периодов приходится анализировать ряды индексов, которые могут быть рассчитаны как цепным, так и базисным способом. Цепные индексы позволяют следить за текущими изменениями изучаемого явления. Базисные индексы характеризуют изменения явления за длительный период времени по отношению к какому-либо одному отправному периоду.

Ряды формируются как из индивидуальных, так и из общих индексов. Индивидуальные индексы можно рассматривать как темпы роста. Поэтому между цепными и базисными индивидуальными индексами существуют те же зависимости, что и между цепными и базисными темпами роста:

1. Произведение последовательных цепных индексов равно базисному индексу последнего периода.

2. Отношение базисного индекса отчетного периода к предшествующему базисному индексу равно цепному индексу отчетного периода.

Эти зависимости позволяют исчислять базисные индексы по данным о цепных и наоборот.

Выясним на примере агрегатных индексов цен и физического объема, применимы ли эти два правила для рядов общих индексов.

В каждом отдельном агрегатном индексе и в числителе, и в знаменателе веса относятся к одному и тому же периоду. При исчислении агрегатных индексов за несколько последовательных отчетных периодов веса могут быть постоянными – одни и те же для всех периодов, или переменными – для каждого периода свои веса. Запишем формулы цепных и базисных агрегатных индексов цен для четырех периодов, один из которых является базисным (обозначим данные этого периода подстрочной цифрой 0), а три остальных являются отчетными (обозначим данные этих периодов подстрочными цифрами 1, 2, 3). Индексы цен, исчисленные по формуле Ласпейреса, образуют ряд с постоянными весами, а индексы цен, исчисленные по формуле Пааше, образуют ряд с переменными весами.

Цепные индексы цен:

с постоянными весами (по формуле Ласпейреса):

$$I_{p\,1/0}=\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0};\quad I_{p\,2/1}=\frac{\Sigma p_2 q_0}{\Sigma p_1 q_0};\quad I_{p\,3/2}=\frac{\Sigma p_3 q_0}{\Sigma p_2 q_0};$$

с переменными весами (по формуле Пааше):

$$I_{p\,1/0}=\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1};\quad I_{p\,2/1}=\frac{\Sigma p_2 q_2}{\Sigma p_1 q_2};\quad I_{p\,3/2}=\frac{\Sigma p_3 q_3}{\Sigma p_2 q_3}.$$

Базисные индексы цен:

с постоянными весами (по формуле Ласпейреса):

$$I_{p\,1/0}=\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0};\quad I_{p\,2/0}=\frac{\Sigma p_2 q_0}{\Sigma p_0 q_0};\quad I_{p\,3/0}=\frac{\Sigma p_3 q_0}{\Sigma p_0 q_0};$$

с переменными весами (по формуле Пааше):

$$I_{p\,1/0}=\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1};\quad I_{p\,2/0}=\frac{\Sigma p_2 q_2}{\Sigma p_0 q_2};\quad I_{p\,3/0}=\frac{\Sigma p_3 q_3}{\Sigma p_0 q_3}.$$

На основе построенных рядов убедимся в том, что для агрегатных индексов качественных показателей первое из приведенных выше правил взаимосвязи цепных и базисных индексов применимо, если ряды построены с постоянными весами, и не применимо, если ряды построены с переменными весами:

$$\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}\times\frac{\Sigma p_2 q_0}{\Sigma p_1 q_0}\times\frac{\Sigma p_3 q_0}{\Sigma p_2 q_0}=\frac{\Sigma p_3 q_0}{\Sigma p_0 q_0};$$

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}\times\frac{\Sigma p_2 q_2}{\Sigma p_1 q_2}\times\frac{\Sigma p_3 q_3}{\Sigma p_2 q_3}\neq\frac{\Sigma p_3 q_3}{\Sigma p_0 q_3}.$$

Справедливость аналогичного вывода по второму правилу читатель легко может проверить самостоятельно. Такие же выводы придется сделать, если рассмотреть взаимосвязь между цепными и базисными агрегатными индексами объемных показателей (см. задачу 5.6.1).

Таким образом, и для качественных, и для объемных показателей оба правила применимы для рядов агрегатных индексов с постоянными весами и не применимы для рядов агрегатных индексов с переменными весами.

Рассмотрим методику исчисления цепных и базисных индексов на примере.

Задача 5.6.1

Имеются следующие данные о ценах и количестве товаров, проданных на городском рынке:

Товар	Продано за квартал, ц			Цена 1 ц в квартале, сом.		
	I q_0	II q_1	III q_2	I p_0	II p_1	III p_2
А	150	200	300	50	40	30
Б	8	10	13	600	650	680

1. Определите цепные и базисные агрегатные индексы количества товаров с постоянными и переменными весами.

2. Проверьте взаимосвязь между вычисленными цепными и базисными индексами.

Решение

Вычислим цепные и базисные агрегатные индексы количества товаров с учетом того, что индексы объемных показателей принято исчислять по ценам базисного периода.

Цепные индексы с переменными весами:

$$I_{q_{1/0}} = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} = \frac{200 \cdot 50 + 10 \cdot 600}{150 \cdot 50 + 8 \cdot 600} = 1{,}3008;$$

$$I_{q_{2/1}} = \frac{\Sigma q_2 p_1}{\Sigma q_1 p_1} = \frac{300 \cdot 40 + 13 \cdot 650}{200 \cdot 40 + 10 \cdot 650} = 1{,}4103.$$

Цепные индексы с постоянными весами:

$$I_{q_{1/0}} = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} = \frac{200 \cdot 50 + 10 \cdot 600}{150 \cdot 50 + 8 \cdot 600} = 1{,}3008;$$

$$I_{q_{2/1}} = \frac{\Sigma q_2 p_0}{\Sigma q_1 p_0} = \frac{300 \cdot 50 + 13 \cdot 600}{200 \cdot 50 + 10 \cdot 600} = 1{,}4250.$$

Базисные индексы:

$$I_{q_{1/0}} = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} = \frac{200 \cdot 50 + 10 \cdot 600}{150 \cdot 50 + 8 \cdot 160} = 1{,}3008;$$

$$I_{q_{2/0}} = \frac{\Sigma q_2 p_0}{\Sigma q_0 p_0} = \frac{300 \cdot 50 + 13 \cdot 600}{150 \cdot 50 + 8 \cdot 600} = 1{,}8536.$$

Проверим взаимосвязь между цепными и базисными индексами:

с постоянными весами: 1,3008 × 1,4250 = 1,8536, т.е. произведение цепных индексов равно базисному;

с переменными весами: 1,3008 × 1,4103 = 1,8345, т.е. произведение цепных индексов не равно базисному (1,8536).

5.7. Территориальные индексы

Территориальные индексы позволяют сравнивать между собой экономические или социальные показатели предприятий, городов, областей и прочих территорий. Территориальные индексы так же, как и динамические, подразделяются на индивидуальные и общие. Индивидуальные территориальные индексы представляют собой относительные показатели сравнения, вычисление которых не вызывает трудностей. При вычислении общих территориальных индексов каждая из двух территорий может быть принята и в качестве сравниваемой, и в качестве базы сравнения. Поэтому веса в этих индексах могут быть разные. В зависимости от выбора весов получаются разные и даже противоречивые результаты. Один из способов устранения такой неоднозначности результатов заключается в применении следующих правил.

При вычислении агрегатных территориальных индексов качественных показателей за веса принимаются суммы соответствующих объемных показателей по обоим сравниваемым объектам. При вычислении агрегатных территориальных индексов объемных показателей за веса принимаются средние уровни соответствующих качественных показателей. Если при сравнении территорий А и Б руководствоваться этими правилами, то расчетные формулы будут иметь следующий вид:

Территориальный индекс цен:

$$I_{p_{А/Б}} = \frac{\Sigma p_А q}{\Sigma p_Б q}, \text{ или } I_{p_{Б/А}} = \frac{\Sigma p_Б q}{\Sigma p_А q},$$

где q – общее по двум территориям количество товара каждого вида:

$$q = q_А + q_Б.$$

Территориальный индекс количества товаров:

$$I_{q_{А/Б}} = \frac{\Sigma q_А \overline{p}}{\Sigma q_Б \overline{p}}, \text{ или } I_{q_{Б/А}} = \frac{\Sigma q_Б \overline{p}}{\Sigma q_А \overline{p}},$$

где $\overline{p}$ – средняя по двум территориям цена товара каждого вида:

$$\overline{p} = \frac{p_А q_А + p_Б q_Б}{q_А + q_Б}.$$

При использовании этой методики исчисления общих территориальных индексов результаты не противоречат один другому независимо от выбора базы сравнения. Рассмотрим пример исчисления территориальных индексов.

Задача 5.7.1

Имеются следующие данные о ценах и количествах реализации товаров на рынках двух городов:

Товар	Город А		Город Б	
	цена 1 кг, сом. $p_А$	продано, ц $q_А$	цена 1 кг, сом. $p_Б$	продано, ц $q_Б$
Картофель	6	20	7	10
Морковь	5	15	4	25

Определите:
1) территориальный общий индекс цен;
2) территориальный общий индекс количества товаров.

Решение

Территориальный индекс цен:

$$I_{p_{А/Б}} = \frac{\Sigma p_А q}{\Sigma p_Б q} = \frac{6 \cdot 30 + 5 \cdot 40}{7 \cdot 30 + 4 \cdot 40} = 1,027, \text{ или } 102,7\%.$$

Цены в городе А на 2,7% выше, чем в городе Б.

Вычислим средние по двум городам цены:

на картофель $$\overline{p}_к = \frac{6 \cdot 20 + 7 \cdot 10}{20 + 10} = 6,33 \text{ сом.},$$

на морковь $$\overline{p}_м = \frac{5 \cdot 15 + 4 \cdot 25}{15 + 25} = 4,38 \text{ сом.}$$

Территориальный индекс количества товаров:

$$I_{q_{А/Б}} = \frac{\Sigma q_А \overline{p}}{\Sigma q_Б \overline{p}} = \frac{20 \cdot 6,33 + 15 \cdot 4,38}{10 \cdot 6,33 + 25 \cdot 4,38} = 1,113, \text{ или } 111,3\%.$$

Следовательно, объем продажи товаров, выраженный в сопоставимых средних ценах, в городе А был на 11,3% больше, чем в городе Б.

Глава 6

ВЫБОРОЧНЫЙ МЕТОД

6.1. Генеральная и выборочная совокупности

Статистическое наблюдение, т.е. сбор статистических данных, бывает сплошным, когда обследуются все единицы изучаемой совокупности, и несплошным, когда обследуется часть единиц совокупности. Одной из разновидностей несплошного наблюдения является выборочное наблюдение, получившее широкое применение в статистической практике. Выборочным называется наблюдение, при котором характеристика всей совокупности единиц дается по результатам обследования некоторой их части, отобранной в случайном порядке.

Выборочное наблюдение имеет ряд преимуществ перед сплошным, важнейшими из которых являются:

- экономия времени и средств из-за сокращения объема работ;
- сведение к минимуму повреждения обследуемых единиц совокупности;
- достижение в некоторых случаях большей точности, чем при сплошном наблюдении, из-за уменьшения ошибок регистрации.

К этому следует добавить, что иногда вообще статистическое наблюдение можно провести только выборочным методом.

Вся совокупность единиц, из которой производится отбор некоторой части для обследования, называется генеральной. Часть генеральной совокупности, которая подвергается выборочному обследованию, называется выборочной. Соответственно показатели совокупности единиц (средняя, доля, дисперсия) в первом случае называются генеральными, а во втором – выборочными.

Задача выборочного метода заключается в том, чтобы на основе выборочных обобщающих показателей охарактеризовать обобщающие показатели генеральной совокупности. Расхождение между соответствующими обобщающими показателями выборочной и генеральной совокупностей называется ошибкой репрезентативности, или ошибкой выборки. Ошибки выборочного наблюдения подразделяются на случайные и систематические, преднамеренные и непреднамеренные[1]. Особую опасность для результатов выборочного исследования представляют преднамеренные систематические ошибки.

Для того чтобы выборка была репрезентативной, выборочная совокупность должна по своему составу и структуре соответствовать генеральной совокупности, т.е. выбранный массив данных должен представлять уменьшенную модель всей совокупности единиц. Для выполнения этого требования при организации выборочного наблюдения необходимо соблюдать принцип случайности отбора: все единицы генеральной совокупности должны обладать равновероятной возможностью попасть в выборочную совокупность.

В связи с тем что величина изучаемого признака у отдельных единиц генеральной совокупности варьирует и отбор единиц в выборочную совокупность производится случайно, существует вероятность, что состав и структура этих совокупностей могут несколько отличаться. Поэтому охарактеризовать обобщающие показатели генеральной совокупности через обобщающие показатели выборочной совокупности можно только с некоторой вероятностью.

Введем условные обозначения обобщающих показателей генеральной и выборочной совокупностей:

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

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

$p = \frac{M}{N}$ – генеральная доля (доля единиц, обладающих изучаемым признаком, в общей численности генеральной совокупности);

M – число единиц генеральной совокупности, обладающих изучаемым признаком;

N – общая численность единиц генеральной совокупности.

$\tilde{x}$ – выборочная средняя;

s^2 – выборочная дисперсия;

$w = \frac{m}{n}$ – выборочная доля;

m – число единиц выборочной совокупности, обладающих изучаемым признаком;

n – общая численность единиц выборочной совокупности.

[1] Подробнее об ошибках статистического наблюдения можно прочитать в теме “Статистическое наблюдение” [1, 6].

Между генеральной и выборочной средними величинами имеется зависимость:

$$\bar{x} = \tilde{x} \pm \Delta,$$

т.е. генеральная средняя находится в интервале:

$$\tilde{x} - \Delta \le \bar{x} \le \tilde{x} + \Delta,$$

где Δ – предельная ошибка выборки.

Генеральная и выборочная доли тоже связаны между собой:

$$p = w \pm \Delta,$$

т.е. генеральная доля находится в интервале:

$$w - \Delta \le p \le w + \Delta.$$

Предельная ошибка выборки исчисляется по формуле:

$$\Delta = t\mu,$$

где t – коэффициент доверия; μ – средняя ошибка выборки.

Величина коэффициента доверия t зависит от вероятности, с которой требуется определить границы интервала, где находится генеральный обобщающий показатель. В математической статистике доказывается, что при $t = 1$ вероятность равна 0,683, при $t = 2$ вероятность составляет 0,954, а при $t = 3$ она достигает 0,997. Вероятность, соответствующая другим значениям t, определяется при помощи специальных справочных таблиц [1; 3; 5].

Величина средней ошибки выборки и методика ее расчета зависят от способа отбора единиц из генеральной совокупности в выборочную.

Для того чтобы распространить результаты выборочного наблюдения на генеральную совокупность применяются способ прямого пересчета и способ поправочных коэффициентов. Способ прямого пересчета заключается в том, что выборочная средняя или доля выборочной совокупности умножается на численность единиц генеральной совокупности. В результате определяется доверительный интервал величины общего объема изучаемого признака. Способ поправочных коэффициентов применяется, когда возникает необходимость проверить или уточнить данные сплошного наблюдения. На основе выборочного обследования генеральной совокупности определяется так называемый “процент недоучета”, а затем данные сплошного наблюдения умножаются на этот поправочный коэффициент.

6.2. Способы отбора единиц из генеральной совокупности

Различают следующие основные способы отбора единиц из генеральной совокупности в выборочную: собственно-случайный, механический, типический и серийный, а также некоторые их разновидности и сочетания. Механический отбор всегда бесповторный, а все остальные способы отбора могут быть организованы в виде повторной и бесповторной выборки. При повторном отборе каждая единица, попавшая в выборочную совокупность, после регистрации ее признаков возвращается в генеральную совокупность и, следовательно, может быть отобрана несколько раз. При бесповторном отборе выбранная единица после регистрации в генеральную совокупность не возвращается. В практике социально-экономических исследований применяется в основном бесповторный отбор. Однако в некоторых случаях (обследование пассажиров, покупателей, пациентов) применение повторного отбора вполне целесообразно.

Рассмотрим сущность каждого из четырех основных способов отбора единиц в выборочную совокупность. Собственно-случайной называется выборка, при которой отбор единиц для обследования производится из всей генеральной совокупности посредством жеребьевки или по таблице случайных чисел. Типичным примером собственно-случайного отбора является тираж выигрышей лотереи, когда из общего количества номеров выпущенных билетов в случайном порядке отбирается определенная часть “выигравших” номеров. Недостаток этого способа заключается в том, что для каждой единицы генеральной совокупности требуется изготовление жребиев (фишек). При большой численности единиц генеральной совокупности это весьма трудоемко. К тому же численность единиц генеральной совокупности не всегда известна.

Механической называется выборка, при которой отбор единиц производится через определенный интервал в порядке расположения единиц в генеральной совокупности или в каком-либо перечне. При этом способе отбора все единицы генеральной совокупности предварительно располагаются по случайному признаку, нейтральному по отношению к изучаемому показателю. Затем в соответствии с долей отбора вся совокупность разбивается на равные по численности группы и из каждой группы отбирается одна единица. Так, если требуется провести 10%-ную выборку, то отбирается каждая десятая единица, начиная с любой единицы из первого десятка (например, 5, 15, 25, 35 и т.д.). При 5%-ной выборке отбирается каждая двадцатая единица, а при 2%-ной – каждая пятидесятая.

Если выборочная совокупность велика, то при механическом отборе ошибка выборки не больше, чем при собственно-случайном. При механической выборке легче проверить правильность отбора единиц из генеральной совокупности.

Типический отбор применяется, когда статистическая совокупность неоднородна по изучаемому признаку. В этом случае генеральная

совокупность единиц предварительно распределяется на однородные типические группы по признаку, от которого зависит исследуемый показатель. Затем из каждой группы механическим или собственно-случайным способом отбирается определенное количество единиц. Типический отбор может быть пропорциональным и непропорциональным. При пропорциональном отборе из каждой типической группы отбирается число единиц, пропорциональное доле этой группы в численности генеральной совокупности. Если это условие не соблюдается, то типический отбор будет непропорциональным.

Типический отбор дает более репрезентативную выборку, чем механический или собственно-случайный, так как обеспечивает представительство каждой типической группы в выборочной совокупности.

Серийной называется выборка, при которой из генеральной совокупности отбираются не отдельные единицы, а их группы (серии). Отбор отдельных серий производится собственно-случайным или механическим способом. Внутри каждой отобранной серии обследуются все единицы. Серийный отбор применяется, например, при проверке качества товара, упакованного в пачках, коробках, ящиках и т.п. В этом случае проще и дешевле полностью проверить несколько стандартных упаковок (серий), чем отбирать для проверки некоторое количество единиц товара из каждой упаковки.

Серийная выборка дает более значительную ошибку репрезентативности, чем другие способы отбора. Это объясняется тем, что состав единиц отдельных серий может существенно отличаться от состава единиц генеральной совокупности.

6.3. Средняя ошибка выборки при собственно-случайном или механическом отборах единиц

Расхождение между обобщающими показателями генеральной и выборочной совокупностей характеризуется средней ошибкой выборки. В курсе математической статистики доказывается, что при собственно-случайном повторном отборе средняя ошибка выборки рассчитывается по формуле:

для средней величины количественного признака

$$\mu_x = \sqrt{\frac{\sigma^2}{n}};$$

для доли альтернативного признака

$$\mu_p = \sqrt{\frac{p(1-p)}{n}}.$$

Как видно из этих формул, величина средней ошибки выборки зависит от вариации изучаемого признака и от численности выборочной совокупности. Чем меньше дисперсия и чем больше численность выборки, тем меньше средняя ошибка.

Дисперсия признака в генеральной совокупности обычно неизвестна, поэтому при практических расчетах ее заменяют выборочной дисперсией, которую исчисляют по результатам обследования отобранных единиц. Между генеральной и выборочной дисперсиями имеется соотношение:

$$\sigma^2 = s^2 \frac{n}{n-1}.$$

Поскольку при большой численности выборки (n > 30) величина $n/(n-1)$ близка к единице, то принимается $\sigma^2 \approx s^2$ и $p(1-p) \approx w(1-w)$. Поэтому для решения практических задач используются следующие формулы средней ошибки выборки при собственно-случайном повторном отборе:

для средней величины количественного признака

$$\mu_x = \sqrt{\frac{s^2}{n}};$$

для доли альтернативного признака

$$\mu_w = \sqrt{\frac{w(1-w)}{n}}.$$

Расчетные формулы средней ошибки выборки при собственно-случайном бесповторном отборе имеют вид:

для средней величины количественного признака

$$\mu_x = \sqrt{\frac{s^2}{n}\left(1-\frac{n}{N}\right)};$$

для доли альтернативного признака

$$\mu_w = \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}.$$

Из сравнения формул средней ошибки выборки при повторном и бесповторном отборах видно, что при бесповторном отборе ошибка всегда меньше, так как величина $(1-n/N)$ меньше единицы. При небольшом проценте выборки множитель $(1-n/N)$ близок по величине к единице и им можно пренебречь. Поэтому для определения средней ошибки выборки при бесповторном отборе можно применять формулы повторного отбора, если n составляет не более 5% от N. Поправкой $(1-n/N)$ можно пренебречь и в том случае, когда численность единиц генеральной совокупности неизвестна или очень велика.

Для расчета средней ошибки выборки при механическом отборе единиц применяются те же самые формулы, что и при собственно-случайном отборе[1].

Решение типовых задач

Задача 6.3.1

Для определения среднего возраста рабочих предприятия было опрошено 100 рабочих, которых отобрали методом случайного бесповторного 10%-ного отбора. Установлено, что средний возраст рабочих в выборочной совокупности составил 34 года при среднем квадратическом отклонении 11 лет.

Определите с вероятностью 0,997 пределы, в которых находится средний возраст всех рабочих предприятия.

Решение

Определим среднюю ошибку выборки:

$$\mu_x = \sqrt{\frac{s^2}{n}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{11^2}{100}(1-0,1)} \approx 1 \text{ год}.$$

Вероятности 0,997 соответствует значение t = 3.

Следовательно, предельная ошибка выборки при определении среднего возраста рабочих предприятия составит:

$$\Delta = t\mu = 3 \times 1 = 3 \text{ года}.$$

Средний возраст всех рабочих предприятия будет находиться в пределах (лет):

$$34 - 3 \le \bar{x} \le 34 + 3.$$

Таким образом, с вероятностью 0,997 можно утверждать, что средний возраст всех рабочих предприятия составляет от 31 до 37 лет.

[1] Методики применения типического и серийного отборов единиц в настоящем пособии не рассматриваются, интересующиеся этими вопросами отсылаются к дополнительной литературе [1; 6].

Задача 6.3.2

Из 4000 штук изготовленных на заводе деталей было отобрано механическим способом 200 штук для проверки их соответствия стандарту. Среди проверенных деталей 8 штук были забракованы.

Определите с вероятностью 0,954 пределы, в которых находится доля брака среди всей партии изготовленных деталей.

Решение

Определим долю брака среди проверенных деталей:

$$w = \frac{m}{n} = \frac{8}{200} = 0,04, \text{ или } 4\%.$$

Так как выборка механическая, то среднюю ошибку выборки определим по формуле для бесповторного отбора:

$$\mu_w = \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{0,04(1-0,04)}{200}\left(1-\frac{200}{4000}\right)} = 0,014.$$

Вероятности 0,954 соответствует значение $t = 2$.

Предельная ошибка доли брака составит:

$$\Delta = t\mu = 2\times 0,014 = 0,028, \text{ или } 2,8\%.$$

Таким образом, доля брака среди всей партии изготовленных деталей находится в пределах, %:

$$4-2,8 \le \bar{x} \le 4+2,8.$$

Следовательно, с вероятностью 0,954 можно гарантировать, что доля брака среди всей партии изготовленных деталей составляет от 1,2 до 6,8%.

6.4. Определение необходимой численности выборки

Как уже отмечено выше, с ростом численности выборочной совокупности величина ошибки выборки уменьшается. Однако при увеличении численности выборки возрастают затраты труда и средств на сбор информации. Поэтому при организации выборочного наблюдения необходимо определить численность выборки, которая с заданной вероятностью обеспечит достаточную точность результатов исследования при оптимальных затратах.

При собственно-случайном повторном отборе численность выборки рассчитывается по формуле:

для определения средней величины количественного признака

$$n = \frac{t^2 s^2}{\Delta^2};$$

для определения доли альтернативного признака

$$n = \frac{t^2 w(1-w)}{\Delta^2}.$$

При механическом или собственно-случайном бесповторном отборе численность выборки рассчитывается по формуле:

для определения средней величины количественного признака

$$n = \frac{t^2 s^2 N}{\Delta^2 N + t^2 s^2};$$

для определения доли альтернативного признака

$$n = \frac{t^2 w(1-w) N}{\Delta^2 N + t^2 w(1-w)}.$$

Если планируется, что доля отобранных единиц не должна превышать 5%, то при расчете численности выборки для бесповторного отбора можно применить формулу для повторного отбора, тем самым незначительно завышая число единиц выборочной совокупности и увеличивая точность результатов наблюдения.

Во всех этих формулах значение допустимой предельной ошибки и уровень вероятности (или соответствующее значение *t*) задаются исследователем. Дисперсия принимается на основе проводимых ранее обследований данной совокупности или же на основе предварительного выборочного обследования небольшого числа ее единиц. Если у исследователя нет необходимых данных для расчета дисперсии альтернативного признака, то принимается максимально возможное ее значение – 0,25.

Решение типовых задач

Задача 6.4.1

С целью определения средних затрат времени при поездках на работу населением города планируется выборочное наблюдение на основе случайного повторного отбора. Сколько людей должно быть обследовано, чтобы с вероятностью 0,954 ошибка выборочной средней не превышала 1 мин. при среднем квадратическом отклонении 15 мин.?

Решение

Вероятности 0,954 соответствует t = 2. Так как планируется случайный повторный отбор, то численность выборки определим по формуле:

$$n = \frac{t^2 s^2}{\Delta^2} = \frac{2^2 \cdot 15^2}{1^2} = 900 \text{ чел.}$$

Следовательно, чтобы с вероятностью 0,954 гарантировать, что ошибка выборочной средней не превысит 1 мин., необходимо опросить 900 жителей города.

Задача 6.4.2

В городе, где проживает 10 тыс. семей, предполагается на основе механической выборки определить долю семей с детьми школьного возраста. Какова должна быть численность выборки, чтобы с вероятностью 0,997 ошибка выборки не превышала 5%?

Решение

Вероятности 0,997 соответствует t = 3. Поскольку по данным условия задачи определить дисперсию альтернативного признака невозможно, то примем максимально возможное ее значение, т.е. $w(1-w) = 0,25.$ Так как выборка механическая, то ее численность определим по формуле для бесповторного отбора:

$$n = \frac{t^2 w(1-w) N}{\Delta^2 N + t^2 w(1-w)} = \frac{3^2 \times 0,25 \times 10000}{0,05^2 \times 10000 + 3^2 \times 0,25} = 826 \text{ семей.}$$

Следовательно, чтобы с вероятностью 0,997 утверждать, что ошибка выборочной доли не превысит 5%, необходимо опросить 826 семей.

Глава 7

СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ВЗАИМОСВЯЗИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

7.1. Виды связей между явлениями

Любое социально-экономическое явление связано с рядом различных факторов, зависит от их воздействия. Одна из важнейших задач статистики – выявить, какие факторы и как влияют на изучаемое социально-экономическое явление, измерить степень их воздействия.

Связи между явлениями могут быть функциональными и стохастическими. Функциональной называется связь, при которой каждому значению факторного признака соответствует определенное значение результативного признака. Примерами функциональных связей могут служить зависимости между радиусом и площадью круга, между производительностью труда работника и количеством произведенной им продукции. Стохастической называется связь, при которой зависимость результативного признака от факторного носит вероятностный, случайный характер; при изменении факторного признака меняется распределение значений результативного признака. В отличие от функциональной стохастическая связь проявляется не в каждой отдельной единице, а в совокупности единиц.

Частным случаем стохастической связи является корреляционная связь, при которой изменение факторного признака – причина изменения среднего значения результативного признака. Примером корреляционной связи может служить зависимость веса человека от его роста: чем выше человек, тем больше его вес. Существуют различные формулы для определения веса человека по его росту. Однако расчеты по этим формулам не дают точных результатов, так как вес человека зависит еще и от других факторов: пола, возраста, состояния здоровья, типа питания и т.д. Каждый из этих факторов у разных людей может по-разному влиять на величину веса, да и сочетаться между собой факторы могут по-разному. Поэтому люди с одинаковым ростом имеют разный вес. Этот пример указывает на важнейшую особенность корреляционных связей – они неполные, так как не учитывают всех факторов, воздействующих на результативный признак. Еще одна особенность корреляционных связей заключается в том, что им свойственна необратимость зависимостей. Если существует корреляционная зависимость веса человека от его роста, то это не означает наличие зависимости роста от веса. Увеличение веса человека, выздоравливающего после болезни, не отразится на его росте; при неизменном росте можно существенно увеличить вес за счет обильного питания и т.д.

И функциональные, и корреляционные связи бывают прямыми и обратными. Прямой называется связь, при которой с возрастанием факторного признака результативный признак также возрастает. Обратной называется связь, при которой с возрастанием факторного признака результативный признак уменьшается.

По аналитическому выражению различают прямо- и криволинейные связи. Прямолинейной называется связь, которая может быть выражена уравнением прямой линии. Все остальные связи криволинейные.

Корреляционные связи делятся на парные (простые) и множественные. Парная корреляция характеризует зависимость результативного признака от одного, а множественная от нескольких факторных признаков.

7.2. Методы изучения стохастических связей между явлениями

При изучении стохастических связей применяются методы параллельных рядов и аналитических группировок, корреляционно-регрессионный анализ и другие методы.

Метод параллельных рядов заключается в сопоставлении двух рядов статистических показателей. Данные, характеризующие факторный признак, располагаются в виде ряда в возрастающем или убывающем порядке. Параллельно приводится ряд соответствующих значений предполагаемого результативного признака. Сопоставляя тенденцию изменения результативного признака с изменениями значений факторного признака, можно выявить наличие связи между явлениями и ее направление (или отсутствие связи).

Применение метода группировок при изучении стохастических связей заключается в построении аналитической группировки единиц совокупности на основе факторного признака и вычислении средней величины результативного признака для каждой группы (см. главу 1, §2). Характер изменения средних величин результативного признака с возрастанием факторного признака укажет на наличие связи между явлениями и ее направление (или на отсутствие связи). При использовании этого метода корреляционная связь проявляется более четко, чем при методе параллельных рядов, так как групповые средние меньше подвержены влиянию случайных факторов.

В отличие от рассмотренных методов корреляционно-регрессионный анализ позволяет не только выявить наличие и направление связи, но и количественно оценить тесноту связи, определить форму (аналитическое выражение) связи. Сущность этого метода заключается в решении следующих задач: 1) построение и анализ уравнения связи результативного признака с одним или несколькими факторными признаками; 2) измерение тесноты связи между результативным и факторным признаками. Использование корреляционно-регрессионного анализа предполагает наличие ряда условий [6]. Одно из условий заключается в том, что факторный и результативный признаки должны иметь количественное выражение.

7.3. Уравнение однофакторной регрессии

Уравнение, характеризующее корреляционную связь между признаками, называется уравнением регрессии. Парная корреляционная связь между результативным и факторным признаками характеризуется уравнением однофакторной регрессии. В процессе построения уравнения однофакторной регрессии можно выделить три последовательных этапа: 1) предварительный анализ свойств исследуемой совокупности единиц; 2) выбор типа уравнения регрессии; 3) вычисление значений параметров выбранного уравнения.

Рассмотрим каждый из этих этапов.

Первый этап. На основе положений экономической теории, конкретной экономики или социологии выясняют, возможна ли связь между изучаемыми показателями в принципе. Анализ сущности изучаемого явления позволяет выявить причинную связь между признаками, т.е. определить, какой признак является факторным и какой результативным. Затем решается вопрос о характере корреляции: парная или множественная, прямая или обратная. На этом этапе анализа применяют метод параллельных рядов и метод аналитических группировок.

Второй этап. Требуется определить тип аналитической функции (прямая, гипербола, парабола, степенная, показательная или какая-нибудь другая), наилучшим образом отражающей связь между факторным и результативным признаками. Выбор типа уравнения регрессии осуществляется на основе сочетания теоретического анализа сущности изучаемой зависимости и исследования исходных эмпирических данных. На этом этапе важное значение имеет графический метод исследования, т.е. изображение изучаемой зависимости в прямоугольной системе координат. Существуют рекомендации, позволяющие выбрать тип уравнения связи в зависимости от характера изменения исходных данных. Так, если с возрастанием значений факторного признака наблюдается пропорциональное возрастание или убывание результативного признака, то связь линейная. Если при равномерном возрастании факторного признака наблюдается ускоренное изменение результативного признака, то можно выбрать уравнение параболы второго порядка. (Рекомендации по выбору уравнений другого типа приводятся в справочной литературе.)

На практике при изучении связей между общественными явлениями часто используется уравнение прямой. Нередко прямолинейная связь достаточно хорошо выражает зависимость между изучаемыми явлениями даже тогда, когда на самом деле связь между ними оказывается более сложной. Это объясняется тем, что в определенном интервале значений факторного признака самые сложные зависимости могут носить приближенно линейный характер.

Уравнение прямолинейной регрессии имеет вид:

$$\bar{y}_x = a_0 + a_1 x,$$

где $\bar{y}_x$ – результативный признак; x – факторный признак; a_0, a_1 – параметры уравнения прямой.

Третий этап. Для определения числовых значений параметров выбранного уравнения применяется метод наименьших квадратов. Сущность этого метода заключается в нахождении таких постоянных параметров уравнения регрессии, при которых минимизируется сумма квадратов отклонений фактических значений результативного признака (y) от соответствующих значений, рассчитанных по выбранному уравнению связи ($\bar{y}_x$):

$$\sum \left(y - \bar{y}_x \right)^2 \to \min .$$

Параметры a_0 и a_1 уравнения прямолинейной регрессии, соответствующие требованию метода наименьших квадратов, находятся решением системы двух нормальных уравнений:

$$\begin{cases} a_0 n + a_1 \Sigma x = \Sigma y; \\ a_0 \Sigma x + a_1 \Sigma x^2 = \Sigma xy, \end{cases}$$

где n – число единиц исследуемой совокупности.

В линейном уравнении регрессии параметр a_0 характеризует усредненное влияние на результативный признак факторов, неучтенных при исследовании изучаемой парной связи. Параметр a_1 называется линейным коэффициентом регрессии, он показывает, насколько в среднем изменяется результативный признак при изменении факторного признака на единицу.

С методикой построения однофакторных уравнений криволинейной связи и уравнений множественной регрессии можно ознакомиться в учебниках теории статистики [1; 2; 6].

Рассмотрим построение уравнения однофакторной регрессии на примере.

Задача 7.3.1

Имеются следующие данные о сменной выработке рабочих бригады и производственном стаже их работы (см. табл.). Определите вид корреляционной зависимости и постройте уравнение регрессии.

№ п/п	Стаж работы, лет x	Выработка рабочего за смену, шт. y	x^2	xy	$\bar{y}_x$
1	2	8	4	16	8,14
2	3	10	9	30	9,43
3	4	11	16	44	10,72
4	5	11	25	55	12,01
5	6	13	36	78	13,30
6	7	15	49	105	14,59
7	8	16	64	128	15,88
Итого	35	84	203	456	84,07

Решение

Исходя из экономической сущности исследуемых признаков, можно считать, что производственный стаж работы – это факторный признак (x), а выработка рабочего за смену – результативный (y). Сопоставление параллельных рядов значений этих признаков показывает, что между ними существует прямая связь. Поскольку с ростом стажа работы сменная выработка рабочих возрастала почти пропорционально, можно предположить, что связь между признаками прямолинейная. Это подтверждает графическое изображение исследуемой зависимости в прямоугольной системе координат.

Для того чтобы определить значения параметров уравнения прямолинейной регрессии, составим систему нормальных уравнений на основе расчетных данных, представленных в таблице:

$$\begin{cases} 7a_0 + 35a_1 = 84; \\ 35a_0 + 203a_1 = 456. \end{cases}$$

Решив эту систему уравнений, получим:

$a_0 = 5{,}56;\ a_1 = 1{,}29.$

Следовательно, $\bar{y}_x = 5{,}56 + 1{,}29x.$

Это уравнение показывает, что с увеличением стажа работы на один год сменная выработка рабочего бригады возрастала в среднем на 1,3 изделия. Подставляя в это уравнение конкретные значения *x*, найдем для каждого из 7 рабочих теоретическую сменную выработку (см. последнюю графу таблицы). Согласно методу наименьших квадратов, суммы эмпирических и теоретических значений результативного признака должны быть одинаковы; в рассматриваемом примере они незначительно отличаются из-за округления расчетных данных.

7.4. Измерение тесноты парной корреляционной связи

Для измерения тесноты корреляционной связи между факторным и результативным признаками применяются теоретическое корреляционное отношение, линейный коэффициент корреляции и другие показатели.

Теоретическое корреляционное отношение применяется для измерения тесноты связи между признаками при любой математической форме уравнения регрессии. Рассчитывается этот показатель по формулам:

$$\eta = \sqrt{\frac{\sigma^2_{y_x}}{\sigma^2_y}} = \sqrt{\frac{\sum(\bar{y}_x - \bar{y})^2}{\sum(y - \bar{y})^2}},$$

где $\sigma^2_{y_x} = \frac{\sum(\bar{y}_x - \bar{y})^2}{n}$ – дисперсия, характеризующая вариацию результативного признака под влиянием фактора *x*; $\sigma^2_y = \frac{\sum(y - \bar{y})^2}{n}$ – дисперсия, характеризующая вариацию результативного признака под влиянием всех факторов, вызвавших эту вариацию; *y* – фактические значения результативного признака; $\bar{y}$ – средняя величина фактических значений результативного признака; $\bar{y}_x$ – теоретические значения результативного признака, рассчитанные по уравнению регрессии; *n* – число единиц исследуемой совокупности.

Как видно из этих формул, теоретическое корреляционное отношение можно вычислить только после того, как построено уравнение регрессии и рассчитаны теоретические значения результативного признака.

Теоретическое корреляционное отношение изменяется от 0 до 1. Если $\eta = 0$, то связь между факторным и результативным признаками отсутствует; если $\eta = 1$, то связь функциональная. Чем ближе корреляционное отношение к единице, тем теснее связь между признаками.

В тех случаях, когда корреляционная связь между признаками прямолинейная, для измерения тесноты связи можно воспользоваться линейным коэффициентом корреляции. Вычислить коэффициент корреляции гораздо проще, чем теоретическое корреляционное отношение, так как не нужно предварительно определять уравнение регрессии и рассчитывать теоретические значения результативного признака. При строго прямолинейной зависимости числовые значения корреляционного отношения и коэффициента корреляции совпадают. Если же эти показатели различаются, то связь между изучаемыми признаками криволинейная.

Для расчета линейного коэффициента корреляции существуют различные формулы, которые легко трансформируются друг в друга. Одна из наиболее применяемых формул линейного коэффициента корреляции имеет вид:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y},$$

где $\overline{xy} = \frac{\Sigma xy}{n}$ – средняя величина из попарных произведений изучаемых признаков;

$\bar{x} = \frac{\Sigma x}{n}$ – средняя величина факторного признака;

$\bar{y} = \frac{\Sigma y}{n}$ – средняя величина результативного признака;

$\sigma_x = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}$ – среднее квадратическое отклонение факторного признака;

$\sigma_y = \sqrt{\frac{\sum(y - \bar{y})^2}{n}}$ – среднее квадратическое отклонение результативного признака;

Линейный коэффициент корреляции может принимать значения от –1 до +1. Знак “минус” указывает на обратную связь между признаками, а знак “плюс” – на прямую. При $|r| = 1$ имеется функциональная связь между признаками, а при $r = 0$ связь отсутствует. Чем ближе $|r|$ к единице, тем теснее связь между результативным и факторным признаками.

Квадрат линейного коэффициента корреляции называется коэффициентом детерминации, который показывает, какая часть общей вариации результативного признака объясняется данным фактором. Коэффициент детерминации может принимать значения от 0 до 1.

Для качественной оценки тесноты связи между признаками можно пользоваться следующей таблицей:

Теснота связи	До 0,3	0,3–0,5	0,5–0,7	0,7–0,9
Характеристика связи	практически отсутствует	слабая	умеренная	сильная

С методикой измерения тесноты связи при множественной корреляции можно ознакомиться в учебниках теории статистики [1; 2; 6].

Рассмотрим расчет теоретического корреляционного отношения и линейного коэффициента корреляции на примере.

Задача 7.4.1

По данным задачи 7.3.1 определите тесноту связи между сменной выработкой рабочих и их производственным стажем.

Решение

Для расчета теоретического корреляционного отношения необходимо предварительно определить $\sum(\bar{y}_x - \bar{y})^2$ и $\sum(y-\bar{y})^2$. Вычисление этих величин представлено в нижеследующей таблице. Теоретическое корреляционное отношение составляет:

$$\eta = \sqrt{\frac{\sum(\bar{y}_x - \bar{y})^2}{\sum(y-\bar{y})^2}} = \sqrt{\frac{46,59}{48}} = 0,985.$$

Поскольку предполагается, что связь между признаками прямолинейная (см. решение задачи 7.3.1), то тесноту связи между ними можно охарактеризовать величиной линейного коэффициента корреляции. Для расчета этого коэффициента предварительно определим $\bar{x}$, $\bar{y}$, $\overline{xy}$, σ_x и σ_y (см. табл.):

№ п/п	Стаж работы, лет x	Выработка рабочего за смену, шт. y	$\bar{y}_x$	$x-\bar{x}$	$(x-\bar{x})^2$	$y-\bar{y}$	$(y-\bar{y})^2$	$\bar{y}_x-\bar{y}$	$(\bar{y}_x-\bar{y})^2$
1	2	8	8,14	–3	9	–4	16	–3,86	14,90
2	3	10	9,43	–2	4	–2	4	–2,57	6,60
3	4	11	10,72	–1	1	–1	1	–1,28	1,64
4	5	11	12,01	0	0	–1	1	0,01	0,00
5	6	13	13,30	1	1	1	1	1,30	1,69
6	7	15	14,59	2	4	3	9	2,59	6,71
7	8	16	15,88	3	9	4	16	3,88	15,05
Итого	35	84	84,07	–	28	–	48	–	46,59

$$\bar{x} = \frac{\Sigma x}{n} = \frac{35}{7} = 5; \quad \bar{y} = \frac{\Sigma y}{n} = \frac{84}{7} = 12; \quad \overline{xy} = \frac{\Sigma xy}{n} = \frac{456}{7} = 65,14;$$

$$\sigma_x = \sqrt{\frac{\sum(x-\bar{x})^2}{n}} = \sqrt{\frac{28}{7}} = 2; \qquad \sigma_y = \sqrt{\frac{\sum(y-\bar{y})^2}{n}} = \sqrt{\frac{48}{7}} = 2,62.$$

Линейный коэффициент корреляции составляет:

$$r = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sigma_x \cdot \sigma_y} = \frac{65,14 - 5\cdot 12}{2\cdot 2,62} = 0,981.$$

Теоретическое корреляционное отношение и линейный коэффициент корреляции указывают на очень тесную связь между сменной выработкой рабочих бригады и их производственным стажем. Так как η и r практически совпали, то предположение о прямолинейной форме корреляционной связи между этими признаками можно считать верным.

Квадрат линейного коэффициента корреляции равен 0,962. Это значение коэффициента детерминации показывает, что в рассматриваемом примере вариация выработки рабочих на 96,2% обусловлена вариацией стажа работы и на 3,8% – прочими факторами.

ЛИТЕРАТУРА

1. *Гусаров В.М.* Статистика. – М.: ЮНИТИ, 2002.
2. *Елисеева И.И., Юзбашев М.М.* Общая теория статистики. – М.: Финансы и статистика, 1995.
3. *Ефимова М.Р. и др.* Практикум по общей теории статистики. – М.: Финансы и статистика, 2006.
4. Практикум по статистике / Под ред. В.М. Симчеры. – М.: Финстатинформ, 1999.
5. Практикум по теории статистики / Под ред. Р.А. Шмойловой. – М.: Финансы и статистика, 2004.
6. Теория статистики / Под ред. Р.А. Шмойловой. – М.: Финансы и статистика, 2006.

ОГЛАВЛЕНИЕ

Виктор Иванович Марук

ТЕОРИЯ СТАТИСТИКИ

Конспекты лекций
и решение типовых задач

Редактор *И.В. Верченко*
Технический редактор *М.Р. Фазлыева*
Компьютерная верстка *Э.Ю. Вислевской*

Подписано в печать 17.12.07. Формат $60 \times 84^1/_{16}$
Офсетная печать. Объем 7,0 п.л.
Тираж 150. Заказ 243.

Издательство Кыргызско-Российского
Славянского университета
720000, Бишкек, ул. Киевская, 44.

Отпечатано в типографии КРСУ
720000, Бишкек, ул. Шопокова, 68.