ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
КЫРГЫЗСКО-РОССИЙСКИЙ СЛАВЯНСКИЙ УНИВЕРСИТЕТ
ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра экономической теории

К.А. Алиев

СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ

Учебное пособие

Бишкек 2016

УДК 311.31(075).8
А 50

Рецензент *В.И. Гусева*, д-р экон. наук, проф.

Рекомендовано к изданию
кафедрой экономической теории и НТС КРСУ

Алиев К.А.
А 50 СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ: учебное пособие. Бишкек: КРСУ, 2016.112 с.

В учебном пособии изложены не только показатели абсолютных, относительных, средних величин и их вариации, но и основные показатели анализа динамических рядов, выборочного метода, показатели анализа при корреляционной связи и индексах.

Данное учебное пособие предназначено бакалаврам и магистрам по направлению «Экономика», а также изучающим курс статистики, при написании курсовых и дипломных работ, когда необходимо использовать статистические показатели для анализа массовых социально-экономических явлений и процессов.

СОДЕРЖАНИЕ

1. СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ АБСОЛЮТНЫХ, ОТНОСИТЕЛЬНЫХ И СРЕДНИХ ВЕЛИЧИН

1.1. Абсолютные величины

Изучение социально-экономических явлений и процессов с количественной стороны предполагает необходимость использования совокупности числовых показателей, среди которых заметное место занимает абсолютные, относительные и средние величины.

Абсолютные величины выражают объем, размер, численность изучаемых явлений и процессов. Абсолютные величины в статистике являются именованными, т. е. они имеют определенные единицы измерения, с помощью которых получают свою качественную характеристику.

Единицы измерения в большинстве случаев непосредственно связаны с естественно-природными свойствами явления и предмета. Так, жидкости получают единицу измерения в литрах, твердые вещества в кг и т.д.

Наряду с этим широко используется сложные единицы измерения, охватывающие два свойства того или иного явления, процесса. В экономике широко используется измерение затраты труда в человеко-часах, человеко-днях, а объем грузооборота измеряется в т-км.

Для обобщенной характеристики (оценки) финансово-хозяйственной деятельности предприятий используются стоимостные единицы измерения. Стоимостные показатели находят как произведение количества произведенной продукции на их цены. Стоимостные показатели используются для выведения таких обобщающих показателей, как прибыль, рентабельность, доход и т. д.

1.2. Относительные величины

При изучении и анализе структуры, динамики и интенсивности изменения явлений и процессов применяются в основном относительные величины. Относительную величину получают в результате сопоставления двух абсолютных величин, при котором один из них выступает в качестве базы сравнения, а другой выступает в качестве сравниваемой величины.

Относительные величины выражаются в виде коэффициентов, процентов, промиллей, продецимиллей, в зависимости от того в каких единицах представлены базы сравнения. Если базу сравнения умножают на единицу, получают коэффициенты, если на 100 – проценты, если на 1000 – промилли, если на 10 000 – продецимилли.

Относительные величины имеют следующие виды:

Относительная величина динамики – характеризует изменения явления во времени и показывает, во сколько раз увеличился или уменьшился уровень показателя по сравнению с базисным периодом:

$$Ob_d = \frac{y_t}{y_0},$$

где Ob_d – относительная величина динамики;

y_t – текущий, сравниваемый уровень;

y_0 – базисный уровень.

Относительная величина планового задания – находят как отношение запланированного уровня к фактическому уровню:

$$Ob_{пл} = \frac{y_{пл}}{y_ф},$$

где $Ob_{пл}$ – относительная величина планового задания;

$y_{пл}$ – плановое задание на текущий период;

$y_ф$ – фактический уровень показателя.

Относительная величина выполнения плана – находят как отношение фактического уровня к плановому уровню:

$$Ob_{вып} = \frac{y_ф}{y_{пл}},$$

где $Ob_{вып}$ – относительная величина выполнения плана.

Относительная величина структуры – находят как отношение отдельных частей к общей совокупности. Наряду с соотношением части и целого (относительная величина структуры) находят соотношение между двумя частями одного целого. Относительную величину, характеризующее соотношение частей между собой называют **относительной величиной координации**.

$$Ob_{ст} = \frac{f_i}{\sum f_i},$$

где $Ob_{ст}$ – относительная величина структуры;

f_i – части (удельный вес) единиц совокупности;

$\sum f_i$ – численность общей совокупности.

Относительная величина сравнения показывает результаты сопоставления одноименных данных, относящихся к одному периоду времени, но к разным объектам, территориям, странам. Например, производство промышленной или сельскохозяйственной продукции по отдельным территориям, станам за определенный период времени.

Относительная величина интенсивности характеризует степень распространения или уровень развития изучаемого явления, процесса в определенной среде. Относительную величину интенсивности определяют путем сравнения двух разноименных величин, находящихся в определенной взаимосвязи между собой. Примером могут быть: плотность населения (чел./кв. м.), численность медицинского персонала на 10 000 чел. фондовооруженность или энерговооруженность труда и т.д.

1.3. Средние величины

При анализе социально-экономических явлений исследователи сталкиваются с таким фактом, когда индивидуальные значения изучаемых признаков варьируют в достаточно больших пределах.

На индивидуальные значения признака воздействует как постоянные, так и случайные факторы. Если элиминировать влияние случайных факторов, то мы получим наиболее общие, типичные показатели, которые выражаются через средние величины.

При нахождении средней величины из массы единиц влияние случайных величин взаимопогашаются и средняя, абстрагируясь от индивидуальных особенностей отдельных единиц, выражает наиболее общие, типичные и характерные свойства, присущие всем единицам совокупности. Поэтому средние величины представляют собой не просто количественный статистический метод измерения, а категорию объективной действительности.

В статистике выделяют две категории средних величин: степенные средние и структурные средние. К степенным средним относятся: средняя арифметическая, средняя гармоническая, средняя геометрическая, средняя квадратическая. К структурным средним относятся статистическая мода и медиана.

Выбор того или иного вида средней зависит от особенности распределения изучаемого признака, от ее экономической сущности и от особенности задач, которую необходимо решать.

Средняя арифметическая широко используются в экономических расчетах. Средняя арифметическая бывает двух видов: средняя арифметическая простая и взвешанная. Среднюю арифметическую простую находят как отношение суммы индивидуальных значений признака на число единиц данной совокупности:

$$\bar{X}_{пр} = \frac{\sum x_i}{n},$$

где $\bar{X}_{пр}$ – средняя арифметическая простая;

x_i – индивидуальное значение признака;

n – численность единиц совокупности.

Если имеется повторение отдельных значений признака, то их предварительно группируют с учетом повторяемости, а затем находят среднюю с учетом их весов (повторяемости) по следующей формуле:

$$\bar{X}_{вз.} = \frac{\sum x_i f_i}{\sum f_i},$$

где $\bar{X}_{вз}$ – средняя арифметическая взвешанная;

f_i – весы (частота) признака.

Данную формулу можно видоизменить, если весы (частоты выразить в виде относительных величин, т. е. в виде коэффициента или же процента.

Тогда $\sum f = \sum d = 100~\%$; $\sum f = \sum w = 1.$

Таблица 1.1. – Пример для расчета средней арифметической взвешанной

Выработка деталей за смену, шт. (x_i)	Число рабочих (чел.) (f_i)	Удельный вес (раб.) в % (d_i)	Удельный вес (раб.) в коэф. (w_i)
4	2	20	0,2
7	2	20	0,2
8	1	10	0,1
12	5	50	0,5
Итого	10	100	1,0

1) средняя выработка, применяя абсолютное значение весов:

$$\bar{X}_{вз} = \frac{\sum x_i f_i}{\sum f_i} = \frac{4\cdot 2 + 7\cdot 2 + 8\cdot 1 + 12\cdot 5}{10} = \frac{90}{10} = 9;$$

2) средняя выработка, применяя в качестве весов у.в. рабочих в %:

$$\bar{X}_{d} = \frac{\sum x_i d_i}{\sum d_i} = \frac{4\cdot 20 + 7\cdot 20 + 8\cdot 10 + 12\cdot 50}{100} = \frac{900}{100} = 9;$$

3) средняя выработка, применяя в качестве весов у.в. рабочих в коэф.:

$$\bar{X}_{w} = \frac{\sum x_i w_i}{\sum w_i} = \frac{4\cdot 0{,}2 + 7\cdot 0{,}2 + 8\cdot 0{,}1 + 12\cdot 0{,}5}{1} = \frac{9}{1} = 9.$$

Средняя арифметическая обладает следующими свойствами:

1. Средняя арифметическая постоянной величины равна этой постоянной величине.
2. Сумма индивидуальных отклонений признака от ее средней всегда равна нулю.
3. Если каждое значение признака уменьшить или увеличить на какое- либо постоянное число, то средняя арифметическая соответственно уменьшится или увеличится на ту же величину.
4. Если каждое значение признака увеличить или уменьшить в кратный раз, то средняя арифметическая увеличится или уменьшится в кратный раз.
5. Сумма квадратов отклонений индивидуальных значений признака от средней арифметической всегда составляет минимальную величину.

Средняя гармоническая:

Если веса (частота) у каждого значения признака отсутствует или же они равны между собой, то применяют среднюю гармоническую простую по следующей формуле:

$$\bar{X}_{гар.пр.} = \frac{n}{\sum \frac{1}{x_i}}.$$

В тех случаях, когда у признаков разные частоты (веса), то применяют среднюю гармоническую взвешанную по следующей формуле:

$$\bar{X}_{гар.вз.} = \frac{\sum x_i f_i}{\sum \frac{x_i f_i}{x_i}}.$$

Таблица 1.2. – Пример для расчета средней гармонической взвешанной

Группа рабочих	Средняя заработная плата по группам (x_i)	Фонд заработной платы $(x_i f_i)$
1	940	9400
2	1020	40800
3	1150	34500
4	1400	28000

Итого 112700

$$\bar{X}_{гар.вз.} = \frac{\sum x_i f_i}{\sum \frac{x_i f_i}{x_i}} = \frac{944 + 40800 + 34500 + 28000}{\frac{9400}{940} + \frac{40800}{1020} + \frac{34500}{1150} + \frac{28000}{1400}} = \frac{112700}{100} = 1127.$$

Средняя степенная имеет два вида:

Средняя степенная простая и средняя степенная взвешанная и их находят по следующей формуле:

1) $\bar{X}_{ст.пр} = \sqrt[z]{\frac{x_i^z}{n}}$ – простая средняя степенная;

2) $\bar{X}_{ст.вз.} = \sqrt[z]{\frac{\sum x_i^z f_i}{\sum f_i}}$ – взвешанная средняя степенная.

При этом, если z = 1, то получаем среднюю арифметическую; если z = -1, то получим среднюю гармоническую. При z = 2, то получим среднюю квадратическую.

Формула средней квадратической:

1) $\bar{X}_{кв.пр.} = \sqrt{\frac{\sum x_i^2}{n}}$ – простая средняя квадратическая;

2) $\bar{X}_{кв.вз.} = \sqrt{\frac{\sum x_i^2 f_i}{\sum f_i}}$ – взвешанная средняя квадратическая.

Среднюю квадратическую используют в тех случаях, когда варианты представляют собой отклонение фактических величин от их средней арифметической или от принятой нормы.

Пример. По результатам измерения отклонений фактической длины изделий от заданной нормы необходимо определить среднее отклонение в целом:

Таблица 1.3.

Отклонение фактич. длины от нормы, мм (x_i)	Число изделий (f_i)	(x_i^2)	$(x_i^2 f_i)$
–1,8	1	3,24	3,24
–0,8	3	0,64	1,92
+0,2	4	0,04	0,16
+1,2	1	1,44	1,44
+2,2	1	4,84	4,84
Итого	10	–	11,60

$$\bar{X}_{кв} = \sqrt{\frac{\sum x_i^2 f_i}{\sum f_i}} = \sqrt{\frac{11,6}{10}} = 1,08 \text{ мм}.$$

Среднее отклонение фактической длины изделий от заданной нормы составляет 1,08 мм. В данном случае нельзя применить среднюю арифметическую, поскольку сумма отклонений индивидуальных значений от заданной нормы (средней) равна нулю.

Наряду с приведенными выше средними показателями, применяются средние хронологические и средние геометрические, которые являются обобщающими характеристиками показателей анализа динамических рядов.

1. Средняя хронологическая определяется по следующей формуле:

$$\bar{X}_{хр.} = \frac{1/2x_1 + x_2 + x_3 + \ldots + x_{n-1} + 1/2x_n}{n-1};$$

2. Формула средней геометрической:

$$\bar{X}_{геом.} = \sqrt[m]{x_1 \cdot x_2 \cdot x_3 \ldots x_n}.$$

Статистическая мода и медиана

Мода и медиана, как и средние величины, являются обобщающими показателями изучаемой совокупности. При этом моду в статистике находят по ее наибольшей частоте, а медиану по ее месту расположения в упорядоченном (ранжированном) ряде распределения.

Моду и медиану находят для дискретного и интервального ряда распределения в отдельности. Так, моду в дискретном ряде распределения только по ее наибольшей частоте. При определении моды в интервальном ряде распределения использует следующую формулу:

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

где x_0 – нижняя граница модального интервала;

i – величина модального интервала;

f_2 – модальная частота;

f_1 – частота, предшествующая модальной частоте;

f_3 – частота, последующая за модальной частотой.

Медиану в дискретном ряде распределения находят по месту расположения, чтобы она данный упорядоченный ряд разделяла на равные части. Местонахождение медианы в дискретном ряде распределения находят, используя соотношение: $\frac{n+1}{2}$, где n – численность вариантов признака в дискретном ряде распределения.

Медиану в интервальном ряде распределения определяют по следующей формуле:

$$Me = x_0 + i\frac{1/2\sum f - S_{n-1}}{f_{me}},$$

где x_0 – нижняя граница медианного интервала;

i – величина медианного интервала;

$1/2\sum f$ – половина суммы частот;

S_{n-1} – сумма накопленных частот, предшествующее медианной частоте;

f_{me} – медианная частота.

Рассчитаем значение моды и медианы, используя данные следующего условного примера (таблица 1.4.).

Таблица 1.4. – Группировка кредитных организаций по величине активов

Активы, млн сомов	Число кредитных организаций	Накопленная частота
105–115	4	4
115–125	9	13
125–135	21	34
135–145	49	83
145–155	28	111
155–165	18	129
165–175	11	140
Итого	140	-

Интервал с границами 135–145 в данном распределении будет модальным, так как он имеет наибольшую частоту. Используя приведенную выше формулу, рассчитаем моду:

$$Mo = 135 + 10 \cdot \frac{49-21}{(49-21)+(49-28)} = 140,71\ \text{млн сом.}$$

Для определения медианного интервала необходимо определить накопленную частоту каждого последующего интервала до тех пор. Пока она не превысит ½ суммы накопленных частот. В данном случае медианным является интервал с границами 135–145. Используя эти данные, рассчитаем значение медианы:

$$Me = 135 + 10 \cdot \frac{(140/2) - 34}{49} = 142,35\ \text{млн сом.}$$

Контрольные вопросы, определение понятий и показателей

1. Назначение и особенности использования абсолютных, относительных величин.
2. Виды относительных величин и их показатели.
3. Средние величины, виды средних величин.
4. Показатели, принятые при определении различных видов средней.
5. Определение статистической моды и медианы.

2. СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ ВАРИАЦИОННЫХ РЯДОВ

2.1. Вариационные ряды

Ряды распределения – это упорядоченное распределение единиц совокупности на группы по изучаемому варьирующему признаку. В зависимости от вида признака, положенного в основу группировки, – качественной или количественной – различают соответственно два типа рядов распределения – атрибутивные и вариационные. Ряды распределения, построенные по качественным признакам, называют *атрибутивными.* Примером атрибутивных рядов может служить распределение населения по полу, возрасту, национальности, уровню образования и т.д. Ряды распределения, построенные по количественному признаку, называют *вариационными*. Величины того или иного количественного признака у отдельных единиц совокупности различаются между собой. Такое различие в величине изучаемого признака называют вариацией. Изучение характера и степени вариации признаков единиц совокупности – особая задача статистического исследования.

При наличии большого количества вариантов значений изучаемого признака трудно сделать какие-либо суждения о характере распределения вариантов значений признака в совокупности. Начальным этапом в упорядочении исходных (первичных) данных вариации признаков единиц совокупности является его ранжирование, т. е. расположение всех вариантов ряда в возрастающем или убывающем порядке.

Ранжированный ряд данных позволяет определить наименьшее и наибольшее значение признака в данной совокупности, а также величину между крайними значениями признака и выделить наиболее часто повторяющиеся значения в изучаемой совокупности. Все эти данные в последующем используются при формировании различных видов групп и подгрупп.

Число повторений отдельных вариантов значений признаков называют частотой повторений, если их обозначить f_i, то сумма частот, равная объему изучаемой совокупности, – $\sum_{i=1}^{k} f_i$, или n ($\sum_{i=1}^{k} f_i = n$),
где k – число вариантов значений признака.

По характеру вариации различают *дискретные и непрерывные признаки. Дискретные* признаки выражаются в виде некоторой конечной величины, т. е. в виде прерывных конечных чисел, например, число рабочих на предприятии, тарифный разряд рабочих и т.д. Непрерывные признаки могут отличаться один от другого на малую величину и чтобы округлить (обобщить) их, величину заключают в определенные интервалы, например, заработная плата рабочих, среднедушевой денежный доход семьи и т.д.

Способы построения вариационного ряда для этих видов признаков различны. Для построения дискретного ряда с небольшим числом вариантов достаточно перечислить все имеющиеся варианты значений признака x_i и подсчитать частоту повторения каждого варианта f_i и результаты оформить в виде следующей таблицы:

Таблица 2.1.

Тарифный разряд рабочего X_i	Число рабочих	Частость W_i	Накопленная Частота S_i
2	2	0,08	2
3	5	0,2	7
4	7	0,28	14
5	8	0,32	22
6	3	0,12	25
Итого	25	1,00	

Из таблицы видно, что ряд первичных данных, характеризующих квалификацию 25 рабочих, сгруппированы в пять групп, где число рабочих, имеющих определенный разряд, дополнен их относительной величиной, т. е. долей рабочих, имеющих тот или иной разряд.

Частоты, представленные в относительном выражении, называют частостями и обозначают w_i: $W_i = \frac{f_i}{\sum_{i=1}^{k} f_i}$. частости могут быть выражены в долях единицы или в процентах. Замена частот частостями позволяет сопоставлять вариационные ряды с различным числом наблюдений.

В тех случаях, когда число вариантов дискретного признака достаточно велико, а также при анализе вариации непрерывного признака, когда значение признака у отдельных единиц могут вообще не повторяться, строятся интервальные ряды распределения. Интервал указывает определенные пределы значений варьирующего признака с обозначением нижней и верхней границы интервала.

При построении интервальных рядов распределения следует установить число групп (интервалов), на которые следует разбить все единицы изучаемой совокупности. При группировке однородных совокупностей применяются равные интервалы, величина которых зависит от вариации признака и количества единиц совокупности.

Величина интервала определяется по формуле: $d = \frac{x_{\max} - x_{\min}}{n}$,
где d – величина интервала;
$x_{\max}$, $x_{\min}$ – максимальное и минимальное значение признака;
n – количество выделяемых групп. Количество групп определяется по формуле Стэрджесса: n = 1+3,322lg s; s – общее число изучаемых единиц.

Если вариационный ряд с неравными интервалами, то для обеспечения сопоставимости рядов распределения необходимо рассчитать абсолютную или относительную плотности распределения, для определения которых находят отношение частот или частостей к величине интервала:

абсолютная плотность $m_i^{(a)} = \frac{f_i}{d_i}$;

относительная плотность $m_{d_i}^{(o)} = \frac{w_i}{d_i}$.

2.2. Показатели центра распределения

Для характеристики среднего значения признака в вариационном ряду применяется средняя арифметическая, мода, медиана. (Подробные характеристики об этих величинах изложены в предыдущих разделах).

Средняя арифметическая для дискретного ряда распределения рассчитывается по формуле: $\bar{x} = \sum_{i=1}^{k} x_i f_i \Big/ \sum_{i=1}^{k} f_i$,

где x_i – варианты значений признака;

f_i – частота повторения данного варианта.

В интервальном вариационном ряду средняя арифметическая определяется по формуле: $\bar{x} = \sum_{i=1}^{k} x_i^{'} f_i \Big/ \sum_{i=1}^{k} f_i$,

где $x_i^{'}$ – середина соответствующего интервала.

Пример расчета средней арифметической для интервального вариационного ряда приведен в таблице 2.3.

В отличие от средней арифметической, рассчитываемой на основе использования всех вариантов значений признака, мода и медиана характеризуют величину варианта, занимающего определенное положение в ранжированном вариационном ряду. Медиана соответствует варианту, расположенному в середине ранжированного ряда и местоположение медианы при дискретном ряде распределения определяется формулой:

$$N_{Me} = \frac{n+1}{2},$$

где n – число единиц в совокупности.

При интервальном ряде распределения **медиана** определяется следующей формулой: $Me = x_{Me} + d\,\dfrac{\dfrac{n+1}{2} - S_{(-1)}}{f_{Me}}$,

где x_{Me} – нижняя граница медианного интервала;

d – величина интервала;

$S_{(-1)}$ – накопленная частота интервала, предшествующего медианному;

f_{Me} – частота медианного интервала.

Мода (Мо) – наиболее часто встречающееся значение признака в совокупности. Моду для дискретного ряда находят по наибольшей частоте вариантов изучаемой совокупности, а моду для интервального ряда распределения находят, применяя следующую формулу:

$$Mo = x_{Mo} + d\,\frac{f_{Mo} - f_{(-1)}}{\left[f_{Mo} - f_{(+1)}\right] + \left[f_{Mo} - f_{(-1)}\right]},$$

где x_{Me} – начало модального интервала;

f_{Me} – частота, соответствующая модальному интервалу;

$f_{(-1)}$ – предмодальная частота;

$f_{(+1)}$ – послемодальная частота

Таблица 2.2.

Средняя дневная выработка, тыс. сом. (x_i)	Число рабочих (f_i)	Накопленные частоты (S_i)
3,7–4,6	3	3
4,6–5,5	4	7
5,5–6,4	6	13
6,4–7,3	3	16
7,3–8,2	4	20
Итого	20	

Используя исходные данные таблицы 2.2., а также приведенные выше формулы, рассчитаем значения медианы и моды при интервальном ряде распределения:

$$Me = x_{Me} + d\,\frac{\dfrac{n+1}{2} - S_{Me}}{f_{Me}} = 5,5 + 0,9\,\frac{\dfrac{20+1}{2} - 7}{6} = 6,025.;$$

$$Mo = x_{Mo} + d\,\frac{f_{Mo} - f_{-1}}{(f_{Mo} - f_{-1}) + (f_{Mo} - f_{+1})} = 5,5 + 0,9\,\frac{6-4}{(6-4)+(6-3)} = 5,86.$$

Основной характеристикой центра распределения является средняя арифметическая, опирающаяся на все единицы изучаемой совокупности. Однако в ряде случаев средняя величина должна быть дополнена модальным значением или же медианой. Так, в статистическом контроле качества продукции удобнее пользоваться медианой, а не средней арифметической, так как определение медианы для ранжированного ряда данных не требует специального расчета. В рядах с открытыми интервалами также целесообразно пользоваться в качестве характеристики центра модой и медианой. Мода широко применяется при изучении спроса населения на товары широкого потребления, пользующиеся наибольшим спросом. Наконец, следует указать, что в симметричных рядах распределения между средней арифметической, модой и медианой соблюдается равенство, т. е.: $\bar{X} = Mo = Me$.

2.3. Показатели вариации признака

Средняя арифметическая, которая выступает в качестве центра в рядах распределения отдельных единиц изучаемой совокупности, одновременно является наиболее типичным, обобщающим, устойчивым представителем всей совокупности, в которой взаимно исключены отдельные случайные отклонения. Поэтому для характеристики степени вариации, колеблемости отдельных значений единиц совокупности сравнивают со средней величиной и находят абсолютные и относительные отклонения отдельных вариантов, единиц изучаемой совокупности.

Для измерения вариации признака применяются различные абсолютные и относительные показатели. К абсолютным показателям относятся: размах вариации, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение и квартильное отклонение.

Размах вариации представляет собой разность между максимальными и минимальными значениями признака в изучаемой совокупности:

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

Достоинством этого показателя является простота расчета. Поскольку размах вариации зависит от величины крайних значений признака, то его применяют при характеристике степени колеблемости однородных совокупностей.

Наиболее объективную характеристику вариации признаков дают показатели, основанные на учете колеблемости всех

значений признака. Поскольку средняя арифметическая является обобщающей, типичной характеристикой признаков всей совокупности, то большинство показателей вариации основано на определение отклонений отдельных значений признака от этой средней величины. К таким показателям относятся среднее линейное отклонение, дисперсия и среднее квадратическое отклонение, которое является средней арифметической из отклонений индивидуальных значений признака от средней арифметической. Среднее линейное отклонение рассчитывается из отклонений в первой степени, дисперсия и среднее квадратическое отклонение – из отклонений во второй степени. Учитывая, что алгебраическая сумма отклонений индивидуальных значений признака от средней арифметической всегда равна нулю, то для расчета среднего линейного отклонения используется арифметическая сумма (абсолютных) отклонений без учета знака.

Среднее линейное отклонение ($\bar{l}$) определяется по следующим формулам:

для несгруппированных данных:

$$\bar{l} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n};$$

для сгруппированных данных:

$$\bar{l} = \frac{\sum_{i=1}^{k} |x_i - \bar{x}| f_i}{\sum_{i=1}^{k} f_i}.$$ [1]

1. При расчете среднего линейного (квадратического) отклонения для интервального вариационного ряда определяют отклонения центральных значений интервала от средней арифметической, т. е. величины $x_i' - \bar{x}$.

Дисперсия (σ^2) – средняя из квадратов отклонений вариантов значений признака от их средней величины. Дисперсия рассчитывается по следующим формулам:

для несгуппированных данных:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n},$$

для сгуппированных данных (вариационного ряда):

$$\sigma^2 = \frac{\sum_{i=1}^{k}(x_i - \overline{x})^2 f_i}{\sum_{i=1}^{k} f_i}.$$

Дисперсия обладает следующими свойствами:

– дисперсия постоянной величины равна нулю;

– если все варианты значений признака уменьшить на одно и тоже постоянное число (к), то дисперсия не уменьшится;

– если все варианты значений признака уменьшить в (к) кратный раз, то дисперсия уменьшится в к^2 раза.

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

для несгруппированных данных:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}};$$

для вариационного ряда:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \overline{x})^2 f_i}{\sum_{i=1}^{k} f_i}}.$$

Размах вариации, среднее линейное и среднее квадратическое отклонение являются величинами именованными. Они имеют те же единицы измерения, что и индивидуальные значения признака.

Исходные данные для расчета показателей вариации приводятся в таблице 2.3.

Таблица 2.3.

Произведено продукции в 1 смену, тыс. сом.	Количество рабочих, чел.	x_i'	$x_i' f_i$	$x_i' - \overline{x}$	$\lvert x_i' - \overline{x}\rvert f_i$	$(x_i' - \overline{x})^2 f_i$
3,2–4,0	2	3,6	7,2	–1,68	3,36	5,64
4,0–4.8	3	4,4	13,2	–0,88	2,64	2,32
4,8–5,6	8	5,2	41,6	–0,08	0,64	0,05
5,6–6,4	5	6,0	30,0	0,72	3,6	2,6
6,4–7,2	2	6,8	13,6	1,52	3,04	4,62
Итого	20		105,6		13,25	15,23

$$\overline{x} = \frac{105,6}{20} = 5,28 \text{ } \textit{тыс. сом.};$$

$$\overline{l} = \frac{13,25}{20} = 0,66 \text{ } \textit{тыс. сом.};$$

$$\sigma^2 = \frac{15,23}{20} = 0,76 \text{ } \textit{тыс. сом.};$$

$$\sigma = \sqrt{0,76} = 0,87 \text{ } \textit{тыс. сом.}$$

Среднее квадратическое отклонение показывает, как расположена основная масса единиц совокупности относительно средней арифметической. В соответствии с теоремой П.Л. Чебышева, можно допускать, что, независимо от формы распределения, 75 % значений признака попадают в интервал $\overline{x} \pm 2\sigma$, а около 89 % всех значений признака попадают в интервал $\overline{x} \pm 3\sigma$.

Если в качестве показателя центра распределения используется медиана, то для характеристики вариации признаков в совокупности можно применить квартильное отклонение (Q). Этот показатель также можно применить вместо размаха вариации, чтобы устранить недостатки, связанные с использованием крайних значений.

$$Q = \frac{Q_3 - Q_1}{2},$$

где Q_1 и Q_3 соответственно первая и третья квартили распределения.

Квартили – это значения признака в ранжированном ряду распределения, выбранные таким образом, что 25 % единиц совокупности будут меньше по величине Q_1; 25 % единиц будут заключены между Q_1 и Q_2; 25 % – между Q_2 и Q_3 и остальные 25 % превосходят Q_3. Квартили определяются по формулам, аналогичным приведенной выше формуле для расчета медианы:

$$Q_1 = x_{Q_1} + d\frac{\frac{n+1}{4} - S_{(-1)}}{f_{Q_1}},$$

где x_{Q_1} – нижняя граница интервала, в котором находится первая квартиль;
$S_{(-1)}$ – сумма накопленных частот интервалов, предшествующих интервалу, в котором находится первая квартиль;
f_{Q_1} – частота интервала, в котором находится первая квартиль:

$$Q_1 = x_{Q_1} + d\frac{\frac{n+1}{4} - S_{(-1)}}{f_{Q_1}},$$

где x_{Q_3} – нижняя граница интервала, в котором находится третья квартиль; $S_{(-1)}$ – сумма накопленных частот интервалов, предшествующих интервалу, в котором находится третья квартиль; f_{Q_3} – частота интервала, в котором находится третья квартиль. Используя приведенные выше формулы и исходные данные таблицы 2.3. определим соответствующие показатели:

$$Q_1 = 4,0 + 0,8\frac{21/4-3}{4} = 4,45 \textit{ тыс. сом.};$$

$$Q_2 = 4,8 + 0,8\frac{21/2-5}{8} = 5,35 \textit{ тыс. сом.};$$

$$Q_3 = 5,6 + 0,8\frac{21*3/4-13}{5} = 6,04 \textit{ тыс. сом.};$$

$$Q = \frac{Q_3 - Q_1}{2} = \frac{6,04-4,45}{2} = 0,795 \textit{ тыс. сом.};$$

$$M_e = 4,8 + 0,8\frac{21/2-5}{8} = 5,35 \textit{ тыс. сом.}$$

В симметричных или умеренно асимметричных распределениях $Q = 2/3\sigma$., поскольку на квартильное отклонение не влияют

отклонения всех значений признака, то его использование следует ограничить случаями, когда определение среднего квадратического отклонения затруднительно, например, для рядов распределения с открытыми интервалами.

При сравнении колеблемости различных признаков в одной и той же совокупности или же при сравнении колеблемости одного и того же признака в нескольких совокупностях с различной величиной средней арифметической пользуются относительными показателями вариации. Эти показатели определяют как отношение абсолютных показателей вариации к средней арифметической (или к медиане). Используя в качестве абсолютного показателя вариации размах вариации, среднее линейное отклонение, среднее квадратическое отклонение или квартильное отклонение получают относительные показатели вариации (они в основном выражаются в процентах).

Коэффициент осцилляции: $K_R = \frac{R}{\bar{x}} \cdot 100\ \% = \frac{7,2-3,2}{5,28} \cdot 100\ \% = 75,75\ \%;$

Коэффициент среднего линейного отклонения:

$$K_{\bar{l}} = \frac{\bar{l}}{\bar{x}} \cdot 100\ \% = \frac{0,66}{5,28} \cdot 100\ \% = 12,5\ \%;$$

Коэффициент вариации: $V = \frac{\sigma}{\bar{x}} \cdot 100\ \% = \frac{0,87}{5,28} \cdot 100\ \% = 16,5\ \%;$

Коэффициент квартильной вариации:

$$K_Q = \frac{Q}{M_e} \cdot 100\ \% = \frac{0,795}{5,35} \cdot 100\ \% = 14,85\ \%.$$

Коэффициенты вариации используются не только для сравнительной оценки вариации признака в изучаемой совокупности, но и для сравнительной характеристики их однородности. Совокупность считается однородной, если коэффициент вариации не превышает 33 % для распределений близких к нормальной. Коэффициенты вариации, рассчитанные на основе данных таблицы 2.3., показывают, что уровень и степень колеблемости признаков в данной совокупности сравнительно небольшие, и они являются однородными.

Если изучаемая статистическая совокупность распределена по какому-либо признаку на отдельные части, группы, тогда возникает необходимость рассчитать с одной стороны соответствующие обобщающие показатели в виде средней величины, с другой стороны необходимо оценить степень вариации, отдельных вариантов, единиц, из которой состоит изучаемая совокупность. Для решения этих задач необходимо рассчитать как средние показатели, так и показатели вариации в каждой отдельно выделенной группе, между выделенными типичными группами, а также по всей изучаемой совокупности.

Для оценки степени вариации внутри каждой выделенной группы используется частная или же внутригрупповая дисперсия:

$$\sigma_j^2 = \frac{\sum_{i=1}^{n_{ji}} (x_j - \overline{x}_j)^2}{n_j};$$

Обобщенная характеристика степени вариации признаков в выделенных частных группах определяется в виде средней из частных (внутригрупповых) дисперсий:

$$\overline{\sigma}^2 = \frac{\sum_{j=1}^{k} \sigma_j^2 n_j}{\sum_{j=1}^{k} n_j};$$

Вариацию, обусловленную влиянием фактора, положенного в основу группировки, характеризует межгрупповая дисперсия (δ^2), которая является мерой колеблемости частных средних по группам $(\overline{x}_j)$ по сравнению с общей средней ($\overline{x}_o$) и определяется по формуле:

$$\delta^2 = \frac{\sum_{j=1}^{k} (\overline{x}_j - \overline{x}_o)^2 n_j}{\sum_{j=1}^{k} n_j};$$

Между общей дисперсией (σ_o^2), средней из частных (внутригрупповых) дисперсий ($\overline{\sigma}^2$) и межгрупповой дисперсией (δ^2) имеется определенное соотношение, определяемое правилом сложения дисперсий:

$$\sigma_o^2 = \overline{\sigma}^2 + \delta^2.$$

Обозначения:

σ_j^2 – внутригрупповая, частная дисперсия;

$\overline{\sigma}^2$ – средняя из внутригрупповых, частных дисперсий;

σ_o^2 – общая дисперсия;

k – число групп;

n_j – число единиц в j–й группе;

$\overline{x}_j$ – частная средняя по j–ой группе;

$\overline{x}_o$ – общая средняя по совокупности единиц.

Пример: при изучении влияния стажа работы рабочих на производительность труда в обследованной группе получены следующие данные, которые приведены в таблице 2.4.

Таблица 2.4.

Группа рабочих по стажу работы, лет	Число рабочих, чел.	Средняя часовая выработка рабочих в группах (в условных единицах)	Дисперсия по выработке в группах
До 5	30	20,0	3,0
5 и более	40	23,0	2,0

Средняя часовая выработка всех рабочих:

$$\overline{x}_o = \frac{20,0*30 + 23,0*40}{70} = 21,7;$$

Вариация признака за счет изменения стажа работы определяется показателем межгрупповой дисперсии:

$$\delta^2 = \frac{(20,0-21,7)^2*30 + (23,0-21,7)^2*40}{70} = 2,43;$$

Вариация всех прочих внутригрупповых факторов характеризуется средней из внутригрупповых (частных) дисперсий:

$$\overline{\sigma}_j^2 = \frac{3,0830 + 2,0*40}{70} = 2,43;$$

Вариация выработки под воздействием всех факторов характеризуется величиной общей дисперсии:

$$\sigma_o^2 = 2,43 + 2,20 = 4,63.$$

Это означает, что на 47,6 % $\left(\frac{2,20}{4,63}*100\ \%\right)$ дисперсия по выработке обусловлена различиями в стаже рабочих, а на 52,4 % $\left(\frac{2,43}{4,63}*100\ \%\right)$ – влиянием прочих внутригрупповых факторов.

Наряду с изучением и оценкой степени вариации количественных признаков, в статистике выделяют способы оценки вариации качественных признаков, которые непосредственно не поддаются количественным измерениям. Если имеются два взаимоисключающих варианта значений признака, то это свидетельствует о наличии альтернативной изменчивости качественных признаков. Так, при изучении качества изготовленной продукции можно разделить на две группы: стандартную и не стандартную, т. е. бракованную. В целях обобщенной характеристики распределения качественных признаков производят следующее обозначение: наличие качественного признака – 1, а отсутствие – 0, общее число единиц совокупности – n, число единиц, обладающих данным качественным признаком – m, а число единиц, не обладающих данным признаком, будет равно – n-m. Используя эти произведенные обозначения, построим ряд распределений по качественному признаку:

Таблица 2.5.

Значение признака	Частота повторений	Частость повторений
1	m	$p = m/\ n$
0	n–m	$q = n - m/n$

Средняя арифметическая данного альтернативного ряда распределения равна:

$$\bar{x} = \frac{1*m + 0*n - m}{n} = \frac{m}{n} = p,$$

т. е. средняя арифметическая альтернативного признака распределения равна относительной частоте (частости): $\bar{x} = p$.

Если доля единиц, обладающих данным признаком равна p; соответственно доля единиц, не обладающих данным признаком, равна q, и p+q = 1. Тогда дисперсия альтернативного признака определяется по формуле:

$$\sigma^2 = \frac{(1-p)^2 p + (0-p)^2 q}{p+q} = \frac{q^2 p + p^2 q}{p+q} = pq.$$

Среднее квадратическое отклонение альтернативного признака равна:

$$\sigma = \sqrt{pq}.$$

Пример: при проверке качества 100 единиц изготовленной продукции 20 оказались бракованными, т. е. доля бракованной продукции составляет $2\ \%\left(\frac{2}{100}*100\right)$, а доля стандартной продукции $98\ \%\left(\frac{98}{100}*100\right)$. Тогда дисперсия данного альтернативного признака равна: $\sigma_p^2 = pq = 0,02*0,98 = 0,0196$, а среднее квадратическое отклонение: $\sigma_p = 0,14$.

2.4. Моменты распределения

Моментом распределения называется средняя арифметическая тех или иных степеней отклонений индивидуальных значений признака от определенной исходной величины и рассчитывается по формуле:

$$\frac{\sum_{i=1}^{k}(x_i - A)^a f_i}{\sum_{i=1}^{k} f_i},$$

где A – величина, от которой определяются отклонения, a – степень отклонения (порядок момента).

В зависимости от того, что принимается за величину A, различают три вида моментов:

начальные моменты M_a получает при A = 0:

$$M_a = \frac{\sum_{i=1}^{k} x_i^a f_i}{\sum_{i=1}^{k} f_i};$$

центральные моменты μ_a получают при $A = \bar{x}$:

$$\mu_a = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^a f_i}{\sum_{i=1}^{k} f_i};$$

условные моменты m_a получают при А, не равной средней арифметической и отличной от нуля:

$$m_a = \frac{\sum_{i=1}^{k} (x_i - A)^a f_i}{\sum_{i=1}^{k} f_i}.$$

В статистической практике пользуются моментами первого, второго, третьего, четвертого порядков и они представлены в таблице 2.6.

Из приведенных формул моментов видно, что начальный момент первого порядка представляет собой среднюю арифметическую и используется как показатель центра распределения. Центральный момент первого порядка (в соответствии с нулевым свойством средней арифметической) всегда равен нулю. Центральный момент второго порядка представляет собой дисперсию и служит основным показателем вариации признака. Центральный момент третьего порядка равен нулю в симметричном распределении и используется для определения показателя асимметрии. Центральный момент четвертого порядка применяется при вычислении показателя эксцесса. Начальные моменты второго, третьего и четвертого порядков так же, как и условные моменты, самостоятельного значения не имеют, а используются для упрощения вычислений центральных моментов. Например, используя начальные моменты первого и второго порядка, можно получить дисперсию по такой формуле:

$$\sigma^2 = \mu_2 = M_2 - M_1^2.$$

Таблица 2.6. – Моменты распределения порядка

Моменты распределения порядка	Начальные	Центральные	Условные
Первого	$M_1 = \frac{\sum x_i f_i}{\sum f_i}$	$\mu_i = \frac{\sum (x_i - \bar{x}) f_i}{\sum f_i}$	$m_1 = \frac{\sum x_i - A) f_i}{\sum f_i}$
Второго	$M_1 = \frac{\sum x_i^2 f_i}{\sum f_i}$	$\mu_2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}$	$\cancel{m}_2 = \frac{\sum (x_i - A)^2 f_i}{\sum f_i}$
Третьего	$M_3 = \frac{\sum x_i^3 f_i}{\sum f_i}$	$\mu_3 = \frac{\sum (x_i - \bar{x})^3 f_i}{\sum f_i}$	$m_3 = \frac{\sum (x_i - A)^3 f_i}{\sum f_i}$
Четвертого	$M_4 = \frac{\sum x_i^4 f_i}{\sum f_i}$	$\mu_4 = \frac{\sum (x_i - \bar{x})^4 f_i}{\sum f_i}$	$m_4 = \frac{\sum (x_i - A)^4 f_i}{\sum f_i}$

2.5. Формы распределения

Число наблюдений, по которому сроится эмпирическое распределение, обычно невелико и представляет выборку из изучаемой генеральной совокупности. Поэтому эмпирические данные в определенной степени связаны со случайными ошибками наблюдения, величина которых неизвестна. Влияние этих случайностей искажает основную закономерность изменения величины признака. С увеличением числа наблюдений и уменьшением величины интервала влияние случайных факторов в определенной степени устраняется, вследствие чего возникает плавная кривая линия, которая называется кривой распределения.

Кривая распределения характеризует теоретическое распределение, т. е. распределение, которое получилось бы при полном погашении всех случайных причин, искажающих основную закономерность. Изучение закономерности или формы распределения предполагает решение следующих последовательных задач: выяснение общего характера распределения; выравнивание эмпирического распределения, которое состоит в том, что на основании эмпирического распределения строится кривая $y = f(x)$

с заданной формой; проверку соответствия найденного теоретического распределения эмпирическому.

Следует отметить, что однородная совокупность характеризуется одновершинными распределениями. Наличие множества вершин свидетельствует о неоднородности изучаемой совокупности. Появление двух и более вершин говорит о необходимости перегруппировки данных для выделения более однородных групп. Выяснения общего характера распределения предполагает оценку степени его однородности, а также определение показателей асимметрии и эксцесса.

Симметричным является распределение, в котором частоты любых двух вариантов, равностоящих в обе стороны от центра распределения, равны между собой. При симметричном распределении имеет место равенство средней арифметической, моды и медианы. Исходя из этих соотношений, для характеристики асимметрии сравнивают показатели центра распределения: чем больше разница между средними $(\bar{x} - M_o)$ тем больше асимметрия ряда. Для сравнительного анализа степени асимметрии нескольких распределений рассчитывают относительный показатель асимметрии (A_S): $A_s = \dfrac{\bar{x} - M_o}{\sigma}$.

Показатель асимметрии может быть положительным и отрицательным. Положительное значение показателя асимметрии указывает на наличие правосторонней асимметрии (правая ветвь относительно максимальной ординаты вытянута больше, чем левая, (рисунок 1). При правосторонней асимметрии между показателями центра распределения существует соотношение: $M_o < M_e < \bar{x}$. Отрицательный знак показателя асимметрии свидетельствует о наличии левосторонней асимметрии (рисунок 2). Между показателями центра распределения в этом случае имеется такое соотношение: $M_o > M_e > \bar{x}$.

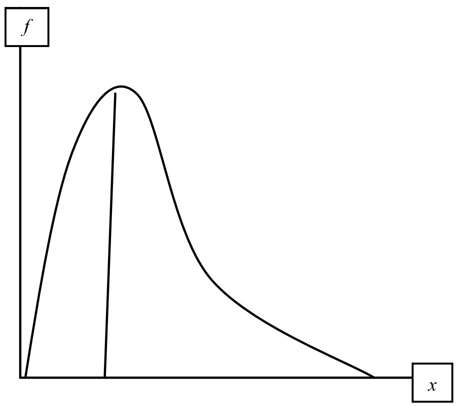

Рисунок 1 – Ряды распределения с правосторонней асимметрией

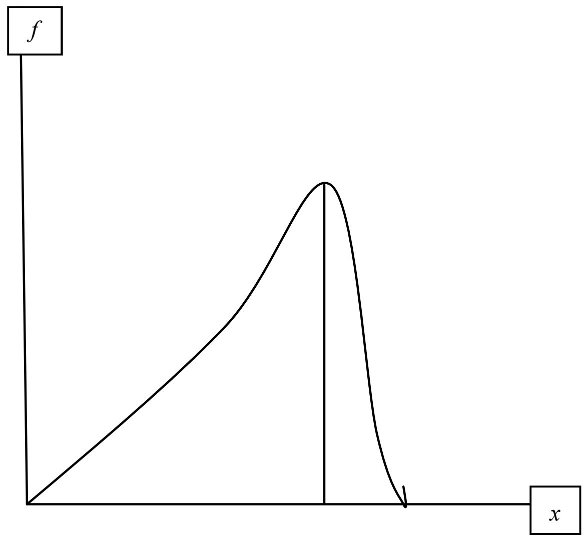

Рисунок 2 – Ряды распределения с левосторонней асимметрией

Другой показатель асимметрии, предложенный шведским математиком Линдбергом, рассчитывают по формуле: $A_s = П - 50$, где

$П$ – процент тех значений признака, которые превосходят по величине среднюю арифметическую.

Наиболее точным и распространенным является показатель, основанный на определении центрального момента третьего порядка (в симметричном распределении его величина равна нулю):

$$A_s = \frac{\mu_3}{\sigma_3}.$$

Применение этого показателя дает возможность не только определить степень асимметрии, но и ответить на вопрос о наличии или отсутствии асимметрии в распределении признака в генеральной совокупности. Оценка степени существенности этого показателя дается с помощью средней квадратической ошибки, которая зависит от объема наблюдений и рассчитывается по формуле:

$$\sigma_{A_S} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}}.$$

Если отношение $\frac{|A_S|}{\sigma_{AS}} > 3$, асимметрия несущественна, ее наличие может быть объяснено влиянием различных случайных обстоятельств.

Для симметричных распределений рассчитывается показатель эксцесса (островершинности). Линдбергом был предложен следующий показатель для оценки эксцесса: $E_x = П - 38{,}29$, где

П = доля (%) количества вариантов, лежащих в интервале, равном половине среднего квадратического отклонения в ту и другую сторону от средней арифметической.

Наиболее точным является показатель, основанный на использовании центрального момента четвертого порядка:

$$E_x = \frac{\mu_4}{\sigma^4} - 3.$$

Эксцесс представляет собой выпад вершины эмпирического распределения вверх или вниз от вершины кривой нормального распределения. В нормальном распределении отношение $\frac{\mu_4}{\sigma^4} = 3$.

Средняя квадратическая ошибка эксцесса рассчитывается по формуле:

$$\sigma_{Ex}\sqrt{\frac{24n(-2)(n-3)}{(n-1)^2(n+3)(n+5)}},$$

где n – число наблюдений.

Оценка существенности показателей асимметрии и эксцесса позволяет сделать вывод, можно ли отнести данное эмпирическое распределение к типу кривых нормального распределения.

Если непрерывная случайная величина имеет плотность распределения

$$f(x) = \frac{1}{\sqrt{2\pi}.\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}},$$

то она подчиняется закону нормального распределения. Для построения кривой нормального распределения надо знать два параметра: $\bar{x}, \sigma$.

Если средняя арифметическая не меняется, но увеличивается среднее квадратическое отклонение, тогда распределение имеет плосковершинный характер (рисунок 3), если же среднее квадратическое отклонение остается постоянной, но с разными значениями средней величины ($\bar{x}_1 < \bar{x}_2 < \bar{x}_3$), в этом случае кривая, не меняя своей формы, сдвигается вправо вдоль оси абсцисс.

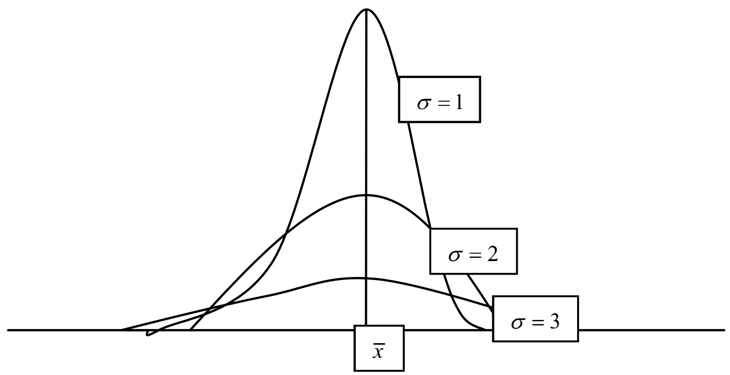

Рисунок 3 – Кривые нормального распределения
с одинаковой $\bar{X}$, но разными $\sigma(\sigma_1 < \sigma_2 < \sigma_3)$

Особенности кривой нормального распределения:

1. Кривая симметрична относительно максимальной ординаты. Максимальная ордината соответствует значению $x = M_o = M_e$, ее величина равна $\frac{1}{\sqrt{2\pi}.\sigma}$.

2. Кривая асимптотически приближается к оси абсцисс, продолжаясь в обе стороны до бесконечности. Следовательно, чем больше значения отклоняются от $\bar{x}$, тем реже они встречаются. Одинаковые по абсолютному значению, но противоположные по знаку отклонения значений переменной x от $\bar{x}$ равновероятны.

3. Кривая имеет две точки перегиба, находящиеся на расстоянии $\pm\sigma$ от $\bar{x}$.

4. При $\bar{x} = const$ с увеличением σ кривая становится более пологой. При $\sigma = const$ с изменением $\bar{x}$ кривая не меняет свою форму, а лишь сдвигается вправо или влево по оси абсцисс.

5. В пределах $\bar{x} \pm \sigma$ находится 68,3 % всех значений признака. В пределах $\bar{x} \pm 2\sigma$ находится 95,4 % всех значений признака. В пределах $\bar{x} \pm 3\sigma$ находится 99,7 % всех значений признака.

Контрольные вопросы, определение понятий и показателей

1. Система показателей для характеристики рядов распределения.
2. Основные показатели и коэффициенты вариации.
3. Виды дисперсий; показатели асимметрии и эксцесса.

3. ВЫБОРОЧНЫЙ МЕТОД СТАТИСТИКИ

3.1. Особенности и теоретические основы выборочного метода

Изучение статистических совокупностей, состоящих из множества единиц, связано со значительными затратами материальных, трудовых и финансовых ресурсов. Задача научно-практических исследователей многих поколений заключалась в том, чтобы на основе обследования определенной части совокупности определить основные характеристики всей генеральной совокупности в целях уменьшения общих затрат.

С развитием математической статистики, а также совершенствования организации и проведения статистических обследований стало возможно на основе изучения определенной части совокупности определить основные характеристики всей совокупности и допустимые расхождения между выборочной и генеральной совокупностями.

Широкое применение выборочного метода на практике стало возможно благодаря тому, что он дает ряд экономических выгод:

1) выборочный метод применяется в тех случаях, когда сплошное наблюдение практически невозможно, поскольку обследование может быть связано с порчей или же с уничтожением материальных ценностей, как например, при контроле качества произведенной продукции;

2) выборочный метод применяется в тех случаях, когда обследуемая совокупность безмерна велика, как например, обследование морского дна на предмет наличия полезных ископаемых;

3) применение выборочного обследования (метода) дает большую экономию во времени и затратах материальных и финансовых средств;

4) применение выборочного метода дает возможность шире и глубже изучить единицы и объект обследования;

5) выборочный метод применяется с целью проверки результатов сплошного наблюдения.

В последнее время выборочные обследования стали широко применяться в различных социально-экономических обследованиях. Так, крупные и средние предприятия охватываются сплошным наблюдением за их деятельностью, а наблюдение за деятельностью малых предприятий производится с помощью выборочных обследований. В ряде случаев выборочные наблюдения применяются в сочетании со сплошными переписями и учетами. Например, программа переписи населения содержат как вопросы сплошного наблюдения, относящиеся ко всему населению, так и вопросы выборочного наблюдения определенной части населения для характеристики занятости, безработицы, трудовой миграции, а также вопросы рождаемости, смертности и уровня жизни населения.

Между основными характеристиками выборочной совокупности и генеральной совокупности всегда имеется определенное расхождение, которое в статистике называют ошибками репрезентативности (представительства). Общая величина возможной ошибки состоит из ошибки регистрации и ошибки репрезентативности.

Ошибки регистрации характерны любому статистическому наблюдению и возникновение их может быть вызваны несовершенством измерительных приборов, недостаточным уровнем квалификации лиц, которые проводят наблюдение и т.д.

Ошибки репрезентативности присущи только не сплошным наблюдениям и представляют собой расхождение между величиной полученных по выборке показателей и величиной этих же показателей, которые могли бы быть получены при проведении сплошного наблюдения.

Ошибки репрезентативности могут быть систематическими и случайными. Систематические ошибки могут возникать в связи с особенностями принятой системы отбора и обработки данных наблюдений, или в связи с нарушением установленных правил отбора. Возникновение случайных ошибок репрезентативности

объясняется недостаточно равномерным представлением в выборочной совокупности различных категорий единиц генеральной совокупности, в силу чего распределение отобранной совокупности единиц не вполне точно воспроизводит распределение единиц генеральной совокупности.

Определение возможной и фактически допущенной ошибки выборки имеет большое значение с применением выборочного метода. Величина ошибки характеризует степень надежности результатов выборки; знание этой величины необходимо при оценке параметров генеральной совокупности. Оценки возможной величины и состава ошибок репрезентативности ложатся в основу планирования проектируемого выборочного наблюдения.

Величина случайной ошибки репрезентативности зависит:

1) от принятого способа формирования выборочной совокупности, его выбор связан с решением вопросов о единице отбора, способе отбора единиц, способе размещения всего объема отбираемых единиц по различным группам генеральной совокупности;

2) от объема выборки;

3) от степени колеблемости изучаемого признака в генеральной совокупности.

Теоретические основы выборочного метода содержатся в теоремах Чебышева и Ляпунова.

Основной предпосылкой применения выборочного метода является возможность судить о характеристиках генеральной (общей) совокупности по отобранной, т. е. выборочной совокупности. Основным принципом в применении выборочного метода является обеспечение равной возможности всем единицам, входящим в состав генеральной совокупности быть избранным в состав выборочной совокупность. Одновременно с обеспечением объективности такого принципа, при котором ни единица не обладает преимуществом попасть в отбираемую совокупность по сравнению с другими единицами, характеристики выборочной совокупности при увеличении объема выборки стремятся к характеристикам генеральной совокупности.

Теорема П.Л. Чебышева (применительно к выборочному методу) математически выражается в виде следующей формулы:

$$P\left[(\tilde{x}-\bar{x})<\frac{t\sigma_o}{\sqrt{n}}\right]>1-\frac{1}{t^2},$$

где $\tilde{x}$ – средняя по совокупности отобранных единиц;

$\bar{x}$ – средняя по генеральной (общей) совокупности;

σ_o – среднее квадратическое отклонение в генеральной (общей) совокупности.

Данная теорема формулируется следующим образом: с вероятностью, сколь угодно близко к единице (достоверности), можно утверждать, что при достаточно большом объеме выборки и ограниченной дисперсии генеральной совокупности разность между выборочной средней ($\tilde{x}$) и генеральной средней ($\bar{x}$) будет сколь угодно мала.

Примечания: 1) выражение $\frac{\sigma_o}{\sqrt{n}}$ часто обозначают μ,

2) при практическом использовании теоремы Чебышева генеральную дисперсию σ_o^2, которая неизвестна, заменяют выборочной дисперсией σ^2.

Теорема А.М. Ляпунова нызывает общие условия, при осуществлении которых распределение суммы независимых случайных величин стремится к нормальному распределению вероятностей. В частности, эта теорема дает возможность оценить погрешность приближенных равенств:

$$\frac{n}{m}\approx p \text{ и } \tilde{x}_{выб}\approx \bar{x}_{ген}$$

при достаточно больших n. Если x_1, x_2, ... x_n – независимые случайные величины и $n\to\infty$, то вероятность их средней $\bar{x}$ находится в пределе от a до b и может быть определена равенством:

$$P(t_1<\bar{x}<t_2)=\int_{t_1}^{t_2}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt,$$

где $t_1=\frac{a-E(x)}{\sigma}$,

$t_2=\frac{b-E(x)}{\sigma}$.

Ограничительные условия теоремы Ляпунова сводятся в основном к тому, чтобы среди слагаемых случайных величин не было сильно выделяющихся (колеблемость которых значительно превосходила бы большинство остальных). В приложении к выборочному методу данная теорема может быть сформулирована следующим образом: при достаточно большом объеме выборки и ограниченной дисперсии генеральной совокупности вероятность того, что разность между выборочной средней и генеральной средней будет в пределах:

$$\pm t\mu=t\frac{\sigma_o}{\sqrt{n}},\ \text{равна}\ F(t)=\frac{1}{\sqrt{2\pi}}\cdot\int_{-t}^{+t}e^{-\frac{t^2}{2}}dt.$$

Формулировка Ляпунова придает теореме Чебышева полную определенность и записывается так:

$$P\left[(\tilde{x}-\bar{x})<t\cdot\frac{\sigma_o}{\sqrt{n}}\right]=F(t)=\frac{1}{\sqrt{2\pi}}\cdot\int_{-t}^{+t}e^{-\frac{t^2}{2}}dt.$$

Теорема Я. Бернулли, (1713 г.) послужила началом разработки совокупности теорем закона больших чисел. Она представляет собой частный случай теоремы Чебышева:

$$P\left[(w-p)<t\sqrt{\frac{pq}{n}}\right]>1-\frac{1}{t^2},$$

где w – доля признака среди отобранных единиц (частость);

p – доля признака в генеральной совокупности.

Теорема Бернулли применяется в тех случаях, когда из генеральной совокупности производится отбор единиц и доля признака не меняется от испытания к испытанию. Формулировка теоремы Бернулли применительно к выборке: с вероятностью, сколь угодно близкой к единице, можно утверждать, что разность между частостью и долей в генеральной совокупности при достаточно большом объеме выборки будет сколь угодно мала. При практическом использовании данной теоремы величина $\sqrt{\frac{pq}{n}}$ рассчитывается путем замены p на w и q на $(1-w)$.

Теорема С. Пуассона также является частным случаем теоремы Чебышева, когда доля признака в генеральной совокупности с ходом выборки все время меняется. В этом случае $\mu^2 = \overline{pq}$, тогда:

$$P\left[(w-p) < t\sqrt{\frac{\overline{pq}}{n}}\right]. > 1-\frac{1}{t^2}.$$

Итак, теоретической основой выборочного метода служат теоремы Чебышева, Ляпунова, Бернулли и Пуассона. Неравенство Чебышева в приложении к выборочному методу может быть сформулировано так: при неограниченном увеличении числа независимых наблюдений ($n \to \infty$) в генеральной совокупности с ограниченной дисперсией с вероятностью, сколь угодно близко к единице, можно ожидать, что отклонение выборочной средней от генеральной средней будет сколь угодно мало, т. е. $P\left(\left|\tilde{x}-\overline{x}\right| \le \varepsilon\right) \to 1$ при $n \to \infty$,

где P – вероятность неравенства, стоящего в скобках;

ε – любое сколь угодно малое положительное число;

$\overline{x}$ – генеральная средняя.

Теорема Чебышева доказывает принципиальную возможность определения генеральной средней по данным простой случайной выборки, но вместе с тем нельзя определить вероятность появления ошибок определенной величины.

Согласно предельной теореме Ляпунова, при достаточно большом числе независимых наблюдений в генеральной совокупности с конечной средней и ограниченной дисперсией, вероятность того, что расхождение между выборочной и генеральной средней $\left|\tilde{x}-\overline{x}\right|$ не превзойдет по абсолютной величине некоторую величину $t\mu$, равна интегралу Лапласа:

$$P\left(\left|\tilde{x}-\overline{x}\right| \le t\mu\right) = F(t),$$

где $F(t)$ – нормированная функция Лапласа.

$$F(t) = \frac{1}{\sqrt{2\pi}}\int_{-t}^{+t} e^{-\frac{t^2}{2}}dt = \frac{2}{\sqrt{2\pi}}\int_{0}^{t} e^{-\frac{t^2}{2}}dt.$$

Величина μ есть средняя квадратическая стандартная ошибка выборки. Из этой теоремы непосредственно следует, что при достаточно большом числе независимых наблюдений, распределение выборочных средних и их отклонений от генеральной средней приближается к нормальному.

Частным случаем теоремы Чебышева является теорема Бернулли:

$$P\left[(w-p) < t\sqrt{\frac{p(1-p)}{n}}\right] > 1-\frac{1}{t^2},$$

где w и p – доля признака в выборочной и генеральной совокупности.

Величина средней квадратической стандартной простой случайной повторной выборки определяется формулой:

$$\mu_{\overline{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}. \qquad (1)$$

Приведенная выше формула показывает, что величина μ_x зависит от колебания признака в генеральной совокупности (чем больше вариация, тем больше и ошибка выборки) и от объема выборки (чем больше обследуется единиц, тем меньше будет величина расхождений выборочных и генеральных характеристик).

Величину $t\mu_{\overline{x}}$ – называют предельной ошибкой выборки, и она выражается в виде следующей формулы: $\Delta_{\overline{x}} = t\mu_{\overline{x}}$ (2), т. е. предельная ошибка выборки равна t – кратному числу средних ошибок выборки. Если допустим, что t = 2, тогда

$$P\left(\left|\tilde{x}-\overline{x}\right| \le 2\mu_{\overline{x}}\right) = F_{(t=2)} = 0{,}9545,$$

т. е. с вероятностью равной 0,9545, можно предполагать, что ошибка выборочной средней не превысит удвоенной средней квадратической ошибки выборки. Кроме того, величина предельной ошибки выборки может быть установлена с определенной вероятностью.

Приведем наиболее часто используемые уровни доверительной вероятности и соответствующие значения t для выборок с объемом ($n \ge 30$):

Таблица 3.1.

t	1,00	1,96	2,00	2,58	3,00
F(t)	0,683	0,950	0,954	0,990	0,997

Данные последней графы показывают вероятность появления ошибки, равной или большей утроенной средней ошибки выборки, т. е. $|\Delta_{\bar{x}}| \geq 3\mu_{\bar{x}}$ крайне мала и равна 0,003(1–0,997). Такие маловероятные события считаются практически невозможными и поэтому величину $|\Delta_{\bar{x}}| - 3\mu_{\bar{x}}$ можно принять за предел возможной ошибки выборки.

Таблица 3.2. – Распределение выборочной доли

Выборочная доля w_j	Число выборок с данной выборочной долей f_j	Отклонение выборочной доли от генеральной $w_j - p$	$w_j f_j$	$(w_j - p)^2 f_j$
0,0	4	–0,5	0.0	1,0
0,5	8	0,0	4,0	0
1,0	4	+0,5	4,0	1,0
Итого	16		8,0	2,0

В среднем для всех возможных вариантов выборок величина выборочной доли совпадает с долей признака в генеральной совокупности:

$$\overline{w} = \frac{\sum_{j=1}^{k} w_j f_j}{\sum_{j=1}^{k} f_j} = \frac{8,0}{16} = 0,5 = p.$$

Средняя квадратическая ошибка для доли по выборке:

$$\mu_p = \sqrt{\frac{\sum_{j=1}^{k}(w_j - p)^2 f_j}{\sum_{j=1}^{k} f_j}} = \sqrt{\frac{2,0}{16}} = 0,354.$$

Ранее было указано, что дисперсия доли $\sigma_p^2 = p(1-p)$, поэтому величину средней стандартной ошибки выборочной доли можно определить по следующей формуле:

$$\mu_p = \sqrt{\frac{p(1-p)}{n}}, \qquad (2)$$

а предельную ошибку доли по формуле: $\Delta_p = t\mu_p$. (4)

Исходя из полученных данных предыдущего примера, где $p = 0,5; n = 2;$

определим среднюю ошибку выборки: $\mu_p = \sqrt{\frac{0,5 \cdot 0,5}{2}} = 0,354.$

Ошибка репрезентативности (ε) представляет собой разность между характеристиками выборочной и генеральной совокупности. Генеральная средняя ($\overline{x}$) вычитается из выборочной средне ($\tilde{x}$) или доля признака в генеральной совокупности (p) вычитается из доли признака в выборочной совокупности, т. е. частости (w); $\varepsilon = (\tilde{x} - \overline{x}); \varepsilon = (w - p)$.

Если Δ представляет собой предел, которого не превосходит абсолютная величина ε, это соотношение выражается так:

$$\tilde{x} - \Delta \leq \overline{x} \leq \tilde{x} + \Delta;$$

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

В формулах выборочного метода использована дисперсия генеральной совокупности (σ_o^2). Однако при проведении выборочного обследования характеристики генеральной совокупности неизвестны. В практической работе с учетом небольшой погрешности можно заменить дисперсию генеральной совокупности дисперсией выборочной совокупности (σ^2) и использовать в нахождении следующих показателей:

$\sigma^2 = \frac{\sum (x - \tilde{x})^2}{n}$ – дисперсия выборочной совокупности (невзвешанная);

$\sigma^2 = \frac{\sum (x - \tilde{x})^2 m}{\sum m}$ – дисперсия выборочной совокупности (взвешанная);

$\sigma^2 = w(1 - w)$ – дисперсия выборочной совокупности для частости.

3.2. Виды выборки. Простая случайная выборка

Развитие теории выборочного метода началось с развития и совершенствования простой случайной выборки. Основные понятия и методы, лежащие в основе случайной выборки, послужили основой при разработке других видов выборки.

Отличительная особенность простой случайной выборки в том, что при ней обеспечивается принцип равенства и независимости, при которой каждая единица совокупности имеет равную вероятность быть представленной в выборочной совокупности.

Простая случайная выборка осуществляется на основе повторного и бесповторного отбора. При повторной случайной выборке каждая единица генеральной совокупности имеет равную возможность, вероятность быть включенной в выборочную совокупность, и она принимает участие в формирование выборочной совокупности столько раз, сколько единиц в выборочной совокупности.

При бесповторной выборке каждая последующая единица, которая еще не принимала участие в формировании выборочной совокупности, имеет большую вероятность попадания в выборочную совокупность.

Применение простой случайной повторной выборки связано с определенными недостатками, поскольку практически невыгодно наблюдение одних и тех же единиц по несколько раз. Применение бесповторного отбора взамен повторного повышает степень репрезентативности выборки, поскольку в выборочную совокупность включаются новые еще не обследованные единицы совокупности.

Особое значение имеет определение средней ошибки репрезентативности. Эти ошибки рассчитывают отдельно для среднего значения признак и для доли изучаемого признака с учетом способа отбора(таблица 3.3).

Таблица 3.3. – Формулы средней ошибки при простой случайной выборке

Средняя ошибка выборки (μ)	Повторный способ отбора	Бесповторный способ отбора
Для средней	$\mu=\sqrt{\frac{\sigma^2}{n}}$	$\mu=\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$
Для доли	$\mu=\sqrt{\frac{w(1-w)}{n}}$	$\mu=\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}$

Пример: Из совокупности 10 000 деталей отобрано случайным бесповторным методом 1000 деталей, для которых средний вес детали оказался равным 50г, дисперсия 49; бракованных деталей – 20 штук. Необходимо определить: средние ошибки выборки для средней и для доли.

Дано: $N=10000;\ n=1000;\ \tilde{x}=50;\ \sigma^2=49;$

$$w=\frac{20}{1000}=0,02.$$

Применяя формулы, приведенные в таблице 3.3. определим средние ошибки для среднего веса детали при бесповторном способе отбора:

$$\mu=\sqrt{\frac{49}{1000}\cdot\left(1-\frac{1000}{10000}\right)}=0,21$$

и для доли брака:

$$\mu=\sqrt{\frac{0,02\cdot 0,98}{1000}\left(1-\frac{1000}{10000}\right)}=0,0042\approx 0,42\ \%.$$

3.3. Численность выборки

При проектировании выборочного наблюдения предполагают заранее заданными величину допустимой ошибки выборки и ее вероятность. Неизвестным остается тот минимальный объем выборки, который должен обеспечить требуемую точность. Используя формулу: $\Delta = t\mu$ и формулу средних ошибок выборки устанавливают необходимую численность выборки.

Необходимая численность выборки при случайном способе отбора определяют по следующей формуле:

Таблица 3.4.

Способ отбора	Для средней	Для доли	Для доли, если она неизвестна
Повторный	$n=\frac{t^2\sigma^2}{\Delta^2}$	$n=\frac{t^2 w(1-w)}{\Delta^2}$	$n=\frac{0,25t^2}{\Delta^2}$
Бесповторный	$n=\frac{t^2\sigma^2 N}{N\Delta^2+t^2\sigma^2}$	$n=\frac{t^2 Nw(1-w)}{N\Delta^2+t^2 w(1-w)}$	$n=\frac{0,25t^2 N}{N\Delta^2+0<25t^2}$

Примечание: при проектировании объема необходимой выборки величина σ^2 и w неизвестны, поэтому вместо точного их значения берут приближенные, установленные на основании уже проведенного другого наблюдения или нескольких пробных наблюдений, выбирая из найденных результатов наибольшие значения σ^2, w, $(1-w)$.

Пример: Проектируется выборочное наблюдение, целью которого является установление среднего размера деталей в совокупности, состоящей из 10 000 деталей. Требуемая точность 1 см. Произведенные пробные выборки дали наибольшую дисперсию, равную 49. Требуется определить необходимую численность случайной бесповторной выборки, обеспечивающей с вероятностью 0,95 заданную точность.

Дано: $N=10000$; $\Delta=1$; $F(t)=0,95$; $\sigma^2=49$.

Используя расчетное приложение, находим по $F(t)$ значение t = 1,96 и по формуле для бесповторной выборки, взятой из таблицы 3.4, определяем:

$$n=\frac{t^2\sigma^2 N}{N\Delta^2+t^2\sigma^2}=\frac{1.96^2\cdot 49\cdot 10000}{10000\cdot 1^2+1,96^2\cdot 49}=184,7\approx 185.$$

Механическая выборка. Разновидностью простой случайной выборки является механическая выборка. При механической выборке отбор единиц из генеральной совокупности производится через равные интервалы, например, отбирается каждая десятая

или же сотая единица совокупности или же отбор изучаемых лиц из общей совокупности производится в алфавитном порядке и т.д.

Поскольку в основе механической выборки лежат принципы непредвзятости, случайности и отбор единиц производится непосредственно из генеральной совокупности, то при определении средней и предельной ошибок репрезентативности используются те же формулы, которые используются при простой случайной выборке. Однако, учитывая особенности ее проведения, применяется в основном бесповторный способ отбора.

3.4. Типическая выборка

При типической выборке генеральная совокупность по какому-либо существенному признаку делится на типические группы. Количество отбираемых единиц из каждой типической группы устанавливается в следующем порядке.

При пропорциональном отборе объему типических групп численность отбираемых единиц в каждой типической группе определяется по формуле:

$$n_i=n\cdot\frac{N_i}{N},$$

где n_i – объем выборки из *i–ой* типической группы;

n – общий объем выборки;

N_i – объем *i–ой* типической группы;

N – объем генеральной совокупности.

При отборе с учетом колеблемости признака, дающим наименьшую величину ошибки выборки, процент выборки из каждой типической группы должен быть пропорционален среднему квадратическому отклонению в этой группе (σ_i). Расчет численности (n_i) производится по следующей формуле:

$$n_i=\frac{nN_i\sigma_i}{\sum N_i\sigma_i}$$ – для средней;

$$n_i=\frac{n_i N_i\sqrt{w_i(1-w_i)}}{\sum N_i\sqrt{w_i(1-w_i)}}$$ – для доли.

Для определения средних ошибок при типической выборке используются следующие формулы (таблицы 3.5; 3.6.).

Таблица 3.5. – Средние ошибки выборки (μ)
при повторном типическом методе отбора

Способ отбора	Для средней	Для доли
Не пропорциональный объему групп	$\frac{1}{N}\sqrt{\sum\frac{\sigma_i^2}{n_i}N_i^2}$	$\frac{1}{N}\sqrt{\sum\frac{w(1-w_i)}{n_i}N_i^2}$
Пропорциональный объему групп	$\sqrt{\frac{\overline{\sigma_i^2}}{n}}$	$\sqrt{\frac{\overline{w(1-w)}}{n}}$
Пропорциональный колеблемости в группах	$\frac{1}{N}\frac{\sum\sigma_i N_i}{\sqrt{n_i}}$	$\frac{1}{N}\frac{\sum w_i(1-w_i)N_i}{\sqrt{n_i}}$

Таблица 3.6. – Средние ошибки выборки
при бесповторном типическом методе отбора

Способ отбора	Для средней	Для доли
Непропорциональный объему групп	$\frac{1}{N}\sqrt{\sum\frac{\sigma_i^2}{n_i}N_i^2\left(1-\frac{n_i}{N_i}\right)}$	$\frac{1}{N}\sqrt{\sum\frac{w_i(1-w_i)}{n_i}N_i^2\left(1-\frac{n_i}{N_i}\right)}$
Пропорциональный объему групп	$\sqrt{\frac{\overline{\sigma_i^2}}{n}\left(1-\frac{n}{N}\right)}$	$\sqrt{\frac{\overline{w(1-w)}}{n}\left(1-\frac{n}{N}\right)}$
Пропорциональный колеблемлемости в группах	$\frac{1}{N}\cdot\frac{\sum\sigma_i N_i}{\sqrt{n_i}}\sqrt{1-\frac{n}{N}}$	$\frac{1}{N}\cdot\frac{\sum w_i(1-w_i)N_i}{\sqrt{n_i}}\sqrt{1-\frac{n}{N}}$

Обозначения:

$\overline{\sigma_i^2}$ – средняя из выборочных дисперсий типических групп;

$\overline{w(1-w)}$ – средняя из произведения частот на дополнение их до единицы;

σ_i^2 – выборочная дисперсия *i–й* типической группы;

σ_i – среднее квадратическое отклонение в выборке из i-й типической группы.

Пример: Для определения основных характеристик из совокупности в 10 000 единиц производится выборка типическим методом, Вся совокупность делится на 5 типических групп, отбор единиц внутри типических групп производится случайным бесповторным методом пропорционально объему каждой группы, всего отбирается 2000 единиц. При обследовании отобранных единиц получены следующие результаты: (таблица 3.7.).

Таблица 3.7.

Типические группы (i)	Численность групп (N_i)	Численность выборки (n_i)	Выборочная средняя ($\overline{x}_i$)	Выборочная дисперсия (σ_i^2)
1	2	3	4	5
1)	500	100	10	4
2)	2000	400	12	6
3)	3000	600	15	10
4)	1500	300	18	20
5)	3000	600	20	16
Всего	10 000	2000	16,1	22,89

Определить: а) среднюю ошибку для каждой группы и для всей выборочной совокупности (при простом случайном и типическом способах отбора); б) границы, в которых с вероятностью 0,997 находится генеральная средняя по группам и по всей совокупности (при простом случайном и типическом способах отбора).

Расчет численности отбираемых единиц из каждой типической группы пропорционально ее объему производится в следующей последовательности: численность первой типической группы при заданном объеме всей выборки в 2000 единиц равна:

$n_1 = 2000\frac{500}{10000} = 100$ единиц;

для второй типической группы: $n_2 = 2000\frac{2000}{10000} = 400$ единиц;

для третьей типической группы: $n_3 = 2000\frac{3000}{10000} = 600$ единиц; и т.д.

Для определения средней ошибки выборки по группам и общей средней ошибки выборки при простом случайном бесповторном способе отбора используем формулы таблицы 3 и рассчитаем среднюю ошибку выборки:

для первой типической группы: $\mu_1 = \sqrt{\frac{4}{100}\left(1 - \frac{100}{500}\right)} \approx 0{,}18;$

для второй типической группы: $\mu_2 = \sqrt{\frac{6}{400}\left(1 - \frac{400}{2000}\right)} \approx 0{,}11;$

и т.д. по всем типическим группам.

Для расчета средней ошибки выборки всей совокупности при простом случайном методе отбора и границ генеральной средней необходимо знать общую выборочную среднюю и общую дисперсию выборочной совокупности. Определение общей выборочной средней производится на основе использования следующей формулы:

$$\tilde{x} = \frac{\sum \tilde{x}_i n_i}{\sum n_i} = \frac{10 \cdot 100 + 12 \cdot 400 + 15 \cdot 600 + 18 \cdot 300 + 20 \cdot 600}{2000} = 16{,}1.$$

Для определения общей выборочной дисперсии предварительно необходимо рассчитать среднюю дисперсию из групповых типических дисперсий и среднюю дисперсию между типическими группами. Затем, используя правило сложения дисперсий, находим общую выборочную дисперсию, которая равна сумме средней дисперсии из типических групп и средней межгрупповой типической дисперсии.

Средняя дисперсия из выделенных типических групп:

$$\overline{\sigma_i^2} = \frac{\sum \sigma_i^2 \cdot n_i}{\sum n_i} = \frac{4 \cdot 100 + 6 \cdot 400 + 10 \cdot 600 + 20 \cdot 300 + 16 \cdot 600}{2000} = 12{,}2;$$

Межгрупповая дисперсия:

$$\overline{\delta^2} = \frac{\sum (\tilde{x}_i - \tilde{x})^2 \cdot n_i}{\sum n_i} = \frac{(-6{,}1)^2 \cdot 100 + (-4{,}1)^2 \cdot 400 + (-1{,}1)^2 600 + (1{,}9)^2 300 + (3{,}9)^2 \cdot 600}{2000} = 10{,}69.$$

Общая дисперсия выборочной совокупности:

$$\sigma^2 = \overline{\sigma_i^2} + \overline{\delta^2} = 12{,}2 + 10{,}69 = 22{,}89.$$

Средняя ошибка выборки всей совокупности при случайном методе отбора:

$$\mu = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{22{,}89}{2000}\left(1 - \frac{2000}{10000}\right)} \approx 0{,}096.$$

Предельная ошибка случайной выборки:

$$\Delta = t\mu \approx 3 \cdot 0{,}096 = 0{,}29.$$

Предельная величина генеральной средней при простом случайном методе отбора: $\overline{x} = \tilde{x} \pm \Delta = 16{,}1 \pm 0{,}29;\ 15{,}81 < \overline{x} < 16{,}39.$

Средняя ошибка типической выборки, пропорциональной объему типических групп определяется по формуле:

$$\mu \sqrt{\frac{\overline{\sigma_i^2}}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{12{,}2}{2000}\left(1 - \frac{2000}{10000}\right)} \approx 0{,}070.$$

Величина ошибки типической выборки равна:

$$\Delta = 3 \cdot 0{,}07 = 0{,}21.$$

Предельная граница генеральной средней равна:

$$\overline{x} = 16{,}1 \pm 0{,}21, \text{ т. е. } 15{,}89 < \overline{x} < 16{,}31.$$

3.5. Серийно-гнездовая выборка

При серийной выборке в случайном порядке отбираются группы (серии, гнезда) единиц, которые подвергаются сплошному обследованию. Серийная выборка по сравнению с другими видами выборки широко используется в тех случаях, если генеральная совокупность состоит из обособленных групп единиц и в них необходимо провести обследование как, например, когда изучаемая совокупность состоит из отдельных территориальных единиц и в каждой единице необходимо вести обследование.

При проведении серийного обследования в предварительно отобранных сериях изучение единиц совокупности производится сплошным образом, т. е. все единицы, входящие в состав данной отобранной серии, подвергаются всестороннему изучению без исключения. При этом общая средняя $\tilde{x}_o$, рассматриваемая как оценка генеральной средней, будет зависеть от того, как, насколько точно средние серий $\tilde{x}_i$ будут репрезентировать генеральную среднюю. Это значит, что случайная ошибка средней по всей выборке зависит от величины межсерийной (межгрупповой) дисперсии δ^2, характеризующий степень колеблемости серийных средних.

При равенстве в сериях и при повторном отборе серий:

$$\widetilde{x}_o = \frac{\sum_{i=1}^{s} \widetilde{x}_i}{s}, \ \delta^2 = \frac{\sum_{i=2}^{s} (\widetilde{x}_i - \widetilde{x}_o)^2}{s}, \ \mu_{\widetilde{x}} = \sqrt{\frac{\delta^2}{s}},$$

где s – число отобранных серий.

При бесповторном отборе серий $\mu_{\widetilde{x}} = \sqrt{\frac{\delta^2}{s}\left(\frac{S-s}{S-1}\right)} = \left(1-\frac{s}{S}\right)$,

где S – общее число серий в генеральной совокупности,

s – число отобранных серий.

Из приведенных формул следует, что случайная ошибка серийной выборки будет меньше с уменьшением величины межсерийной дисперсии и чем, следовательно, более однородны единицы признака в сериях.

Поскольку при серийной выборке, как правило, обследуется небольшое число серий, то случайная ошибка серийной выборки несколько больше, чем при других способах отбора, т. е. серийная выборка менее точна, чем выборка, основанная на индивидуальном отборе. Тем не менее, серийный отбор широко применяется в практике выборочных обследований, особенно в тех случаях, когда обследование охватывает обширную территорию и сериями выступают территориальные единицы, что связано со значительной экономией финансовых средств на проведение обследования.

Средняя ошибка выборочной доли при случайной выборке в случае равновеликих серий определяется по формулам:

а) при повторном отборе:

$$\mu_p \sqrt{\frac{\delta_p^2}{s}},$$

где δ_p^2 – межсерийная (межгрупповая) дисперсия доли, определяемая по формуле:

$$\delta_p^2 = \frac{\sum_{i=1}^{s} \left(w_i - \overline{w}\right)^2}{s},$$

где w_i – доли признака i – й серии;

$\overline{w}$ – доля признака по всей выборочной совокупности.

б) при бесповторном отборе:

$$\mu_p = \sqrt{\frac{\delta_p^s}{s}\left(1-\frac{s}{S}\right)}.$$

Пример: Генеральная совокупность состоит из 5000 единиц, разбитых на 50 равных по величине серий (по 100 единиц). Бесповторным методом отобрано 10 серий. Результаты выборки представлены в таблице 3.8.

Таблица 3.8.

Номера серий	1	2	3	4	5	6	7	8	9	10
Средний размер признака ($\widetilde{x}_i$)	51	54	60	61	62	70	73	77	79	92
Частная дисперсия (σ_i^2)	25,0	23,2	16,1	35,9	45,8	91,2	88,3	77,8	110,5	145,1

Необходимо определить среднюю ошибку серийной бесповторной выборки. Для этого предварительно рассчитаем: а) общую среднюю всей выборочной совокупности по серийным средним:

$$\tilde{x} = \frac{\sum \tilde{x}_i}{s} = \frac{679}{10} = 67,9;$$

б) межсерийную (межгрупповую) дисперсию средних:

$$\overline{\delta_{\tilde{x}}^2} = \frac{\sum (\tilde{x}_i - \tilde{x})^2}{s} = \frac{(51-67,9)^2 + (54-67,9)^2 + \cdots + (92-67,9)^2}{10} = \frac{1440,9}{10} = 144,09;$$

в) среднюю ошибку серийной выборки:

$$\mu_{\tilde{x}} = \sqrt{\frac{\overline{\delta_{\tilde{x}}^2}}{s}\left(\frac{S-s}{S-1}\right)} = \sqrt{\frac{144,09}{10} \cdot \left(\frac{50-10}{50-1}\right)} = \sqrt{11,76} \approx 3,43.$$

Необходимая численность отбираемых серий при серийном отборе рассчитаем, используя формулы таблицы 4., в которых N, n, σ^2 заменяем их на $S, s, \overline{\delta_{\tilde{x}}^2}$.

Пример: Совокупность разбита на 50 серий и межсерийная дисперсия равна 16. Сколько серий необходимо отобрать бесповторным методом, чтобы с вероятностью 0,954 утверждать, что ошибка выборочной средней не превысит 2,3?

Дано: $\overline{\delta^2} = 16$; $S = 50$; $t = 2$; $\Delta = 2,3$.

Определим необходимое число серий, отбор которых обеспечит требуемую точность:

$$s = \frac{t^2 \overline{\delta^2} S}{S\Delta^2 + t^2 \overline{\delta^2}} = \frac{4 \cdot 16 \cdot 50}{50 \cdot 5,29 + 4 \cdot 16} = \frac{3200}{328,5} \approx 10 \text{ серий.}$$

Контрольные вопросы, определение понятий и показателей

1. Особенности применения выборочного метода; ошибки выборки.
2. Охарактеризуйте основные виды и формы выборочного метода.
3. Как определяется необходимая численность выборки?
4. Определите величину ошибки для средней и доли при простой выборке.
5. Определите ошибки репрезентативности при типической выборке.

4. СТАТИСИЧЕСКИЕ ПОКАЗАТЕЛИ КОРРЕЛЯЦИОННОЙ СВЯЗИ

4.1. Формы и виды связи. Уравнение регрессии

На изменение социально-экономических явлений и процессов воздействуют совокупность факторов, одни из которых выступают в качестве причины, а другие в качестве следствия. Признаки, которые выступают в качестве причины, в статистике называют *факторными признаками*, а признаки, которые изменяются под воздействием факторных признаков, называют *результативными признаками.*

При изучении взаимосвязи явлений выделяют два вида связи: функциональную и статистическую, или корреляционную. При функциональной связи каждому значению факторного признака строго соответствует одно или вполне ограниченное число результативного признака. Примером функциональной связи является зависимость между площадью и радиусом круга, где $S = \pi R^2$.

При корреляционной связи значение результативного признака устанавливается под воздействием нескольких значений факторного признака. Совокупное воздействие факторных признаков в среднем формирует значение результативного признака. Примером может служит формирование уровня производительности труда, когда она формируется под воздействием совокупности факторных признаков, таких как качество техники и технологии, которые используются в производстве, организация производства и труда, уровень квалификации специалистов и рабочих кадров и т.д.

Изучение корреляционных зависимостей основывается на исследовании таких связей между переменными, при которых значения одной зависимой переменной в среднем изменяются в зависимости от того, какие значения принимает другая переменная, рассматриваемая как причина по отношению к зависимой переменной. Поскольку воздействие различных причин (факторов) и случайностей по степени и направлению могут быть разными, то тенденция и закономерности изменения результативного признака могут быть искаженными. Вычисляя средние значения результативного признака для данной группы значений признака – фактора, в определенной степени элиминируем влияние случайностей. Вычисляя параметры теоретической линии связи, производим дальнейшее их элиминирование и в результате получаем однозначное по форме изменение результативного признака (y) с изменением факторного признака (x).

Теоретической линией регрессии называется та линия, вокруг которой группируются точки корреляционного поля и которая указывает основное направление, основную тенденцию связи. Теоретическая линия регрессии должна отображать изменение средних величин результативного признака (y) по мере изменения величин факторного признака (x) при условии полного взаимопогашения всех прочих случайных по отношению к фактору (x) причин. Следовательно, эта линия должна быть проведена так, чтобы сумма отклонений точек поля корреляции от соответствующих точек теоретической линии регрессии равнялась нулю, а сумма квадратов этих отклонений была минимальной величиной.

Важным этапом регрессионного анализа является определение типа функции, с помощью которой характеризуется зависимость между признаками. Основанием для выбора вида уравнения должен служить теоретический анализ сущности изучаемой зависимости и механизма их действия. На основе теоретического анализа могут быть сделаны общие выводы относительно направления связи и возможности использования линейной зависимости на практике.

Приблизительное представление о линии связи можно получить на основе эмпирической линии регрессии, Эмпирическая линия обычно является ломанной линией со значительным изломом. Это объясняется тем, что влияние прочих неучтенных факторов, оказывающих воздействие на вариацию результативного признака, в средних погашается не полностью, в силу недостаточно большого количества наблюдений, поэтому эмпирической линией связи для

выбора и обоснования типа теоретической кривой можно воспользоваться при условии, что число наблюдений будет достаточно велико.

Наиболее часто для характеристики связей социально-экономических явлений и процессов используется следующие виды взаимосвязи:

линейная – $\hat{y} = a_0 + a_1 x$;

параболическая – $\hat{y} = a_0 + a_1 x + a_2 x^2$;

гиперболическая – $\hat{y} = a_0 + a_1 \dfrac{1}{x}$; и т.д.

4.2. Линейное уравнение связи

Если эмпирическая линия регрессии больше приближается к прямой, то теоретическая линия регрессии может быть представлена линейным уравнением: $\hat{y} = a_0 + a_1 x$.

Для нахождения параметров a_0, a_1 уравнения регрессии используем метод наименьших квадратов. При применении метода наименьших квадратов для нахождения такой функции, которая соответствует эмпирическим данным, считается, что сумма квадратов отклонений эмпирических точек от теоретической линии регрессии должна быть величиной минимальной.

Критерий метода наименьших квадратов выражается в виде следующей формулы:

$$S = \sum_{i=1}^{n} \left(y_i - \hat{y}(x)\right)^2 \to \min;$$

поскольку $\hat{y}_x = a_0 + a_1 x$, $S = \sum_{1}^{n} \left[y_i - (a_0 + a_1 x)\right]^2 \to \min$

Применение метода наименьших квадратов для определения параметров a_0, a_1 прямой, наиболее соответствующей эмпирическим данным, сводится к нахождению экстремума.

Функция двух переменных $S(a_0, a_1)$ может достигнуть экстремума в том случае, если первые частные производные этой функции равняются нулю, т. е. когда

$$\frac{dS}{da_0} = 0; \frac{dS}{da_1} = 0.$$

Вычисляя эти частные производные, получим:

$$\frac{dS}{da_0} = -2\sum_1^n (y_i - a_0 - a_1 x_i) = 0,$$

$$\frac{dS}{da_1} = -2\sum_1^n (y_i x_i - a_0 x_i - a_1 x_i^2) = 0.$$

После соответствующих преобразований получим систему нормальных уравнений способа наименьших квадратов для определения величины параметров a_0, a_1 уравнения прямолинейной корреляционной связи по эмпирическим данным:

$$a_0 n + a_1 \sum_1^n x_i = \sum_1^n y_i,$$

$$a_0 \sum_1^n x_i + a_1 \sum_1^n x_i^2 = \sum_1^n y_i x_i.$$

Примером расчета параметров уравнения и средних значений результативного признака $\hat{y}_x$ могут служить данные группировки по факторному признаку и подсчету средних показателей по результативному признаку.

Группировка предприятий по стоимости основных фондов и подсчет сумм, необходимых для нахождения уравнения связи (таблица 4.1.):

Таблица 4.1.

Стоимость основных фондов (млн сом.)	Выработка продукции на 1 работника (тыс. сом.)	Среднее значение интервала	x^2	yx	$\hat{y}_x$
до 1	4	0,5	0,25	2,00	4,35
1–2	6	1,5	2,25	9,00	5,21
2–3	5,5	2,5	6,25	13,75	6,07
3–4	7,0	3,5	12,25	24,50	6,93
4–5	8,0	4,5	20,25	36,00	7,79
св .5	8,5	5,5	30,25	46,75	8,65
Итого	39,0	18,0	71,50	132,00	39,00

Используя данные таблицы, строим систему двух уравнений с двумя неизвестными:

$$6a_0 + 18a_1 = 39,0; \quad 18a_0 + 71,5a_1 = 132,0.$$

Разделив каждый член в обоих уравнениях на коэффициенты при a_0, получим:

$$a_0 + 3a_1 = 6,5;$$

$$a_0 + 3,97a_1 = 7,33.$$

Вычитая из второго уравнения первое, получим: $0,97a_1 = 0,83$, $a_1 = 0,86$.

Уравнение связи примет вид: $\hat{y}_x = 3,92 + 0,86x$. Подставив в это уравнение соответствующие значения x, получим значения результативного признака, отражающие среднюю зависимость (y) от (x) в виде корреляционной зависимости, при этом суммы, исчисленные по уравнению и фактические должны быть равны между собой.

Изображение фактических и теоретически вычисленных значений выработки продукции, т. е. результативного признака представлен на рисунке 1.

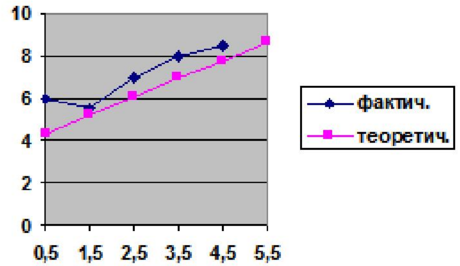

Рисунок 1 – Результативный признак

4.3. Параболическое уравнение связи

Параболическая зависимость, выражаемая уравнением параболы 2-го порядка $\hat{y}_x = a_0 + a_1 x + a_2 x^2$, имеет место при ускоренном возрастании или убывании результативного признака в сочетании с равномерным возрастанием факторного признака.

Параметры уравнения параболы $a_0; a_1; a_2$ вычисляются путем решения системы трех нормальных уравнений:

$$na_0 + a_1 \sum x + a_2 \sum x^2 = \sum y;$$

$$a_0\sum x + a_1\sum x^2 + a_2\sum x^3 = \sum yx;$$

$$a_0\sum x^2 + a_1\sum\ x^3 + a_2\sum x^4 = \sum ex^2.$$

Используя данные выпуска продукции (у) и стоимости основных производственных фондов (х), которые приведены в таблице 4.2. составим уравнение параболы второго порядка, характеризующее взаимозависимость факторных и результативных признаков.

Таблица 4.2

Группа предприятий по стоимости основных фондов (млн сом.)	Выпуск продукции (млн сом.) (у)	Среднее значение интервала (х)	x^2	x^3	x^4	xy	xy^2	$\hat{y}_x$
0,5–1,5	10	1	1	1	1	10	10	8,1
1,5–2,5	12	2	4	8	16	24	48	17,6
2,5–3,5	28	3	9	27	81	84	252	26,7
3,5–4,5	40	4	16	64	256	160	640	35,5
4,5–5,5	42	5	25	125	625	210	1050	44,0
5,5–6,5	52	6	36	216	1296	312	1872	52,2
Итого	184	21	91	441	2275	800	3872	184,1

Параметры уравнения параболы $a_0; a_1; a_2$ определяются на основе решения системы трех нормальных уравнений:

$$na_0 + a_1\sum x + a_2\sum x^2 = \sum y;$$

$$a_0\sum x + a_1\sum x^2 + a_2\sum x^3 = \sum yx;$$

$$a_0\sum x^2 + a_1\sum x^3 + a_2\sum x^4 = \sum yx^2.$$

По данным таблицы 4.2. составим систему уравнений:

$$6a_0 + 21a_1 + 91a_2 = 184;$$

$$21a_0 + 91a_1 + 441a_2 = 800;$$

$$91a_0 + 441a_1 + 2275a_2 = 3872.$$

После деления всех уравнений на коэффициенты при a_0 получим:

$$a_0 + 3,5a_1 + 15,67a_2 = 30,67;$$

$$a_0 + 4,33a_1 + 21,0a_2 = 38,095;$$

$$a_0 + 4,846a_1 + 25,0a_2 = 42,55.$$

Вычитая из второго уравнения первое и из третьего второе, получим два новых уравнения с двумя неизвестными:

$$0,833a_1 + 5,333a_2 = 7,428;$$

$$0,513a_1 + 4,00a_2 = 4,454.$$

Полученные уравнения снова разделим на коэффициенты при a_1:

$$a_1 + 6,40a_2 = 8,91;$$

$$a_1 + 7,80a_2 = 8,68.$$

Вычитая из второго уравнения первое, получим:

$$1,40a_2 = -0,228; a_2 = -0,1631.$$

$$a_1 = 8,914 - 6,400(-0,163) = 9,958;$$

$$a_0 = 30,666 - 34,854 + 2,555 = -1,632.$$

В результате получим уравнение параболы, выражающей связь между факторным признаком (х) и результативным признаком (у):

$$\hat{y}_x = -1,6323 + 9,958x - 0,1631x^2.$$

Графическое изображение фактических данных и данных теоретического расчета показывает, несмотря на небольшое отклонения линий, основное направление их в основном совпадают, т. е. основное направление изменения фактических данных почти совпадают с направлением изменения, найденное по уравнению параболы (рисунок 2).

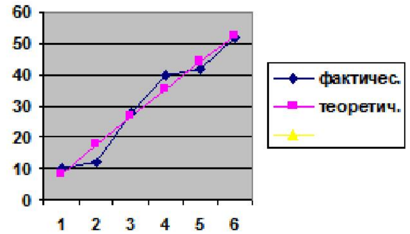

Рисунок 2 – Графическое изображение фактических данных и данных теоретического расчета

В практике изучения связи между признаками, кроме параболы второго порядка, применяются параболы более высоких порядков. Чем выше порядок параболы, тем точнее он воспроизводит фактические данные.

4.4. Уравнение гиперболы

Обратная связь указывает на уменьшение результативного признака при возрастании факторного признака. Обратная зависимость между факторными и результативными признаками выражается уравнением гиперболы:

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

Параметры уравнения гиперболы a_0, a_1 определяют, используя систему нормальных уравнений:

$$na_0 + a_1 \sum \frac{1}{x} = \sum y;$$

$$a_0 \sum \frac{1}{x} + a_1 \sum \frac{1}{x^2} = \sum \frac{y}{x},$$

где $\sum \frac{1}{x}$ – сумма величин, обратных значениям факторного признака, а $\sum \frac{1}{x^2}$ – сумма их квадратов.

Определение влияния на оснащенность предприятий основными производственными фондами на себестоимость единицы выпускаемой продукции на основе применения уравнения гиперболы приведены в таблице 4.3.

Таблица 4.3.

Группа предприятий по стоимости основных фондов (млн сом.)	Себестоимость единицы продукции (у)	x	$\frac{1}{x}$	$\frac{1}{x^2}$	$\frac{y}{x}$	$\hat{y}_x$
до 1	15	0,5	2,00	4,0000	30,00	15,05
1–2	11	1,5	0,67	0,4489	7,33	11,62
2–3	12	2,5	0,40	0,1600	4,80	10,92
3–4	12	3,5	0,28	0,0784	3,43	10,62
4–5	9	4,5	0,22	0,0484	2,00	10,46
св. 5	10	5,5	0,18	0,0364	1,82	10,33
Итого	69	18,0	3,75	4,7721	49,38	69,00

Составив по данным таблицы систему уравнений и разделив каждый член обоих уравнений на коэффициенты при a_0, получим:

$$6a_0 + 3{,}75a_1 = 69; \;\; a_0 + 0{,}62a_1 = 11{,}50;$$

$$3{,}75a_0 + 4{,}77a_1 = 49{,}38; \;\; a_0 + 1{,}27a_1 = 13{,}18.$$

Вычитая из второго уравнения первое, найдем значение a_1.

$$0{,}65a_1 = 1{,}68; a_1 = 2{,}58.$$

Подставив вместо a_1 его значение, получим a_0

$$a_0 = 11{,}50 - 0{,}62 \cdot 2{,}58 = 11{,}50 - 1{,}60 = 9{,}90.$$

Получим уравнение в общем виде: $\hat{y}_x = 9{,}90 + \frac{2{,}58}{x}$,

Подставив каждое значение x в уравнение, находим $\hat{y}_x$ по каждой строке таблицы. Построим ломаную линию по парам x и y, а также по x и $\hat{y}_x$ (рисунок 3).

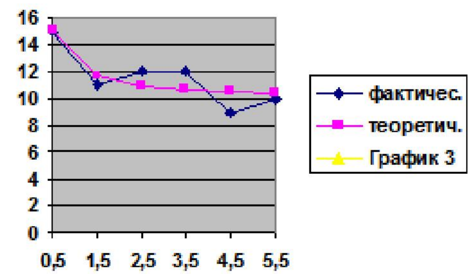

Рисунок 3 – Ломаная линия

4.5. Статистические показатели тесноты связи

Показатели тесноты связи дают возможность охарактеризовать зависимость вариации результативного признака от изменения факторного признака. В определенной степени они дополняют и развивают отмеченные приемы обнаружения связи.

Зная показатели тесноты корреляционной связи, можно решить следующие вопросы:

– необходимость изучения данной группы связи между признаками и целесообразности ее практического применения;

– определить, сопоставляя показатели тесноты связи для различных ситуаций о степени ее проявления в конкретных условиях;

– сопоставляя показатели тесноты связи результативного признака с различными факторными признаками, выявить те основные из них, от которых зависит формирование результативного признака.

В целом качество корреляционного анализа зависит: от однородности единиц изучаемой совокупности; достоверности числа наблюдений, чтобы получить достоверные данные; из множества факторных признаков необходимо отобрать основные, от которых зависит значение результативного признака, при этом отобранные факторные признаки должны быть независимыми друг от друга, иначе они будут дублировать друг друга. Факторные признаки, которые используются для корреляционного анализа, должны иметь фиксированную количественную характеристику. Наконец, установление формы и направление связи, а также определение тесноты связи желательно провести на примере изучения связей двух признаков, т. е. парной зависимости.

Коэффициент Фехнера

К наиболее простым показателям степени тесноты связи относят коэффициент корреляции знаков, который был предложен немецким ученым К. Фехнером. Этот показатель основан на оценке степени согласованности направлений отклонений индивидуальных значений факторного и результативного признаков от соответствующих средних. Для его расчета вычисляют средние значения результативного и факторного признаков, а затем проставляют знаки отклонений для всех значений взаимосвязанных пар признаков.

Коэффициент Фехнера находят по следующей формуле:

$$K_F = \frac{n_a - n_b}{n_a + n_b},$$

где n_a – число совпадений знаков отклонений индивидуальных величин от средней n_b – число несовпадений знаков отклонений индивидуальных величин от средней.

Коэффициент Фехнера может принимать различные значения в пределах от -1 до+1. Если знаки всех отклонений совпадут, то $n_b = 0$ и тогда $K_F = 1$, что свидетельствует о наличии прямой связи. Если же знаки всех отклонений будут разными, тогда $n_a = 0$ и $K_F = -1$, что дает основание предположить наличие обратной связи.

Пример для расчета коэффициента Фехнера (таблица 4):

Таблица 4.4. – Расчет коофециента Фехнера

Энерговооруженность труда, (кват-час/чел.)	Выработка рабочих за м-ц, (тыс. сом/чел)	Знаки отклонений по (x)	Знаки отклонений по (y)	Совпадение – а Несовпадение – в
4,0	8,43	-	-	а
4,3	9,79	-	-	а
6,7	9,06	-	-	а
7,4	11,01	-	-	а
7,7	12,69	-	+	в
8,3	10,55	-	-	а
9,6	10,12	-	-	в
12,1	14,58	+	+	а
15,0	14,18	+	+	а
16,0	20,22	+	+	а
91,1	121,63			

$$\overline{x} = \frac{91{,}1}{10} = 9{,}11;\ \overline{y} = \frac{121{,}63}{10} = 12{,}16.\ n_a = 8; n_d = 2.\ K_F = \frac{n_a - n_d}{n_a + n_d} = \frac{8-2}{10} = 0{,}6.$$

Коэффициент Фехнера, рассчитанный на основе использования данных таблицы 4.4. свидетельствует, что имеется прямая зависимость между анализируемыми признаками. Кроме того, коэффициент Фехнера не зависит от величины отклонений факторного и результативного признаков от соответствующей средней величины. Поэтому данный коэффициент дает только приближенную характеристику о степени тесноты корреляционной связи между изучаемыми признаками.

4.6. Линейный коэффициент корреляции

Более достоверным показателем степени тесноты связи является парный линейный коэффициент корреляции (r), предложенный английским ученым К. Пирсоном.

При расчете этого показателя учитываются не только знаки отклонений индивидуальных значений признака от средней, но и сама величина таких отклонений, т. е. соответственно для факторного и результативного признаков величины $(x_i - \overline{x})$, $(y_i - \overline{y})$. Однако непосредственно сопоставлять между собой полученные абсолютные величины нельзя, так как сами признаки могут быть выражены в разных единицах, а при наличии одних и тех же единиц измерения средние могут быть различными по величине. В этой связи сравнению могут подлежать отклонения, выраженные в относительных величинах, т. е. в долях среднего квадратического отклонения, которые называют нормированными отклонениями. Так, для факторного признака нормированное отклонение выражается: $t_{x_i} = \dfrac{x_i - \overline{x}}{\sigma_x}$, а для результативного признака: $t_{y_i} = \dfrac{y_i - \overline{y}}{\sigma_y}$.

Полученные нормированные отклонения можно сравнивать между собой. Для того чтобы на основе сопоставления рассчитанных нормированных отклонений получить обобщающие характеристику степени тесноты связи между признаками для всей совокупности, рассчитывают среднее произведение нормированных отклонений. Полученная таким образом нормированная средняя называется линейным коэффициентом корреляции (r).

$$r = \frac{\sum_1^n t_{x_i} t_{y_i}}{n} = \frac{\sum_1^n \left(\dfrac{x_i - \overline{x}}{\sigma_x}\right) \cdot \left(\dfrac{y_i - \overline{y}}{\sigma_y}\right)}{n},$$

поскольку σ_x, σ_y для данных рядов являются постоянными, то их можно вынести за скобку, тогда формула линейного коэффициента корреляции примет следующий вид:

$$r = \frac{\sum_1^n (x_i - \overline{x}) \cdot (y_i - \overline{y})}{n \sigma_x \sigma_y}.$$

Пример для расчета линейного коэффициента корреляции (таблица 4.5.):

Таблица 4.5 – Расчет линейного коэффициента корреляции

Энерговооружен. труда, кват-час/чел.	Выработка рабочих тыс. сом/чел	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$	$(y_i = \overline{y})$	$(y_i - \overline{y})^2$	$(x_i - \overline{x})(y_i - \overline{y})$
4,0	8,43	–5,11	26,11	–3,73	13,94	19,08
4,3	9,79	–4,81	23,14	–2,37	5,63	11,41
6,7	9,06	–2,41	5,80	–3,10	9,63	7,48
7,4	11,01	–1,71	2,92	–1,15	1,33	1,97
7,7	11,69	–1,41	1,99	–0,47	0,22	0,66
8,3	12,55	–0,81	0,66	0,39	0,15	–0,31
9,6	10,12	0.49	0,24	–2,04	4,17	–1,00
12,1	14,58	2,99	8,94	2,42	5,84	7,23
15,0	14,18	5,89	34,69	2,02	4,07	11,88
16,0	20,22	6,89	47,47	8,06	64,91	55,51
Всего: 91,1	121,63		151,97		109,90	113,91

Основные показатели для расчета линейного коэффициента корреляции:

$$\bar{x} = \frac{91{,}1}{10} = 9{,}11; \; \bar{y} = \frac{121{,}6}{10} = 12{,}16; \; \sigma_x = \sqrt{\frac{\sum (x-\bar{x})^2}{n}} = \sqrt{\frac{151{,}9}{10}} \approx 3{,}89;$$

$$\sigma_y = \sqrt{\frac{\sum (y-\bar{y})^2}{n}} = \sqrt{\frac{109{,}90}{10}} \approx 3{,}316;$$

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y} = \frac{113{,}905}{10 \cdot 3{,}89 \cdot 3{,}316} = 0{,}881.$$

Полученный линейный коэффициент корреляции, где $r = 0{,}881$ характеризует о достаточно тесной зависимости между рассматриваемыми признаками. Для подтверждения данной оценки можно привлечь другой показатель, который называется коэффициентом детерминации и его находят при возведении в квадрат коэффициента корреляции, т.е. (r^2).

Коэффициент детерминации применительно к данному примеру равен:

$$r^2 = 0{,}881^2 = 0{,}7762 \approx 77{,}62\,\%.$$

Это означает, что на 77,62 % выработки продукции рабочих зависит от энерговооруженности их труда.

Линейный коэффициент корреляции может принимать любые значения в пределах от –1 до +1. Чем ближе коэффициент корреляции по абсолютной величине к 1, тем теснее связь между изучаемыми признаками. Знак при линейном коэффициенте корреляции указывает на направление связи – прямой зависимости соответствует знак плюс, а обратной зависимости – знак минус.

4.7. Корреляционное отношение

Линейный коэффициент корреляции дает достоверную оценку степени тесноты связи при линейной зависимости между признаками. При наличии криволинейной зависимости более достоверную оценку дает применение эмпирического корреляционного отношения (η).

Расчет корреляционного отношения основан на использовании теоремы сложении дисперсии. Как известно общая дисперсия результативного признака (σ_0^2) может быть разложена на две составляющие части. Первая часть – межгрупповая дисперсия (δ^2) характеризует ту часть вариации результативного признака, которая складывается под влиянием изменения признака – фактора, положенного в основу группировки.

$$\delta^2 = \frac{\sum_{j=1}^{k} (\bar{y}_j - \bar{y}_0)^2}{\sum_{j=1}^{k} n_j},$$

где $\bar{y}_j$ – средние значения результативного признака в выделенных группах; $\bar{y}_0$ – общая средняя для всей совокупности;

n_j – число наблюдений в соответствующей группе;

k – число выделенных групп.

Вторая составная часть – средняя из внутригрупповых дисперсий ($\overline{\sigma^2}$) оценивает ту часть вариации результативного признака, которая обусловлена действием других случайных причин.

$$\overline{\sigma^2} = \frac{\sum_{1}^{k} \sigma_i^2 n_j}{\sum_{1}^{k} n_j},$$

где σ_j^2 – дисперсия результативного признака в соответствующей группе.

Общая дисперсия равна: $\sigma_0^2 = \delta^2 + \overline{\sigma^2}$. Зная общую и межгрупповую дисперсии, можно оценить ту долю, которую составляет вариация под действием фактора (x) в общей вариации результативного признака (y), т. е. найти отношение:

$$\eta = \sqrt{\frac{\delta}{\sigma_0^2}}.$$

Величина корреляционного отношения может быть рассчитана и по следующей формуле:

$$\eta = \sqrt{1 - \frac{\overline{\sigma^2}}{\sigma_0^2}}.$$

Величина корреляционного отношения будет равна нулю, если нет колеблемости в величине средних по выделенным группам. В тех случаях, когда внутригрупповая дисперсия близка к нулю, т. е. практически вся вариация результативного признака обусловлена действием фактора (x), величина корреляционного отношения близка к 1.

Пример для расчета корреляционного отношения:

дана корреляционная таблица двух взаимозависимых рядов, в которых отражены три значения факторного признака (x) – количество внесенных удобрений и значение результативного признака (y) – урожайность (таблица 4.6.):

Таблица 4.6. – Корреляционная таблица взаимозависимых рядов

y \ x	20,0	30,0	40,0	Итого
0,8	1	0		1
0,9	2	1		3
1,0	3	3		6
1,1	1	4	1	6
1,2			2	2
1,3			2	2
Итого	7	8	5	20

Каждая группа участков с разной урожайностью имела разное количество внесенных удобрений. Так, при внесении удобрений по 20 т, урожайность на разных участках была разной: на одном участке она составила 0,8 т, на двух участках – 0,9 т, на трех – 1,0 т и на одном – 1,1 т. Используя эти данные, средняя урожайность и дисперсия для этой группы участков:

$$\overline{y}_1 = \frac{0,8\cdot 1 + 0,9\cdot 2 + 1,0\cdot 3 + 1,1\cdot 1}{7} = \frac{6,7}{7} = 0,957 \text{ т.};$$

$$\sigma_1^2 = \frac{(0,8-0,957)^2\cdot 1 + (0,9-0,957)^2\cdot 2 + (1,0-0,957)^2\cdot 3 + (1,1-0,957)^2\cdot 1}{7} =$$

$$= \frac{0,057143}{7} = 0,008163$$

Для группы участков с количеством внесенных удобрений 30,0 т. среднюю урожайность и дисперсию, произведя соответствующие вычисления, как и в первой группе, определим:

– среднюю урожайность для второй группы $\overline{y}_2 = 1,037$ т.

– дисперсию для второй группы $\sigma_2^2 = 0,004844$.

Вычислим аналогичные характеристики для группы участков внесенных удобрений по 40 т.:

– средняя урожайность $\overline{y}_3 = 1,2$ т.;

– дисперсия $\sigma_3^2 = 0,0056$.

Используя эти данные, определим средний урожай всех 20 участков, т.е. общую среднюю:

$$\overline{y}_0 = \frac{0,8\cdot 1 + 0,9\cdot 3 + 1,0\cdot 6 + 1,1\cdot 6 + 1,2\cdot 2 + 1,3\cdot 2}{20} = \frac{21,1}{20} = 1,055 \text{ т};$$

и степень колеблемости (дисперсию) средней урожайности групп в сопоставлении с общей средней урожайностью, который называется межгрупповой дисперсией (δ^2).

$$\overline{\delta^2} = \frac{\sum_j^k (\overline{y}_j - \overline{y}_0)^2 n_j}{\sum_j^k n_j} = \frac{(0,957-1,055)^2\cdot 7 + (1,037-1,055)^2\cdot 8 + (1,220-1,055)^2\cdot 5}{20} = 0,010247.$$

Межгрупповая дисперсия показывает вариацию, возникающую за счет воздействия факторного признака. В данном случае $\overline{\delta^2} = 0,010247$ – это показатель вариации урожайности, возникающего за счет разности в количестве внесенных удобрений.

Кроме межгрупповой дисперсии, рассчитывается и средняя из частных, групповых дисперсий, которая характеризует степень колеблемости результативного признака за счет влияния всех прочих неучтенных факторов, т. е. в данном случае неучтенные кроме удобрений факторы, воздействующие на урожайность. Средняя из групповых (частных) дисперсий $(\overline{\sigma^2})$ рассчитывается по формуле:

$$\overline{\sigma^2} = \frac{\sum_j^k \sigma_j^2 n_j}{\sum_j^k n_j} = \frac{0,008163\cdot 7 + 0,00484\cdot 8 + 0,005200\cdot 5}{20} = 0,006094.$$

Исходя из того, что общая дисперсия определяется как сумма межгрупповой дисперсии и средней из групповых дисперсий, то это соотношение выражается в виде следующей формулы:

$$\sigma_0^2 = \overline{\delta^2} + \overline{\sigma^2} = 0{,}010247 + 0{,}006094 = 0{,}016341 .$$

Полученный расчетным путем показатель общей дисперсии характеризует степень вариации урожайности по всей совокупности, т. е. 20 участкам в целом.

Расчет общей дисперсии, внутригрупповой и межгрупповой дисперсий дает возможность оценить степень влияния факторного признака на колеблемость результативного признака на основе использования корреляционного отношения (η):

$$\eta = \sqrt{\frac{\overline{\delta^2}}{\sigma_0^2}} = \sqrt{\frac{0{,}010247}{0{,}016347}} = \sqrt{0{,}61} = 0{,}78.$$

Данное корреляционное отношение η = 0,78 = 78 % показывает, что колеблемлость урожайности на обследованных 20 участках на 78 % зависит от количества внесенных удобрений.

4.8. Непараметрические методы измерения связи

Непараметрические методы измерения связи широко применяются при характеристике взаимосвязи качественных признаков или же количественных признаков, которые не поддаются определенным законам распределения числовых показателей.

К непараметрическим методам измерения тесноты связи между изучаемыми признаками относятся:

– коэффициенты ассоциации и контингенции;
– коэффициенты взаимной сопряженности;
– коэффициенты корреляции рангов.

Коэффициенты ассоциации и контингенции

а) Коэффициент ассоциации как мера тесноты связи применяется при изучении двух качественных признаков, состоящих только из двух групп. Для его вычисления строится четырехклеточная таблица, которая выражает связь между двумя признаками, каждая из них в свою очередь должна быть альтернативной,

т. е. состоящей только из двух видов, качественно отличных друг от друга. Например, при изучении зависимости урожая от количества внесенных в почву удобрений выделяем по урожайности и по количеству внесенных удобрений лишь по две группы. При этом условии можно построить следующую четырехклеточную таблицу (таблица 4.7).

Таблица 4.7.

Удобрено / Урожайность	Хорошо	Плохо	Всего
Высокая	50 (a)	10 (b)	60 (a + b)
Низкая	10 (c)	30 (d)	40 (c + d)
Всего	60 (a + c)	40 (d + d)	

Числа, стоящие на пересечении строк и граф – a, b, c, d , показывают, сколько участков встречается с тем и другим количеством удобрений, внесенным в почву, с той и другой урожайностью.

Мера тесноты связи – коэффициент ассоциации, предложенный Юлом, определяется по формуле:

$$K_{as} = \frac{ad - bc}{ad + bc} = \frac{50 \cdot 30 - 10 \cdot 10}{50 \cdot 30 + 10 \cdot 10} = 0{,}865.$$

Рассчитанный коэффициент ассоциации показывает довольно тесную связь между внесенным удобрением и урожайностью.

б) Коэффициент контингенции. Четырехклеточная таблица позволяет найти другой показатель, называемый коэффициентом контингенции, разработанный Пирсоном, который рассчитывается по формуле:

$$K_{kon} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} = \frac{50 \cdot 30 - 10 \cdot 10}{\sqrt{60 \cdot 40 \cdot 60 \cdot 40}} = 0{,}583.$$

Коэффициент контингенции оценивает тесноту связи между урожайностью и уровнем внесенных удобрений сравнительно ниже, чем коэффициент ассоциации.

Коэффициент взаимной сопряженности является мерой тесноты связи для качественных признаков, каждый из которых состоит более чем из двух групп. Для расчета коэффициентов взаимной сопряженности применяются формулы, разработанные К. Пирсоном и А. Чупровым.

Коэффициент взаимной сопряженности Пирсона определяется по формуле:

$$C = \sqrt{\frac{\varphi^2}{1+\varphi^2}},$$

где φ^2 – показатель взаимной сопряженности, который определяется суммой отношений квадратов частот каждой клетки таблицы, характеризующий связь качественных признаков к произведению частот итоговых показателей соответствующего столбца и строки. Вычитая из этой суммы единицу, получим φ^2, схема расчета которой показаны в следующей таблице:

Таблица 4.8.

Признак В / Признак А	В–1	В–2	В–3	Итого
А-1	f_1	f_2	f_3	F_1
А-2	f_4	f_5	f_6	F_2
А-3	f_7	f_8	f_9	F_3
Итого	N_1	N_2	N_3	

Для расчета коэффициента сопряженности необходимо знать φ^2, который представляет собой сумму квадратов частот каждой строки, деленных на сумму частот по колонам и, в свою очередь, на сумму частот по строке без единицы.

Расчет φ^2 в символах, взятых из приведенной схемы, можно представить в следующем виде:

по 1–ой строке $\left(\frac{f_1}{N_1}+\frac{f_2}{N_2}+\frac{f_3}{N_3}\right):F_1 = P_1;$

по 2–ой строке $\left(\frac{f_4}{N_1}+\frac{f_5}{N_2}+\frac{f_6}{N_3}\right):F_2 = P_2;$

по 3–ей строке $\left(\frac{f_7}{N_1}+\frac{f_8}{N_2}+\frac{f_9}{N_3}\right):F_3 = P_3.$

Тогда $\varphi^2 = P_1 + P_2 + P_3 = \sum P - 1$.

Примерно по той же схеме исчисляется коэффициент взаимной сопряженности Чупрова:

$$K = \sqrt{\frac{\phi^2}{(k_1-1)(k_2-1)}},$$

где φ^2 имеет одинаковое значение с φ^2 Пирсона и является показателем взаимной сопряженности, k_1– число групп по столбцам таблицы, k_2 − число групп по строкам таблицы.

Пример для расчета взаимной сопряженности между урожайностью и поливом посевов (таблица 4.9.):

Таблица 4.9.

Полив / Урожай	Слабый	Средний	Хороший	Всего
Низкий	30 (900) 10,591	42 (1764) 543	2 (4) 0,056	74 (-) 15.790; 0,213
Средний	53 (2809) 33,059	247 (61009) 177,869	48 (2304) 32,00	348 (-) 242,928;0,698
Высокий	2 (4) 0,037	54 (2916) 8,501	22 (484) 6,722	78 (-) 15,269; 0,196
Всего	85	343	72	500; 1,107

В каждой клетке таблицы записаны частоты, их квадраты, квадраты частот, деленные на сумму частот по столбцу. В итоговых столбцах записаны суммы частот, суммы результатов деления, а также результат деления нижнего числа на верхнее, используя эти данные, определим значение φ = 1,107 – 1,00 = 0,107, а коэффициенты взаимной сопряженности:

а) Пирсона $C = \sqrt{\frac{\varphi^2}{1+\varphi^2}} = \sqrt{\frac{0,107}{1,107}} = 0,311;$

б) Чупрова $K = \sqrt{\frac{\phi^2}{(k_1-1)(k_2-1)}} = \sqrt{\frac{0,107}{2\cdot 2}} = 0,163.$

Полученные коэффициенты сопряженности свидетельствуют о наличии определенной взаимосвязи между урожайностью и поливом посевов.

Коэффициенты корреляции рангов

В основу данного метода положен принцип нумерации значений статистического ряда, когда каждой единице совокупности присваивается порядковый номер в ряду и будет упорядочен по уровню признака, затем ряд значений признака ранжируется, а номер каждой отдельной единицы будет ее рангом.

Можно получить предварительное представление о наличии или отсутствии связи между признаками, если сопоставить последовательность взаимного расположения рангов факторного и результативного признаков. Для этого ранги индивидуальных значений факторного признака располагают в порядке возрастания, и если ранги результативного признака обнаруживают тенденцию к увеличению, то можно предполагать наличие прямой связи; если же с увеличением рангов факторного признака ранги результативного признака уменьшаются, то это свидетельствует о наличии обратной связи между изучаемыми признаками.

Коэффициенты корреляции, основанные на использовании рангов, были предложены К. Спирмэном и М. Кендэлем.

Коэффициент корреляции рангов Спирмэна рассчитывается по формуле: $\rho = 1 - \dfrac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$.

Таблица 4.10.

Порядковый номер кондитерского магазина	Ранг магазинов по оценке экспертов	Ранг магазинов по оценке покупателей	Разница рангов (гр. 2–гр. 3) $\lvert d \rvert$	d_i^2
1	2	3	4	5
1	6	4	2	4
2	3	1	2	4
3	4	7	3	9
4	2	8	6	36
5	7	5	2	4
6	1	3	2	4
7	5	2	3	9
8	8	8	0	0

9	10	9	1	1
10	9	10	1	1
Итого				72

Используя приведенные данные таблицы 4.10., рассчитаем коэффициент корреляции рангов Спирмэна:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6 \cdot 72}{10 \cdot 99} = 1 - 0,436 = 0,564.$$

Коэффициент корреляции рангов, предложенный Кендэлем, рассчитывается по формуле:

$$\tau = \frac{2S}{n(n-1)}, \text{ где } S = P + Q.$$

Для определения τ необходимо упорядочить ряд рангов переменной (х), приведя его к ряду натуральных чисел, затем приводим в соответствие последовательность рангов переменной (у) к переменной (х), результаты которой приводится в таблице 11.

Таблица 4.11.

Ранг экспертов (х)	1	2	3	4	5	6	7	8	9	10
Ранг покупателей (у)	3	6	1	7	2	4	5	8	10	9

Для определения *S* необходимо найти сумму двух слагаемых ($P + Q$). При определении значения *P* необходимо подсчитать сколько чисел, находящихся справа от каждого из элементов последовательности рангов переменной (у), имеет величину ранга, превышающую ранг рассматриваемого элемента. Так, первому значению в последовательности рангов переменной (у), т. е. числу 2 превышают 7 чисел (8,7,4,5,8,9,10); второму значению 6 превышают 4 (7,8,10,9) и т.д. Суммируя полученные в таким способом числа получим значение $P = 35(7+4+7+3+5+4+3+2)$.

Чтобы определить *Q*, необходимо подсчитать, сколько чисел, находящихся справа от каждого из членов последовательности рангов переменной (у), имеет ранг меньше, чем это число. Результаты таких сопоставлений берут со знаком минус. Используя приведенные данные, рассчитаем значение $Q = –10,(–2,–4,0,–3,0,0,0,0,–1)$.

Определив значение $S = P + Q = 35 - 10 = 25$, в целом можно рассчитать коэффициент корреляции рангов Кендэла:

$$\tau = \frac{2S}{n(n-1)} = \frac{2\cdot 25}{10\cdot 9} = 0,555.$$

Коэффициенты корреляции рангов Спирмэна, а также Кендэла показывают сравнительно высокий уровень тесноты связи между изучаемыми признаками.

4.9. Совокупный общий коэффициент корреляции

Показатель тесноты связи, устанавливаемый между результативным и двумя и более факторными признаками, называют совокупным общим коэффициентом корреляции (R).

Этот коэффициент предполагает наличие линейной связи между каждой парой признаков, которая может быть выражена парными коэффициентами корреляции:

$$R = \sqrt{\frac{r_{yx}^2 + r_{yz}^2 - 2r_{yx}r_{yz}r_{xz}}{1 - r_{xz}^2}},$$

где $r_{yx} = \dfrac{\overline{yx} - \overline{y}\cdot\overline{x}}{\sigma_y\sigma_x}$; $r_{yz} = \dfrac{\overline{yz} - \overline{y}\cdot\overline{z}}{\sigma_y\sigma_z}$; $r_{xz} = \dfrac{\overline{xz} - \overline{x}\cdot\overline{z}}{\sigma_x\sigma_z}$.

$$\sigma_y = \sqrt{\frac{\sum(y-\overline{y})^2}{n}};\ \sigma_x = \sqrt{\frac{\sum(x-\overline{x})^2}{n}};\ \sigma_z = \sqrt{\frac{\sum(z-\overline{z})^2}{n}}.$$

Таблица 4.12. – Пример для расчета совокупного общего коэффициента корреляции

№ предприятия	Выпуск продукции, млн сом. (y)	Основные фонды, млн сом. (x)	Числен. рабочих, чел. (z)	y x	y z	x z
1	5,2	3,9	462	20,28	2402,4	1801,8
2	1,5	2,0	120	3	180	240
3	9,4	5,5	581	51,7	5461,4	3195,5
4	4,4	4,9	505	21,56	2222	2474,5
5	5,6	4,5	435	25,2	2436	1957,5
6	1,9	2,2	139	4,18	264,1	305,8
7	5,8	4,0	350	23,2	2030	1400
8	3,9	5,6	450	21,84	1755	2520
9	3,6	3,1	310	11,16	1116	961
10	7,9	4,5	400	35,55	3160	1800
Всего	49,2	40,2	3752	217,67	21026,9	16656,1

(Продолжение таблицы 4.12.)

№ предприятия	$(y-\overline{y})^2$	$(x-\overline{x})^2$	$(z-\overline{z})^2$
1	0,09	0,01	7569
2	11,56	4	65025
3	20,25	2,25	42436
4	0,25	0,81	16900
5	0,49	0,25	3600
6	9	3,24	55696
7	0,81	0	625
8	1	2,56	5625
9	1,69	0,81	4225
10	9	0,25	625
Всего	54,14	14,18	202326

Используя данные таблицы 4.12., рассчитаем промежуточные обобщающие показатели:

1) $\overline{y} = \dfrac{49,2}{10} = 4,9;$

2) $\overline{x} = \dfrac{40,2}{10} = 4,0;$

3) $\overline{z} = \dfrac{3752}{10} = 375,2;$

4) $\overline{yx} = \dfrac{217,67}{10} = 21,77;$

5) $\overline{yz} = \dfrac{21026,5}{10} = 2102,7;$

6) $\overline{xz} = \dfrac{16656,1}{10} = 1665,6;$

7) $\sigma_y = \sqrt{\dfrac{54,14}{10}} = 2,32;$

8) $\sigma_x = \sqrt{\dfrac{14,18}{10}} = 1,19;$

9) $\sigma_z = \sqrt{\dfrac{202326}{10}} = 142,2.$

Расчет парных коэффициентов и совокупного коэффициента корреляции:

$$r_{yx} = \frac{\overline{yx} - \overline{y} \cdot \overline{x}}{\sigma_y \sigma_x} = \frac{21,77 - 4,9 \cdot 4}{2,32 \cdot 1,19} = \frac{2,17}{2,76} = 0,786;$$

$$r_{yz} = \frac{\overline{yz} - \overline{y} \cdot \overline{z}}{\sigma_y \sigma_z} = \frac{2102,7 - 4,9 \cdot 375}{2,32 \cdot 142,2} = \frac{265}{329,9} = 0,804;$$

$$r_{xz} = \frac{\overline{xz} - \overline{x} \cdot \overline{z}}{\sigma_x \sigma_z} = \frac{1665,6 - 4 \cdot 375}{1,19 \cdot 142,2} = \frac{165,6}{169,2} = 0,978.$$

$$R = \sqrt{\frac{r_{yx}^2 + r_{yz}^2 - 2 r_{yx} r_{yz} r_{xz}}{1 - r_{xz}^2}} = \sqrt{\frac{0,786^2 + 0,804^2 - 2 \cdot 0,786 \cdot 0,804 \cdot 0,978}{1 - 0,978^2}} = \sqrt{\frac{0,028}{0,044}} = 0,797.$$

Совокупный общий коэффициент корреляции, характеризующий влияние двух факторных признаков на результативный признак, показывает высокий уровень тесноты связи между ними, т. е. R = 0,797.

Контрольные вопросы, определение понятий и показателей

1. Основные формы и виды связи.
2. Как строится уравнение регрессии?
3. Составьте уравнение прямой и кривой с нахождением их параметров
4. Определение коэффициентов корреляции (парный линейный коэффициент, корреляционное отношение, совокупный коэффициент).
5. Непараметрические методы измерение связи.

5. СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ АНАЛИЗА ДИНАМИЧЕСКИХ РЯДОВ

5.1. Общее понятие о динамических рядах

Динамический ряд представляет собой ряд числовых показателей, расположенных в хронологической последовательности, которые отражают развитие, изменение изучаемых явлений и процессов во времени, и в пространстве. Каждый динамический ряд состоит из периода или момента времени, к которым относятся приводимые статистические данные, кроме того, приводится и те статистические показатели, которые характеризуют анализируемые явления и процессы по состоянию на определенный период или же момент времени.

Статистические показатели, характеризующие изучаемые объекты, являются уровнями данного ряда динамики. Вид ряда динамики зависит не только от характера показателей, оценивающих изучаемый объект, но и от того, за какой период, или же по состоянию на определенный момент времени. Статистические показатели, приводимые в динамическом ряду, могут быть абсолютными, относительными или же средними величинами.

Различают моментные и интервальные виды рядов динамики. Моментным является ряд динамики, уровни которого характеризуют изучаемое явление по состоянию на конкретный момент времени. Такие ряды используют, например, при описании товарных остатков в торговых точках или же остатки средств на счетах клиентов в коммерческих банках и т.д. Интервальным является ряд динамики, уровни которого характеризуют накопленный, суммарный результат изменения явлений за определенные промежутки (интервалы, периоды) времени. Интервальные ряды динамики используют для описания величин в виде экономического потока, операций (операции полученные, проценты уплаченные, доходы и расходы, текущие затраты, выпуск и реализация продукции и т.п.).

Уровни интервальных рядов динамики обладают свойством суммарности, показатели моментных рядов такого свойства не имеют. Так, можно сложить показатели объема промышленной продукции за месяцы, кварталы и получить итог производства за год, если же за год сложить численность рабочих на начало каждого месяца или же квартала, то полученная сумма не будет иметь реального смысла

При построении и анализе динамических рядов необходимо учитывать требование сопоставимости данных. Сопоставимость уровней динамического ряда производится по кругу охватываемых объектов, по территории и по методологии расчета показателей.

Несопоставимость по кругу охватываемых объектов возникает в результате особенности учета изучаемых единиц совокупности или же в результате изменений в классификации единиц совокупности и т.д. Так, данные об объеме промышленной продукции могут быть представлены по крупным и средним предприятиям или же по всем предприятиям, включая малые предприятия. Для построения сопоставимого динамического ряда в качестве источника информации необходимо использовать соответствующие показатели объема промышленной продукции, охватывающие весь круг предприятий, по которым имеется статистическая отчетность.

Несопоставимость статистических данных по территории возникает в результате административно-территориальных изменений. В этом случае для обеспечения сопоставимости показателей проводят прямой пересчет данных по исходному, первичному материалу.

К несопоставимости уровней динамического ряда приводят различия в методике их расчета. Для устранения их производят пересчет предшествующих данных в соответствии с новой методикой.

Несопоставимость показателей, возникающая в силу неодинаковости применяемых единиц измерения, сама по себе очевидна. С различием применяемых единиц измерения приходится встречаться при учете продукции в натуральном выражении. Поэтому приведение к сопоставимому виду разнообразной продукции основывается на ее выражении в ценностных или трудовых измерителях. При анализе показателей объема продукции, измеренных в ценностных единицах, следует учитывать, что существует несколько видов цен: цены производителей и цены потребителей. Поэтому при характеристике стоимостных показателей объема продукции во времени должно быть устранено влияние изменения цен. На практике для решения этой задачи количество продукции, произведенное в разные периоды, оценивают в ценах одного периода, которые называют фиксированными или сопоставимыми.

5.2. Показатели анализа динамических рядов и методы их определения

Динамический ряд представляет собой ряд последовательных уровней, сопоставляя которые можно охарактеризовать скорость и интенсивность развития изучаемого явления. В результате сравнения уровней получается система абсолютных и относительных показателей динамических рядов, к числу которых относятся абсолютный прирост, коэффициент роста, темп прироста, абсолютное значение одного процента прироста. Если сравнению подлежат несколько последовательных уровней, то возможны два варианта сопоставления:

1. Если каждый уровень динамического ряда сравнивается с одним и тем же предшествующим уровнем, принятым за базу сравнения, такое сравнение называют сравнением с постоянной базой. В качестве базисного уровня выбирается либо начальный уровень, или же уровень, с которого начинается определенный этап развития изучаемого явления. Такое сравнение динамического ряда называется с постоянной базой сравнения.

2. Если каждый уровень динамического ряда сравнивается с предшествующим ему уровнем, такое сравнение называют сравнение с переменной базой.

Показатели динамических рядов с постоянной базой сравнения (базисные показатели) характеризуют конечный результат всех изменений в уровнях ряда динамики от периода, к которому относится базисный уровень, до данного (i- того) периода.

Показатели динамических рядов с переменной базой сравнения (цепные показатели) характеризуют интенсивность изменения уровня от периода к периоду (или от даты к дате) в пределах изучаемого промежутка времени.

Абсолютный прирост (Δ_i) определяется, как разность между двумя уровнями динамического ряда и показывает, насколько данный уровень ряда превышает уровень, принятый за базу сравнения:

$$\Delta_i = y_i - y_0, \qquad (1)$$

где Δ_i – абсолютный прирост,

y_i – уровень сравниваемого периода,

y_0 – уровень базисного периода.

При сравнении с переменной базой сравнения абсолютный прирост равен:

$$\Delta_i = y_i - y_{i-1}, \qquad (1.\ а)$$

где y_{i-1} – уровень предшествующего периода.

Коэффициент роста определяется, как отношение двух сравниваемых уровней и показывает, во сколько раз данный уровень превышает уровень базисного периода.

При сравнении с постоянной базой: $K_i = \dfrac{y_i}{y_0}$. (2)

При сравнении с переменной базой: $K_i = \dfrac{y_i}{y_0}$. (2.а)

Если коэффициенты роста выражают в процентах, то их называют темпами роста: $K_p = K_i \cdot 100$ %.

Коэффициент прироста показывает, насколько процентов уровень данного периода больше (или меньше) базисного уровня. Этот показатель рассчитывается в двух вариантах:

1) как отношение абсолютного прироста к уровню, принятому за базу сравнения:

$K_{пр} = \dfrac{y_i - y_0}{y_0} \cdot 100$ %;

или $K_{пр} = \dfrac{y_i - y_{i-1}}{y_{i-1}}$.

2)как разность между темпом роста (в процентах) и 100 %:

$$K_{пр} = K_p - 100\ \%.$$

Между показателями анализа динамических рядов с постоянной и переменной базой сравнения имеется определенная связь. Так, сумма абсолютных приростов с переменной базой дает общий прирост за изучаемый период:

$$\sum_{i=1}^{n-1} \Delta_i = y_n - y_1 = \Delta_{n/1},$$

где n – число уровней динамического ряда.

Также моет быть произведен переход от коэффициентов роста с постоянной базой к коэффициентам роста, вычисленным с переменной базой. Так, если известны коэффициенты роста с постоянной базой, то частные от последовательного деления этих коэффициентов равны соответствующим коэффициентам

роста с переменной базой: $K_{i/i-1} = \dfrac{K_{i/1}}{K_{i-1/1}}$.

При наличии коэффициентов роста с переменной базой соответствующий базисный коэффициент роста определяется перемножением цепных коэффициентов роста:

$$K_{i/1} = K_{2/1} \cdot K_{3/2} \cdot \dots \cdot K_{i-1/i-2} K_{i/i-1}$$

При анализе относительных показателей динамики (темпов роста и темпов прироста) не следуют рассматривать изолированно от их абсолютных показателей (уровней ряда и абсолютных приростов.

Сравнение абсолютного прироста и темпа прироста за одни и те же периоды времени показывают, что замедление темпов прироста не всегда сопровождается уменьшением абсолютных приростов. Поэтому их следует рассматривать в сопоставлении с показателем абсолютного прироста и этот показатель называется *абсолютным значением одного процента прироста* (A_i):

$A_i = \dfrac{y_i - y_{i-1}}{T_{прi/i-1}}$, т. е. этот показатель рассчитывают, как отношение абсолютного прироста (в процентах) за тот же период времени.

Если преобразовать эту формулу, то получим следующее выражение:

$$A_i = \frac{y_i - y_{i-1}}{\dfrac{y_i - y_{i-1}}{y_{i-1}} \cdot 100\ \%} = \frac{y_{i-1}}{100} = 0{,}01 y_{i-1}.$$

При сопоставлении динамики развития двух явлений можно использовать показатели, представляющие собой отношения темпов роста или темпов прироста за одинаковые отрезки времени по двум динамическим рядам. Эти показатели называют *коэффициентами* опережения:

$$K_{оп} = \frac{K_p^{i-1}}{K_р^i} \text{ или } K_{оп} = \frac{K_{пр}^{i-1}}{K_{пр}^i},$$

где K_p^{i-1}, K_p^i, $K_{пр}^{i-1}$, $K_{пр}^i$ соответственно темпы роста и темпы прироста сравниваемых динамических рядов.

С помощью этих коэффициентов могут сравниваться динамические ряды одинакового содержания, но относящиеся к разным территориям (странам, регионам, районам) или к различным организациям, предприятиям, министерствам и т.д.

5.3. Средние характеристики показателей ряда динамики

Для обобщающей характеристики изменения изучаемого явления за ряд периодов определяют средние показатели для интервального ряда динамики, моментного ряда динамики, а также для каждого вида показателей анализа рядов динамики.

Для обобщенной характеристики изменения интервального ряда динамики применяется средняя арифметическая простая:

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n},$$

где y_i – уровни интервального ряда динамики;

n – число уровней интервального ряда динамики.

Для обобщенной характеристики изменения моментного ряда динамики учитываются особенности моментного ряда динамики, когда уровни их характеризуют состояние изучаемого явления или же объекта на определенный момент времени, как например, данные на начало и конец периода(месяца)необходимо обобщить в виде средней величины за период (месяц). Для обобщенной характеристики изучаемого явления за продолжительный период времени, как год или полугодие, необходимо из этих месячных средних определить обобщенную среднюю за год. Эта обобщенная средняя величина для характеристики изменения моментного ряда динамики называют средней хронологической и находят по следующей формуле:

$$\overline{y}_{хр} = \frac{\frac{y_1}{2} + y_2 + y_3 + \ldots + y_{n-1} + \frac{y_n}{2}}{n-1},$$

где n – количество моментного ряда динамики;

$y_1, y_2, \ldots, y_n$ – уровни моментного ряда динамики.

Средний абсолютный прирост рассчитывается как средняя арифметическая из показателей абсолютного роста за отдельные промежутки времени по следующей формуле:

$$\overline{\Delta} = \frac{\sum_{i=1}^{n-1} \Delta_i}{n-1},$$

где n – число уровней ряда;

Δ_i – абсолютные изменения по сравнению с предшествующим уровнем.

Средний коэффициент роста вычисляется по формуле средней геометрической из показателей коэффициентов роста за отдельные периоды:

$$\overline{K}_p = \sqrt[n-1]{K_1 \cdot K_2 \cdot \ldots \cdot K_{n-1}},$$

где $K_1, K_2, \ldots, K_{n-1}$ – коэффициенты роста по сравнению с уровнем предшествующего периода;

n – число уровней ряда.

Используя взаимосвязь между коэффициентами роста, вычисленными с переменной базой и коэффициентами роста, рассчитанными с постоянной базой и учитывая, что $K_1 \cdot K_2 \cdot \ldots \cdot k_{n-1} = \frac{y_n}{y_1}$, средний коэффициент роста можно определить по формуле:

$$\overline{K}_p = \sqrt[n-1]{\frac{y_n}{y_1}}.$$

Средний коэффициент прироста можно определить, используя данные коэффициентов роста за определенный период по следующей формуле:

$$\overline{K}_{пр} = \overline{K} - 1,$$

где $\overline{K}_{пр}$ – средний коэффициент прироста,
$\overline{K}_{р}$ – средний коэффициент роста.

Таблица 5.1 – Пример для расчета показателей анализа динамических рядов и их средние показатели (данные условные)

Годы	Произ. прод., тыс.сом	Δ_i к 2008г	Δ_i к пр. уровню	$K_р$ к 2008г	$K_р$ к пр.ур.	$K_{пр}$ к 2008г	$K_{пр}$ к пр.ур.	A_i
2008	120	-	-	-	-	-	-	-
2009	135	15	15	1,125	1,125	0,13	0,13	1,2
2010	154	34	19	1,283	1,140	0,28	0,14	1,35
2011	169	49	15	1,41	1,097	0,41	0,10	1,54
2012	187	67	18	1,59	1,106	0,56	0,11	1,69
Средние показатели	153	41,3	16,8	1,33	1,12	0,33	0,12	1,44

5.4. Статистические методы определения основной тенденции развития динамических рядов

При определении основной тенденции изменения динамических рядов за относительно продолжительный период времени применяются следующие статистические методы: метод укрупнения интервалов, метод скользящей подвижной средней и метод аналитического выравнивания рядов динамики.

Если имеются фактические помесячные данные изучаемой совокупности в течение года или же за продолжительный период времени, используя метод укрупнения интервалов можно определить основную тенденцию изменения изучаемой совокупности.

Для этого ежемесячные данные следует укрупнить поквартально, суммируя их значения за три месяца и разделив их на три, т. е. определить их среднее значение по кварталам. При укрупнении интервала и нахождении их среднего значения в определенной степени элиминируется (устраняется) влияние случайных факторов и выявляется основное направление и тенденция изменения изучаемой совокупности.

Применение метода скользящей подвижной средней предполагает нахождение средней величины, где первые члены единиц совокупности, принимающие участие в нахождении средней в последующем исключаются, а члены, которые еще не принимали участия, включаются в число единиц, из которых определяется подвижная средняя. Так, если мы имеем фактические данные за 12 месяцев, то скользящая подвижная средняя определяется в следующем порядке: из фактических ежемесячных данных находят трехчленную подвижную среднюю, т. е. из y_1, y_2, y_3, ..., y_{12} в следующей последовательности:

$$\overline{y}_1 = \frac{y_1 + y_2 + y_3}{3};\ \overline{y}_2 = \frac{y_2 + y_3 + y_4}{3};\ \overline{y}_3 = \frac{y_3 + y_4 + y_5}{3};\ \overline{y}_{10} = \frac{y_{10} + y_{11} + y_{12}}{3}.$$

При определении значение подвижной средней сокращается на определенное их число единиц, т. е. в данном случае число общей подвижной средней сокращается на две единицы, которые необходимо найти методом экстраполяции недостающих членов данного динамического ряда.

Для определения основной тенденции развития динамических рядов часто используют метод аналитического выравнивания с использованием уравнения прямой, параболы.

Уравнение прямой выражается в виде следующего соотношения:

$$\overline{Y}_x = a_0 + a_1 x,$$

где a_0, a_1 находят по способу наименьших квадратов и используя систему нормальных уравнений:

$$na_0 + a_1 \sum x = \sum y$$

$$a_0 \sum x + a_1 \sum x^2 = \sum yx$$

При определении основной тенденции изменения динамических рядов в систему уравнений и в уравнение прямой вносят некоторое изменение, связанное с заменой значений x на t, который означает подобранные отрезки времени, где t подбирается таким образом, чтобы $\sum t = 0$.

Тогда получим измененный вид уравнений;

$\overline{Y}_t = a_0 + a_1 t$ и соответственно:

$$na_0 + a_1 \sum t = \sum y$$
$$a_0 \sum t + a_1 \sum t^2 = \sum yt$$

Учитывая, что $\sum t = 0$, то получим упрощенную систему уравнений: $na_0 = \sum y$.

$$a_1 \sum t^2 = \sum yt.$$

Из этой системы уравнений легко можно найти параметры a_0, a_1:

$$a_0 = \frac{\sum y}{n}; a_1 = \frac{\sum yt}{\sum t^2}.$$

Подставив в уравнение $\overline{Y}_t = a_0 + a_1 t$, конкретные числовые значения a_0, a_1 можно вычислить их теоретическое распределение, характеризующее основное направление изменения динамических рядов (таблица 5.2.).

Графическое изображение фактической помесячной численности рабочих и теоретически вычисленные с применением метода аналитического выравнивания основной тенденции изменения численности рабочих представлено на рисунке 4.

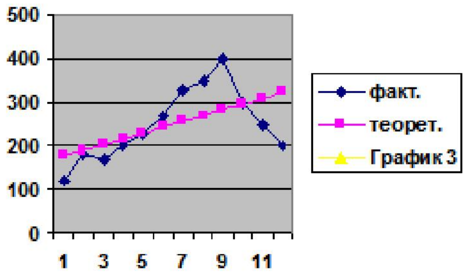

Рисунок 4. – Графическое изображение помесячной численности рабочих

Таблица 5.2. – Пример для определения основной тенденции изменения динамических рядов методом скользящей подвижной средней, аналитического выравнивания и индекса сезонности:

Месяц	Кол-во рабочих, чел ($y_ф$)	3-х член. подвижная сред. ($\overline{y}_x$)	Индекс сезонности $I = \frac{y_ф}{\overline{y}_x}$	t	t^2	yt	$\overline{y}_t$	Индекс сезонности $I = \frac{y_ф}{\overline{y}_t}$
1	120	–	-	–11	121	–1320	176,1	68,14
2	180	156,7	114,8	–9	81	–1620	189,6	94,94
3	170	183,3	92,7	–7	49	–1190	203,0	83,74
4	200	200,0	100	–5	25	–1000	216,5	92,38
5	23	233,3	98,6	–3	9	–690	230,0	100,0
6	270	276,7	97,6	–1	1	–270	243,3	110,9
7	330	316,7	104,2	1	1	330	256,7	128,5
8	350	360	97,2	3	9	1050	270,0	129,6
9	400	350	114,3	5	25	2000	283,6	141,0
10	300	316,7	94,7	7	49	2100	297,0	101,0
11	250	250	100	9	81	2250	310,4	80,54
12	200	–	-	11	121	2200	323,8	61,76
Итого	3000				572	–6090 +9930 3840	3000	

Контрольные вопросы, определение понятий и показателей

1. Общая характеристика и определение видов ряда динамики.
2. Показатели анализа динамических рядов и методы их определения.
3. Средние характеристики показателей ряда динамики.
4. Методы определения основной тенденции развития динамических рядов.

6. СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ ИНДЕКСОВ

6.1. Общая характеристика индексных показателей

Индексы, наряду со средними величинами, являются наиболее распространенными статистическими показателями. С помощью индексных показателей характеризуются развитие национальной экономики в целом и ее отдельных отраслей, анализируются результаты финансово-хозяйственной деятельности предприятий, фирм, учреждений и организаций. Индексы также используется для определения уровня жизни населения и мониторинга деловой активности в экономике.

Индекс представляет собой относительную величину, получаемую в результате сопоставления уровней сложных социально-экономических показателей во времени и пространстве или с запланированным уровнем.

Индексы характеризуют явления, состоящие из разнородных элементов, которых нельзя непосредственно суммировать в силу их несоизмеримости. Например, продукцию промышленности, сельского хозяйства, чтобы получить их общий объем, в силу их разнообразия непосредственно суммировать нельзя.

Для того чтобы получить суммарный объем разнообразной продукции необходимо свести их к одному единству через соизмерения с помощью таких показателей, как цена, себестоимость, трудоемкость единицы продукции.

С помощью индексных показателей решаются следующие задачи:

1) характеристика общего изменения сложного показателя (например затрат на производство продукции, стоимости произведенной или реализованной продукции т.д.);

2) выделение в изменении сложного показателя влияния одного из факторов через элиминирование влияния других факторов

(например, увеличение выручки, связанное с ростом цен или увеличением объема выпуска продукции в натуральном выражении).

6.2. Виды индексов и их показатели

По степени охвата элементов изучаемой совокупности различают индивидуальные и общие (сводные) индексы. Индивидуальными называются индексы, характеризующие изменение только одного элемента совокупности. Индивидуальный индекс обозначается – i.

Общий индекс характеризует изменение элементов всей изучаемой совокупности и обозначается – I. Если индексы охватывают не все элементы общей совокупности, а лишь часть их, то их называют групповыми или субиндексами.

При построении индексов используют стандартные обозначения, так количество произведенной продукции обозначается – q_i; цена единицы продукции – p_i; себестоимость единицы продукции – z_i; трудоемкость – t_i.

В зависимости от содержания и характера индексируемой величины различают индексы количественных (объемных) показателей, как например, индексы физического объема продукции и индексы качественных показателей, как индексы цен, себестоимости.

При построении индексов различают сравниваемый уровень и уровень, с которым производится сравнение, называемый базисным. Базисные индексы получают, когда уровни периода сравниваются с уровнем одного периода, принятого за базу сравнения.

Если индексы получают сопоставлением текущих (отчетных) уровней с предшествующим уровнем, то их называют цепным индексом. При построении цепного индекса база сравнения непрерывно меняется.

Общие сводные индексы применяются, когда необходимо дать обобщенную характеристику сложным явлениям, элементы которых несоизмеримы между собой. Так, если имеются данные об экспорте продукции по отдельным предприятиям, то обобщенную количественную характеристику экспорта продукции в целом по стране нельзя дать, используя только натуральные

показатели, поскольку предприятия выпускают и экспортируют разнородные продукции.

Для получения обобщенной характеристики необходимо данные по этим различным видам продукции привести к единой общей мере, используя стоимостную оценку экспортируемой продукции, т .е. вместо показателя: $\frac{\sum q_1}{\sum q_0}$, необходимо использовать показатель: $\frac{\sum q_1 p_1}{\sum q_0 p_0}$, в числителе которой стоимость продукции отчетного периода, а в знаменателе – стоимость продукции базисного периода.

В данном случае получаем индекс, на уровень которого влияют уровни цен и количество продукции. Данный индекс не может характеризовать изменение объема продукции. Для этого необходимо элиминировать влияние цен, используя в индексах цены одного периода.

Тогда получим следующий общий (агрегатный) индекс физического объема:

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

где q_1, q_0 – количество продукции отчетного и базисного периодов, кроме того они же являются индексируемыми величинами, изменение которых измеряются;

p_0 – уровень цен базисного периода и весы, через который обеспечивается сопоставимость разных видов продукций и в тоже время элиминируется влияние уровня цен на общий уровень индекса физического объема.

Индекс физического объема (I_q) с весами цен базисного периода был предложен Ласпейресом в 1864г.

В экономико-статистическом анализе не ограничивается исчислением отдельных изолированных индексов характеризующие изменение изучаемых явлений за какой-то один период времени, а исчисляют несколько индексов, охватывающие ряд последовательные периоды времени (таблица 6.1.).

При таком исчислении индексов применяют в качестве соизмерителя (весов) цены одного периода. Такие цены называются сопоставимыми, фиксированными или неизменными, которые применяются на протяжении длительного времени в условиях стабильного развития экономики.

Применяя в качестве соизмерителя неизменные цены, получаем следующую формулу индекса физического объема продукции:

$$I_q = \frac{\sum q_1 p_н}{\sum q_0 p_н},$$

где $p_н$ – неизменная оптовая цена единицы продукции.

Применение неизменных (постоянных) цен в качестве соизмерителя дает возможность путем суммирования получить итоговые данные стоимости продукции за месяц, квартал, полугодие, год по предприятиям и отрасли в целом.

Если в качестве соизмерителя (весов) использовать цены отчетного периода, тогда индекс физического объема продукции можно построить в следующем виде:

$$I_q = \frac{\sum q_1 p_1}{\sum q_0 p_1}$$

Данный общий индекс физического объема с соизмерителем цен отчетного периода был предложен в 1874г. Г. Паше (Paoshe).

Таблица 6.1. – Пример для расчета индивидуального и общего (агрегатного) индексов

Виды продукций	Выпуск продукции в базисном периоде (q_0)	Выпуск продукции в отчетном периоде (q_1)	Цена за 1 продукцию в базисном периоде (p_0)	Цена за 1 продукцию в отчетном периоде (p_1)
А	21	20	75,0	82,5
В	115	120	8,75	10,10

1. Индивидуальные индексы физического *объема*:

$$i_{q(A)} = \frac{q_1}{q_0} = \frac{20}{21} = 0,95; \quad i_{q(B)} = \frac{q_1}{q_0} = \frac{120}{115} = 1,04.$$

2. Общие (агрегатные) индексы физического *объема* Ласпейреса и Паше:

$$I_q^L = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{20 \cdot 74,0 + 120 \cdot 8,75}{21 \cdot 75,0 + 115 \cdot 8,75} = \frac{2550}{2581} = 0,988;$$

$$I_q^P = \frac{\sum q_1 p_1}{\sum q_0 p_1} = \frac{20 \cdot 82,5 + 120 \cdot 10,1}{21 \cdot 82,5 + 115 \cdot 8,75} = \frac{2862}{2894} = 0,989.$$

1. Индивидуальные индексы *цен*:

$$i_{p(A)} = \frac{p_1}{p_0} = \frac{82,5}{75,0} = 1,1; \quad i_{p(B)} = \frac{10,1}{8,75} = 1,15.$$

2. Общие (агрегатные) индексы *цен* Ласпейреса и Пааше:

$$I_p^L = \frac{\sum p_1 q_0}{\sum p_0 q_0} = \frac{82,5 \cdot 21 + 10,1 \cdot 115}{75,0 \cdot 21 + 8,75 \cdot 115} = \frac{2894}{2581,2} = 1,121;$$

$$I_p^P = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{82,5 \cdot 20 + 10,1 \cdot 120}{75 \cdot 20 + 8,75 \cdot 120} = \frac{2862}{2550} = 1,122.$$

6.3. Среднеарифметический и среднегармонический индексы

Общие индексы, построенные с использованием индивидуальных индексов, принимают форму среднего арифметического или среднего гармонического индекса.

Методика построения среднего арифметического и среднего гармонического индексов физического объема. Если известна стоимость продукции каждого вида в базисном периоде $(p_0 q_0)$ и индивидуальные индексы физического объема, $(i_q = \frac{q_1}{q_0})$, то для построения среднего индекса используем в качестве базы общий (агрегатный) индекс Ласпейреса: $I_q^L = \frac{\sum q_1 p_0}{\sum q_0 p_0}$.

В этом индексе непосредственно можно суммировать фактические данные знаменателя этой формулы. Числитель же может быть получен перемножением стоимости отдельного вида продукции базисного периода на индивидуальный индекс: $q_1 p_0 = p_0 q_0 \cdot i_q = p_0 q_0 \cdot \frac{q_1}{q_0}$, где $i_q = \frac{q_1}{q_0}$; $q_1 = i_q \cdot q_0$.

Тогда форма общего (агрегатного) индекса физического объема Ласпейреса принимает следующий вид: $I_q^L = \frac{\sum i_q \cdot p_0 q_0}{\sum p_0 q_0}$, т. е. среднеарифметический индекс физического объема.

Используя данные предыдущего примера, определим среднеарифметический индекс физического объема продукции:

$I_q^L = \frac{0,95 \cdot 1575 + 1,04 \cdot 1006}{1575 + 1006} = 0,988$, т. е. получаем тот же результат, что и при расчете общего индекса физического объема Ласпейреса.

Если имеются данные об индивидуальных индексах выпуска продукции (i_q), а также о стоимости каждого вида продукции в отчетном периоде ($p_1 q_1$), то, используя формулу Паше, можно построить среднегармонический индекс.

В формуле Паше числитель индекса можно получить суммированием величин ($p_1 q_1$), а знаменатель – делением фактической стоимости каждого вида продукции на соответствующий индекс физического объема продукции: $\left(\frac{p_1 q_1}{i_q}\right) = p_1 q_0$. Тогда получим: $I_q^P = \frac{\sum q_1 p_1}{\sum q_0 p_1} = \frac{\sum q_1 p_1}{\sum \frac{q \ p_1}{i_q}}$.

Используя исходные данные предыдущего примера, рассчитаем общий среднегармонический индекс физического объема Пааше:

$$I_q^P = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{i_q}} = \frac{1650 + 1212}{\frac{1650}{0,95} + \frac{1212}{1,04}} = 0,989.$$

Расчет общих (агрегатных) индексов может производиться на основе данных о стоимостных (а не натуральных) объемах выпуска каждого вида продукции и индивидуальных индексах.

Особенности общих (агрегатных) индексов Ласпейреса и Пааше

Различия в соотношении общих (агрегатных) индексов Ласпейреса и Пааше и их индивидуальных индексов определяются относительной вариацией индивидуальных индексов цен (V_{ip})

индивидуальных индексов физического объема (V_{iq}) и коэффициентом корреляции, оценивающим степень тесноты связи между ними: $\frac{I_p^P}{I_p^L} = 1 + r_{ipiq} \cdot V_{ip} \cdot V_{iq}$;

где I_p^P, I_p^L – общие (агрегатные) индексы цен Пааше и Ласпейреса;

V_{ip}, V_{iq} – коэффициенты вариации индивидуальных индексов цен и физического объема;

$$V_{ip} = \frac{\sigma_{ip}}{I_p^L}; \sigma_{ip} = \sqrt{\frac{\sum (i_{p_i} - I_p^L)^2 p_0 q_0}{\sum p_0 q_0}};$$

$$V_{iq} = \frac{\sigma_{iq}}{I_q^L}; \sigma_{iq} = \sqrt{\frac{\sum (i_{q_i} - I_q^L)^2 p \ q_0}{\sum p_0 q_0}}.$$

Степень тесноты связи между индивидуальными индексами цен и физического объема продукции определяют по следующей формуле:

$$r_{iqip} = \frac{\sum (i_{pi} - I_p^L)(i_{qi} - I_q^L) p_0 q_0}{\sum p_0 q_0 \cdot \sigma_{ip} \sigma_{iq}}.$$

Для характеристики общего изменения уровня анализируемого показателя предпочтение отдают индексу Ласпейреса. Так, при исчислении общего (агрегатного) индекса физического объема продукции достаточно вести мониторинг за изменением физических объемов продукции, тогда как при использовании общего (агрегатного) индекса Пааше необходимо учитывать как изменение уровня цен, так и физического объема продукции.

Расчет общего (агрегатного) индекса физического объема продукции по формуле Ласпейреса получило наибольшее распространение в мировой практике, поскольку, используя стабильную структуру потребления продуктов и услуг, при расчете индекса потребительских цен (ИПЦ) производят индексацию доходов населения. Применение агрегатного индекса Ласпейреса дает возможность перехода от ряда индексов с переменными базами сравнения к ряду индексов с постоянной базой сравнения и наоборот. Если перемножить общие (агрегатные) индексы Ласпейреса и Пааше между собой, то получим формулу, с помощью которой можно определить изменение уровня стоимости продукции. Кроме того, рассчитав среднюю геометрическую из двух общих (агрегатных) индексов получают, так называемый комбинированный общий (агрегатный) индекс Фишера.

Таблица 6.2. – Общий (агрегатный) индексы физического объема и цен Ласпейреса, Пааше и Фишера:

Название индекса	Агрегатный индекс физического объема	Агрегатный индекс цен
Индекс Ласпейреса с базисными весами	$I_q^L = \frac{\sum q_1 p_0}{\sum q_0 p_0}$	$I_p^L = \frac{\sum p_1 q_0}{\sum p_0 q_0}$
Индекс Пааше с весами отчетного периода	$I_q^P = \frac{\sum q_q p_1}{\sum q_0 p_1}$	$I_p^P = \frac{\sum p_1 q_1}{\sum p_0 q_1}$
Комбинированный индекс Фишера	$I_q^F = \sqrt{\frac{\sum q_1 p_0}{\sum q_0 p_0} \times \frac{\sum q_1 p_1}{\sum q_0 p_1}}$	$I_p^F = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}}$

Таблица 6.3. – Расчет общих (агрегатных) индексов Ласпейреса, Пааше и Фишера, используя исходные данные следующего условного примера:

Товар	Продано в базисном периоде, шт. (q_0)	Продано в отчетном периоде, шт. (q_1)	Цена за единицу, сом. в базисном периоде (p_0)	Цена за единицу, сом. в отчетном периоде (p_1)
A	60	55	50	80
B	45	51	15	18
C	50	80	60	70

Расчет общих (агрегатных) индексов физического объема Ласпейреса, Пааше и Фишера:

1) $I_q^L = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{55 \cdot 50 + 51 \cdot 15 + 80 \cdot 60}{60 \cdot 50 + 45 \cdot 15 + 50 \cdot 60} = \frac{8315}{6675} = 1,245;$

2) $I_q^P = \frac{\sum q_1 p_1}{\sum q_0 p_1} = \frac{55 \cdot 80 + 51 \cdot 18 + 80 \cdot 70}{60 \cdot 80 + 45 \cdot 18 + 50 \cdot 70} = \frac{10918}{9150} = 1,193;$

3) $I_q^F = \sqrt{I_q^L \times I_q^P} = \sqrt{1,245 \cdot 1,193} = 1,218.$

Расчет общих (агрегатных) индексов цен Ласпейреса, Пааше и Фишера:

1) $I_p^L = \frac{\sum p_1 q_0}{\sum p_0 q_0} = \frac{9150}{6675} = 1,370;$

2) $I_p^P = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{10918}{8315} = 1,313;$

3) $I_p^F = \sqrt{I_p^L \times I_p^P} = \sqrt{1,370 \cdot 1,313} = 1,341.$

6.4. Индексы переменного, фиксированного составов и структурных сдвигов

Индекс переменного состава измеряет изменение среднего уровня индексируемого показателя в отчетном периоде по сравнению с базисным периодом. При нахождении их средних значений соизмерителями (весами) выступают объемы произведенной продукции соответственно отчетного и базисного периодов.

Используя исходные данные таблицы 6.3, рассчитаем индекс переменного состава:

$$I_p^{пс} = \frac{\overline{p}_1}{\overline{p}_0} = \frac{\sum p_1 q_1}{\sum q_1} / \frac{\sum p_0 q_0}{\sum q_0} = \frac{58,70}{43,06} = 1,363.$$

Индекс переменного состава, в целом характеризует изменение среднего уровня затрат отчетного периода по сравнению с базисным периодом, когда на общий уровень затрат влияет уровни цен на отдельных участках и объемы выпускаемой продукции в соответствующих периодах.

Индекс фиксированного состава учитывает изменение индексируемой величины при фиксированном (постоянном) уровне соизмерителя (веса). В данном случае в качестве соизмерителя используется фиксированный объем произведенной продукции отчетного периода. Общий (агрегатный) индекс цен фиксированного состава рассчитывается по следующей формуле, используя исходные данные таблицы 3:

$$I_p^{фс} = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{10918}{8315} = 1{,}313.$$

Индекс структурных сдвигов находят как отношение индекса переменного состава к индексам фиксированного состава. Индекс структурных сдвигов показывает, как влияет изменение объема и структуры выпускаемой продукции на общий уровень стоимости продукции при элиминировании влияния цен. Общий (агрегатный) индекс структурных сдвигов рассчитывается по следующей формуле с использованием предыдущих количественных показателей:

$$I_p^{сс} = \frac{I_p^{пс}}{I_p^{фс}} = \frac{1{,}363}{1{,}313} = 1{,}04.$$

6.5. Использование индексов в экономическом анализе

Индекс используется для характеристики изменения уровня сложных явлений. Их можно использовать также и в аналитических целях для оценки влияния на результативный показатель факторных признаков. При этом, если результативный признак (показатель) является произведением двух или более определяющих его величину показателями (факторами) или суммой таких произведений, как например, стоимостный объем произведенного (реализованного) товара в натуральном выражении, когда стоимость товара зависит от изменения цен или объема товара в натуральном выражении. Поэтому необходимо показать, в какой мере изменение стоимости товара вызвано каждым из этих факторов.

Оценить роль отдельных факторов в изменении результативного показателя можно путем построения системы взаимосвязанных индексов. В основе построения аналитических индексов лежит принцип элиминирования влияния отдельных факторов, кроме анализируемого фактора. При этом необходимо обеспечить выполнения следующих условий:

1) при расчете индексов качественных показателей, веса в общем (агрегатном) индексе фиксируются на уровне отчетного (текущего) периода, т. е. используется общий (агрегатный) индекс Пааше;

2) при расчете количественных общих (агрегатных) индексов соизмерители (веса) принимаются на уровне базисного периода, т. е. используется общий (агрегатный) индекс Ласпейреса.

Используя перечисленные выше общие (агрегатные) индексы Паше и Ласпейреса, а также данные таблицы 6.3. можно определить абсолютный прирост результативного показателя по факторам:

$$\Delta I_{pq} = \Delta I_{pq(p)} + \Delta I_{pq(q)},$$

где ΔI_{pq} – абсолютный прирост стоимости продукции;

$\Delta I_{pq(p)}$ – абсолютный прирост стоимости продукции, обусловленный изменением уровня цен на продукцию;

$\Delta I_{pq(q)}$ – абсолютный прирост стоимости продукции, обусловленный изменением физического объема продукции.

Абсолютные приросты стоимости продукции определяются как разность между числителем и знаменателем соответствующего общего (агрегатного) индекса:

1) $I_{pq} = \dfrac{\sum p_1 q_1}{\sum p_0 q_0}; \Delta I_{pq} = \sum p_1 q_1 - \sum p_0 q_0 = 19180 - 12765 = 6415;$

2) $I_p = \dfrac{\sum p_1 q_1}{\sum p_0 q_1}; \Delta I_{pq(p)} = \sum p_1 q_1 - \sum p_0 q_1 = 19180 - 15200 = 3980;$

3) $I_q = \dfrac{\sum q_1 p_0}{\sum q_0 p_0}; \Delta I_{pq(q)} = \sum q_1 p_0 - \sum q_0 p_0 = 15200 - 12765 = 2435;$

4) $\Delta I_{pq} = \Delta I_{pq(p)} + \Delta I_{pq(q)} = 3980 + 2435 = 6415.$

Приведенные данные показывают, что абсолютный прирост стоимости продукции, когда воздействуют совместно факторы цен и физического объема произведенной продукции, равен сумме абсолютного прироста стоимости продукции; когда факторы цен и физического объема произведенной продукции оказывают воздействия – в отдельности.

6.6. Статистические показатели индекса цен

В условиях рыночных отношений цена является особенно востребованной категорией, от уровня которой зависит уровень жизни населения и непосредственно влияет на уровень и состояние таких макроэкономических показателей, как Валовой внутренний продукт (ВВП). Национальный доход (НД), уровень потребления и сбережения.

Изучение рыночной конъюнктуры начинается с анализа средних цен и их уровней. Различают индивидуальный и средний уровень цен на товары и услуги первой необходимости. Индивидуальный уровень цен – это абсолютная величина цены в денежном выражении за единицу конкретного товара на рынке.

Средний уровень цен – это обобщающий показатель уровня цен, исчисленный по однородным группам товаров во времени и в пространстве. Для расчета средних цен (уровней цен) используются следующие виды средних величин:

1. Средняя арифметическая, взвешанная: $\overline{p} = \dfrac{\sum pq}{\sum q}$,

где p – цена единицы товара;

q – количество товара.

Средняя арифметическая, взвешанная применяется в тех случаях, когда в качестве весов применяется показатели количества товаров в натуральном выражении;

2. Средняя гармоническая, взвешанная: $\overline{p} = \dfrac{\sum pq}{\sum \dfrac{pq}{p}}$,

где pq – товарооборот в денежном выражении.

Средняя гармоническая применяется в том случае, если в качестве весов используются стоимостные данные продажи товара, т. е. товарооборот.

3. Средняя хронологическая, простая:

$$\overline{p} = \frac{1/2p_1 + p_2 + p \ + \ldots + p_{n-1} + 1/2p_n}{n-1},$$

где $p_1 \ldots p_n$ – уровни цен по дням месяца;

n – число дней месяца.

4. Средняя хронологическая, взвешанная: $\overline{p} = \dfrac{\sum \overline{p}_1 t_i}{\sum t_i}$,

где p_i – средняя цена за период;

t_i – число дней в периоде.

При анализе динамики и структуры изменения цен на различные товары и услуги в целом используются общие (агрегатные) индексы Ласпейреса и Пааше:

1. Общий (агрегатный) индекс Ласпейреса: $I_p^L = \dfrac{\sum p_1 q_0}{\sum p_0 q_0}$,

где p_1, p_0 – цена единицы товара в отчетном и базисном периодах;

q_0 – количество товара в базисном периоде.

2. Общий (агрегатный) индекс Пааше: $I_p^P = \dfrac{\sum p_1 q_1}{\sum p_0 q_1}$,

где q_1 –количество товаров в отчетном периоде.

Используя общие (агрегатные) индексы Ласпейреса, Пааше можно определить среднюю геометрическую из этих индексов – общий (агрегатный) индекс Фишера: $I_p^F = \sqrt{I_p^L \times I_p^P}$.

Общие (агрегатные) индексы цен Ласпейреса и Пааше используются в российской статистической практике, а общий (агрегатный) индекс Фишера используется в основном в международных сопоставлениях ВВП.

В системе индексов цен, наряду с общими (агрегатными) формами индексов, широко применяются и индивидуальные индексы цен, индексы средних цен. Индивидуальный индекс цен характеризует динамику изменения цен конкретного товара (услуги):

$i_p = \dfrac{p_t}{p_0}$ – базисный индекс цен;

$i_p = \dfrac{p_t}{p_{t-1}}$ – цепной индекс цен,

где p_t – цена на товар в текущем периоде;

p_{t-1} – цена на товар в предыдущем периоде;

p_0 – цена на товар в базисном периоде.

Используя взаимосвязи индексов, величина базисного индекса цен определяется как произведение цепных индексов:

$$i_{p5/0} = i_{p2/1} \cdot p_{p3/2} \cdot i_{p4/3} \cdot i_{p5/4}$$

Индивидуальные индексы цен применяются при изучении динамики цен разнообразной совокупности товаров и услуг. Кроме того, они используются в различных модификациях общих (агрегатных) индексов цен. Так, например, общие (агрегатные) индексы Ласпейреса и Пааше можно представить в виде среднеарифметической и среднегармонической индексов цен:

$$I_p^L = \frac{\sum i_p p_0 q_0}{\sum p_0 q_0}; I_p^P = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}}.$$

В этих формулах индивидуальные индексы цен позволяют выявить роль отдельных товаров в формировании общего (агрегатного) индекса цен.

Индексный анализ динамики средних цен производится на основе использования общих (агрегатных) индексов цен переменного, фиксированного состава и структурных сдвигов:

$I_{\overline{p}пер} = \frac{\overline{p}_1}{\overline{p}_0} = \frac{\sum p_1 q_1}{\sum q_1} / \frac{\sum p_0 q_0}{\sum q_0}$ – общий индекс переменного состава;

$I_{\overline{p}фик}^P = \frac{\sum p_1 q_1}{\sum p_0 q_1}$ – общий (агрегатный) индекс цен фиксированного состава;

$I_{\overline{p}стр} = \frac{I_{\overline{p}пер}}{I_{\overline{p}фик}}$ – общий индекс структурных сдвигов.

Индексы средних цен целесообразно рассчитывать при изучении цен как одного товара, так и однородных товарных групп по различным территориям и рынкам. В этом случае индекс структурных сдвигов отражает влияние изменения качества товара, перераспределение товарной массы, изменение структуры продаж и денежных доходов населения.

Контрольные вопросы,
определение понятий и показателей.

1. Основное назначение индексов, виды индексов.
2. Особенности построения и использования индивидуальных и общих (агрегатных) индексов цен и физического объема.
3. Среднеарифметические и среднегармонические индексы.
4. Общие (агрегатные) индексы Ласпейреса, Пааше и Фишера.
5. Индексы переменного, фиксированного составов и структурных сдвигов.

ЛИТЕРАТУРА

Ефимова М.Р. Общая теория. Статистика: учебник / М.Р. Ефимова, Е.В. Петрова, В.Н. Румянцев. М., 1996, 1999, 2009.

Статистика: учебник / под ред. И.И. Елисеевой. М., 2006, 2010.

Елисеева И.И, Юзбашев М.М.. Общая теория статистики: учебник. М., 1999.

Венецкий И.Г., Кильдишев Г.С. Основы теории вероятности и математической статистики. М., 1968.

Карасев А.И. Теория вероятностей и математическая статистика. М., 1977г.

Статистика: учебник: КРСУ, Бишкек. 2013.

Качибек Алиевич Алиев

СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ

Учебное пособие

Редактор *Е.М. Кузичева*
Компьютерная верстка *З.Б. Турашевой*

Подписано в печать 31.10.2016
Формат 60×84 $^{1}/_{8}$. Печать офсетная.
Объем 7,0 пл. л. Тираж 100 экз. Заказ 5

Издательство КРСУ
720000, г. Бишкек, ул. Киевская, 44

Отпечатано в типографии КРСУ
720048, г. Бишкек, ул. Горького, 2