ВЫСШЕЕ ОБРАЗОВАНИЕ

серия основана в 1996 г.

Г.Л. ГРОМЫКО

ТЕОРИЯ СТАТИСТИКИ

ПРАКТИКУМ

Издание второе, дополненное и переработанное

...осква
ИНФРА-М
2001

УДК (075.8)311
ББК 60.6я73
Г86

Г86 **Громыко Г.Л.** Теория статистики: Практикум. — 2-е изд., доп. и перераб. — М.: ИНФРА-М, 2001. — 160 с. — (Серия «Высшее образование»).

ISBN 5-16-000583-8

Автор — доктор экономических наук, профессор Галина Леонтьевна Громыко много лет преподает статистику в МГУ им. М.В. Ломоносова.

Практикум содержит методические указания и решения задач по основным темам теории статистики.

Для студентов экономических вузов.

ББК 60.6я73

ISBN 5-16-000583-8
© Г.Л. Громыко, 2001

Корректор *И.А. Лукашевич*
Компьютерная верстка *В.О. Комлев*
Оформление серии *Е.А. Доний*

ЛР № 070824 от 21.01.93 г.

Подписано в печать 06.02.2001.
Формат 60x88/16. Гарнитура NewtonC. Печать офсетная.
Усл. печ. л. 9,8. Тираж 6000 экз. Заказ № 3049.
Цена договорная.

Издательский Дом «ИНФРА-М»,
127214, Москва, Дмитровское ш., 107.
Тел. 485-70-63, 485-74-00.
E-mail: books@infra-m.ru
http://www.infra-m.ru

Отпечатано с готовых диапозитивов в Тульской типографии.
300600, г. Тула, пр. Ленина, 109.

ПРЕДИСЛОВИЕ

Неотъемлемой частью экономического образования являются статистическая грамотность, умение пользоваться статистическими методами исследования, позволяющими обнаружить различные статистические закономерности. Поэтому улучшение экономического образования предполагает обязательное повышение статистической грамотности специалистов в разных отраслях знаний. В этой связи особое значение приобретает изучение в вузах курса «Теория статистики» с элементами математической статистики.

Для изучения данного курса в соответствии с существующей программой рекомендуются следующие учебники*:

1. Теория статистики: Учебник/Под ред. профессора Г.Л. Громыко. — М.: ИНФРА-М, 2000. — 414 с.
2. Теория статистики: Учебник/Под ред. профессора Р.А. Шмойловой. — 2-е изд. — М.: Финансы и статистика, 1998. — 576 с.
3. *Елисеева И.И., Юзбашев М.М.* Общая теория статистики. — М.: Финансы и статистика, 1995. — 356 с.
4. *Ефимова М.Р., Петрова Е.В., Румянцев В.Н.* Общая теория статистики: Учебник. — 2-е изд., испр. и доп. — М.: ИНФРА-М, 2000. — 416 с.

Для самопроверки и закрепления знаний полезно ответить на контрольные вопросы, самостоятельно решить несколько задач, помещенных в конце каждой темы. Большинство задач содержит данные официальных статистических сборников разных лет. К части задач приведены ответы. При этом в отдельных случаях не исключена возможность несовпадений последних из-за различной точности округлений в промежуточных расчетах.

В приложениях приведены основные математико-статистические таблицы, использование которых предусмотрено в ряде задач.

---

* Далее в тексте ссылки на литературу даются цифрами в квадратных скобках, соответствующими порядковому номеру рекомендуемой литературы.

Тема 1
СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ

Статистическое наблюдение — первый этап любого статистического исследования. Оно сводится к сбору данных о массовых явлениях, регистрации различных фактов или признаков у отдельных единиц наблюдения. Это источник первичной статистической информации, являющейся частью экономической информации вообще.

Статистическое наблюдение должно проводиться по заранее составленному плану, в котором следует различать *программно-методологические* и *организационные* вопросы. Первые включают в себя определение цели, объекта, единицы наблюдения, программы (перечня вопросов, на которые надо получить ответы). К организационным вопросам относится выбор формы, места и времени наблюдения, способа и вида наблюдения и пр.

По форме статистическое наблюдение может быть проведено через *отчетность* и *специально организованные обследования*, из которых важнейшими являются переписи, отражающие состояние изучаемой совокупности на определенный момент времени.

Долгие годы в нашей стране основным источником статистических данных была отчетность. В последнее время в связи с переходом к рыночной экономике, разгосударствлением и развитием различных форм собственности роль отчетности заметно снизилась и возросла роль различных единовременных выборочных обследований.

Исследуемые факты могут относиться либо к какому-либо периоду (добыча угля за год, количество родившихся за месяц, год и т.д.), либо к определенному моменту времени (численность населения на начало года, автомобильный парк на конец года и т.д.). Та дата, на которую регистрируются факты, называется **критическим моментом**. Период, в течение которого проводится сбор данных, называется **временем наблюдения.**

Статистические данные могут быть получены разными способами: путем непосредственного измерения (подсчета, взвешивания и т.д.), путем опроса и пр.

По охвату единиц наблюдаемого объекта наблюдение может быть *сплошным* (когда обследуются все единицы совокупности) и *несплошным* (когда обследуется только часть единиц совокупности).

Последнее может быть организовано как исследование основного массива, анкетное, выборочное и монографическое.

При изучении этой темы следует помнить, что от правильно организованного и хорошо продуманного наблюдения зависят полнота получаемых данных и точность выводов в результате обработки собранных данных. Следует особое внимание обратить на составление статистического формуляра — бланка, в котором регистрируются сведения о единицах наблюдения, и на составление инструкций — письменных разъяснений по вопросам заполнения статистических формуляров и организации наблюдения.

В практике статистики используются как карточные бланки (заполняемые на каждую единицу отдельно), так и списочные (заполняемые для нескольких единиц одновременно).

В последние годы в связи с широким внедрением новой вычислительной техники появились новые носители статистической информации в форме магнитных лент и др.

Собранные в процессе наблюдения сведения должны быть подвергнуты проверке путем логического и арифметического контроля.

Литература

[1, с. 18—35], [2, с. 41—60], [3, с. 3—52]

Контрольные вопросы

1. Назовите этапы статистического исследования.
2. В чем суть статистического наблюдения?
3. Каковы основные организационные формы статистического наблюдения в нашей стране?
4. С какой целью составляется план статистического наблюдения?
5. Перечислите программно-методологические вопросы плана статистического наблюдения.
6. Что такое объект и единица наблюдения?
7. Что такое программа статистического наблюдения?
8. Назовите виды носителей статистической информации.
9. Какие вопросы относятся к организационным в плане статистического наблюдения?
10. Какое наблюдение называют текущим и какое прерывным?
11. Как различается наблюдение с точки зрения охвата единиц изучаемой совокупности?
12. Назовите виды несплошного наблюдения.
13. Назовите способы статистического наблюдения.
14. Что такое критический момент и время наблюдения?
15. Какие ошибки могут возникнуть в процессе наблюдения, какие существуют способы их предотвращения и контроля?

Задачи и упражнения для самостоятельной работы

1. Какими наиболее существенными признаками можно охарактеризовать такие единицы статистического наблюдения, как:

а) промышленное предприятие;
б) фермерское хозяйство;
в) торговое предприятие;
г) больница;
д) библиотека;
е) школа;
ж) высшее учебное заведение (государственное, негосударственное);
з) преподаватель вуза;
и) студент вуза;
к) семья;
л) детский сад.

2. Выберите по своему усмотрению интересующую Вас единицу статистического наблюдения и перечислите основные ее признаки: количественные и неколичественные (атрибутивные).

3. Определите цель и разработайте программу:

а) статистического обследования школ города;
б) выборочного обследования читателей библиотеки;
в) выборочного обследования студентов (одного из курсов, факультета, университета, всех вузов города);
г) переписи производственного промышленного оборудования.

4. Составьте анкету опроса студентов в целях выяснения:

а) их возрастного и полового состава, семейного положения, успеваемости, научных интересов;
б) их оценки качества преподавания отдельных дисциплин и мастерства преподавателей;
в) бюджета их времени и использования свободного времени;
г) их бытовых условий и материального положения.

5. Определите форму, вид (по времени регистрации и по охвату единиц наблюдения) и способ следующих статистических наблюдений:

а) всеобщая перепись населения страны;
б) бюджетные обследования семей;
в) инвентаризация основных фондов;
г) изучение цен на рынках;
д) определение качества продукции на отдельном предприятии;
е) перепись скота в стране;
ж) опрос общественного мнения по тем или иным проблемам.

Тема 2
СВОДКА И ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ. ПОСТРОЕНИЕ РЯДОВ РАСПРЕДЕЛЕНИЯ

Вторым этапом статистического исследования является **сводка**, суть которой в обработке первичных материалов наблюдения в целях получения итоговых или упорядоченных определенным образом числовых характеристик той или иной изучаемой совокупности. Основным и важнейшим моментом сводки является **группировка**, т.е. объединение статистических данных в однородные по определенным признакам группы.

Группировки помогают изучать структуру совокупности, взаимосвязь между явлениями.

Изучение структуры той или иной совокупности достигается построением **рядов распределения**, характеризующих распределение единиц совокупности по одному признаку.

Распределение единиц совокупности по количественному признаку называют **вариационным рядом.**

При изучении данной темы надо уделить особое внимание вопросу построения вариационных рядов по дискретному и непрерывному признакам.

Дискретным называется признак, который может принимать определенные значения из конечного набора таких значений, выражаемых, как правило, целыми числами (например, число детей в семье).

Непрерывный признак может принимать любые промежуточные значения (например, урожайность, возраст и др.). Как правило, при построении вариационных рядов по непрерывному признаку последний указывается в виде интервалов «от и до», и ряд называется **интервальным.**

Отдельные значения группировочного признака называются **вариантами** (обозначаются x_i), а числа, показывающие, сколько раз встречается то или иное значение признака, — **частотами**, если они выражены абсолютными величинами (обозначаются m или f), или **частостями**, если это относительные величины (обозначаются w_i).

Рассмотрим построение дискретного ряда на следующем примере.

Задача 2.1

Пусть имеются следующие данные о тарифных разрядах 50 рабочих одного из цехов завода:

3	5	6	3	2	4	3	5	5	6
4	3	2	3	4	5	4	2	4	6
5	3	4	5	4	3	3	6	2	3
4	6	3	4	4	5	4	5	3	4
2	6	3	4	5	3	4	4	5	4

Чтобы показать распределение рабочих по тарифному разряду, построим вариационный ряд, для чего выпишем все значения признака (тарифного разряда) в порядке возрастания и подсчитаем число рабочих в каждой группе:

Тарифный разряд (вариант x_i)	**Численность рабочих (частота f_i)**
2	5
3	13
4	16
5	10
6	6
Всего	**50**

Это дискретный вариационный ряд, у которого вариантами являются значения тарифного разряда, а частотами — число рабочих.

Численность рабочих можно выразить также в долях, тогда последние именуются частостями и обозначаются как $w_i = \frac{f_i}{\sum f_i}$.

Естественно, $\sum w_i = 1$. Относительные численности (частости) выражают и в процентах. Тогда $\sum w_i = 100$.

Задача 2.2

Для построения интервального ряда с равными интервалами воспользуемся следующими данными о стоимости основных фондов (непрерывный признак) у 50 предприятий, млн. руб.:

9,4	8,0	6,3	10,0	15,0	8,2	7,3	9,2	5,8	8,7
5,2	13,2	8,1	7,5	11,8	14,6	8,5	7,8	10,5	6,0
5,1	6,8	8,3	7,7	7,9	9,0	10,1	8,0	12,0	14,0
8,2	9,8	13,5	12,4	5,5	7,9	9,2	10,8	12,1	12,4
12,9	12,6	6,7	9,7	8,3	10,8	15,0	7,0	13,0	9,5

Чтобы показать распределение предприятий по стоимости основных фондов, сначала решим вопрос о количестве групп, которые мы хотим выделить. Предположим, решено выделить 5 групп заводов. Чтобы определить величину интервала в группе, найдем разность между максимальным и минимальным значениями признака и разделим ее на число выделяемых групп. Если обозначить величину интервала через h, то в нашем примере $h = \frac{15 - 5{,}1}{5} \approx 2$ (млн. руб.).

Выделим теперь группы с интервалом 2 млн. руб. и подсчитаем число заводов в каждой группе (частоту):

Стоимость основных фондов, млн. руб.	Число заводов (частоты)	Накопленные (кумулятивные) частоты
5—7	9	9
7—9	16	25
9—11	11	36
11—13	8	44
13—15	6	50
Всего	**50**	—

Это интервальный вариационный ряд с равными интервалами. При такой записи непрерывного признака, когда одна и та же величина встречается дважды (как верхняя граница одного интервала и как нижняя граница другого интервала), единица, обладающая этим значением, обычно относится к той группе, где эта величина выступает в роли верхней границы. Так, в нашем примере завод со стоимостью основных фондов 9 млн. руб. отнесен ко второй группе (а не к третьей).

Кроме обычных частот в вариационном ряду можно рассчитать нарастающим итогом **накопленные** (кумулятивные) частоты, по которым строим суждение о том, какое число единиц в совокупности обладает значением признака «не более» или «не менее» определенного. Так, в нашем примере можно сказать, что 25 заводов из 50, т.е. половина, имеют основные фонды, стоимость которых не превышает 9 млн. руб.

Если вопрос о количестве выделяемых групп, а следовательно, и о величине интервала, вызывает сомнение, то рекомендуется использовать формулу Стерджесса, предусматривающую выделение **оптимального** числа групп ($K = 1 + 3{,}322 \lg N$) при заданной численности совокупности (N):

$$h = \frac{x_{max} - x_{min}}{1 + 3{,}322 \lg N}.$$

В рассмотренном выше примере это составило бы

$$h = \frac{15 - 5{,}1}{1 + 3{,}322 \lg 50} \approx \frac{9{,}9}{6{,}644} \approx 1{,}49 \approx 1{,}5 \text{ (млн. руб.)},$$

т.е. можно было выделить 7 групп заводов ($6{,}644 \approx 7$) с равными интервалами $h = 1{,}5$ млн. руб.

Ряды распределения (вариационные ряды) могут быть построены по самым различным объектам. Так, объектом распределения могут служить и временные периоды (месяцы, годы), и территориальные единицы.

Для наглядности вариационные ряды изображают графически с помощью **полигона** (преимущественно дискретные ряды) и **гистограммы** (интервальные ряды). Нужно хорошо разобраться по учебникам в построении полигона и гистограммы для рядов с равными и неравными интервалами. Для вариационных рядов с неравными интервалами приобретает особое значение расчет **плотности** распределения (частное от деления частоты или частости каждого интервала на его величину). Необходимо обратить внимание на построение диаграммы кумулятивного ряда (по накопленным частотам) в форме **огивы** и **кумуляты**.

Наряду с распределением числа единиц совокупности по какому-либо признаку представляет интерес и изучение распределения по этим же группам определенного **суммарного показателя**. Последнее может быть равномерным (т.е. соответствующим распределению числа единиц) и неравномерным.

Для изучения степени неравномерности распределения определенного суммарного показателя между единицами отдельных групп вариационного ряда в статистике могут быть использованы кривая Лоренца (или кривая концентрации) и рассчитанный на ее основе коэффициент Джини (G).

Рассмотрим их построение на конкретном примере.

<u>Задача 2.3</u>

Пусть имеется следующее распределение городов по числу жителей и распределение населения в этих городах в одном из государств (графы 1, 2, 3):

Города с числом жителей, тыс. чел.	Число городов, % к итогу w_i	Численность населения, % к итогу y_i	Кумулятивные итоги	
			% городов $cum\ w_i$ (или p_i)	% населения $cum\ y_i$ (или q_i)
1	2	3	4	5
До 3	4,2	0,2	4,2	0,2
3—5	4,6	0,2	8,8	0,5
5—10	13,1	1,7	21,9	2,2
10—20	28,3	6,8	50,2	9,0
20—50	28,7	14,8	78,9	23,8
50—100	9,7	10,3	88,6	34,1
100—500	9,7	33,8	98,3	67,9
Свыше 500	1,7	32,1	100,0	100,0
Итого	**100,0**	**100,0**	—	—

В графах 4 и 5 таблицы рассчитаны кумулятивные итоги процентов городов и населения в них.

Чтобы графически показать неравномерность распределения населения по отдельным группам городов, строим квадрат 100 × 100 и на оси абсцисс откладываем значения кумулятивных итогов процента городов, а на оси ординат — значения кумулятивных итогов процента численности населения в них. Для каждой пары значений кумулятивных итогов находим точку пересечения на графике, проводя перпендикуляры к осям. Затем по точкам пересечения перпендикуляров к осям вычерчиваем кривую, которая и носит название **кривой Лоренца** (рис. 1).

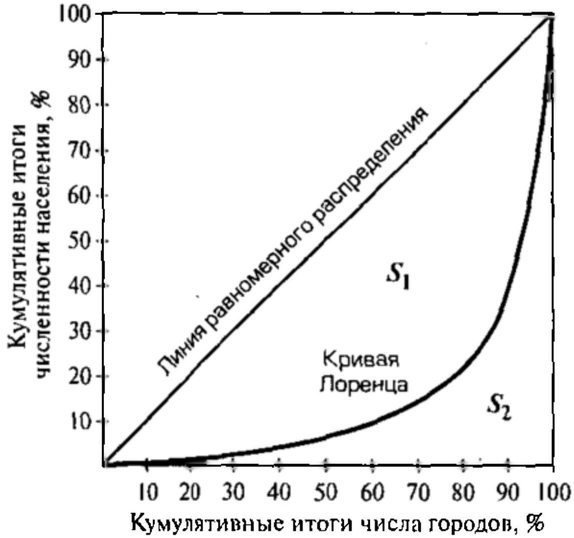

Рис. 1. Кривая Лоренца

Если бы каждому проценту накопленных (кумулятивных) частостей городов соответствовал такой же процент населения в них, то все точки расположились бы по диагонали квадрата и это означало бы равномерное распределение населения по выделенным группам городов. Естественно, чем больше фактическое распределение двух показателей отклоняется от равномерного, тем больше кривая Лоренца удалена от диагонали. Следовательно, чем больше это удаление (вогнутость), тем выше концентрация изучаемого показателя (в нашем примере численности населения) в определенных группах единиц (в нашем примере в крупных городах).

Если значения признака в группах вариационного ряда даны в порядке убывания (от большего к меньшему), то построенная по таким данным кривая Лоренца расположена выше диагонали в форме выпуклости.

Несколько кривых Лоренца, построенных на одном квадрате, позволяют сравнивать уровень концентрации изучаемого показателя в разное время или по разным объектам.

Для количественного измерения степени концентрации имеется ряд показателей. Наиболее часто используется для этой цели так называемый **коэффициент Джини** (G). По своей сути он представляет собой отношение площади (S_1), ограниченной линией равномерного распределения (диагональю квадрата) и кривой Лоренца, к половине площади квадрата ($S_1 + S_2$), т.е. $G = \dfrac{S_1}{S_1 + S_2}$. Если площадь ($S_1 + S_2$) принять за 0,5, то тогда $S_1 = 0{,}5 - S_2$, а $G = \dfrac{0{,}5 - S_2}{0{,}5} = 1 - 2S_2$.

Это отношение можно определить приближенно по формуле

$$G = \sum cum\ w_i\ cum\ y_{i+1} - \sum cum\ w_{i+1}\ cum\ y_i\ ,$$

где $cum\ w_i$ — кумулятивные доли единиц распределения;

$cum\ y_i$ — кумулятивные доли суммарного показателя.

В нашем примере коэффициент Джини

$G = (0{,}042 \cdot 0{,}005 + 0{,}088 \cdot 0{,}022 + 0{,}219 \cdot 0{,}09 + 0{,}502 \cdot 0{,}238 + 0{,}789 \cdot 0{,}341 + 0{,}886 \cdot 0{,}679 + 0{,}983 \cdot 1{,}0) - (0{,}088 \cdot 0{,}002 + 0{,}219 \cdot 0{,}005 + 0{,}502 \cdot 0{,}022 + 0{,}789 \cdot 0{,}09 + 0{,}886 \cdot 0{,}238 + 0{,}983 \cdot 0{,}341 + 1{,}0 \cdot 0{,}679) = 1{,}994975 - 1{,}299396 = 0{,}695579 \approx 0{,}7$.

(Если пользоваться в расчетах не кумулятивными долями, а процентами, то результат вычисления надо разделить на 10 000.)

В последнее время в литературе вместо символа $cum\ w_i$ *используется* p_i, а вместо $cum\ y_i$ — символ q_i и тогда

$$G = \sum p_i q_{i+1} - \sum p_{i+1} q_i\,. \qquad (1)$$

Коэффициент Джини можно выразить и другой формулой. Так, площадь S_2 можно представить как сумму площадей одного треугольника и нескольких прямоугольных трапеций, образованных кривой Лоренца, с основаниями w_i и сторонами q_{i-1} и q_i (см. рис. 1), т.е. приближенно площадь S_2 можно определить как

$$S_2 = \frac{1}{2}\sum_{i=1}^{n} w_i (q_{i-1} + q_i),$$

где $q_0 = 0$, а $q_n = 1$.

Тогда коэффициент Джини выразится следующей формулой:

$$G = 1 - 2S_2 = 1 - \sum w_i(q_{i-1} + q_i). \qquad (2)$$

Используя эту формулу, получим в нашем примере

$G = 1 - [0{,}042 \cdot 0{,}002 + 0{,}046(0{,}002 + 0{,}005) + 0{,}131(0{,}005 + 0{,}022) + 0{,}283(0{,}022 + 0{,}09) + 0{,}287(0{,}09 + 0{,}238) + 0{,}097(0{,}238 + 0{,}341) + 0{,}097(0{,}341 + 0{,}679) + 0{,}017\,(0{,}679 + 1)] = 1 - 0{,}313 = 0{,}697 \approx 0{,}7$, т.е. результат тот же, что и при расчете по формуле (1).

Значение коэффициента Джини $G = 0{,}7$ свидетельствует о сильной концентрации населения в крупных городах.

Задача 2.4

Имеются следующие данные о распределении общего объема денежных доходов населения РФ в январе–июле 1997 г.:

Группы населения по уровню доходов	Численность населения, % к итогу w_i	Доля общего объема доходов по группам, % к итогу	Накопленные итоги, %	
			доли численности населения p_i	доли общего объема доходов q_i
1	2	3	4	5
Первая (с наименьшими доходами)	20	6,2	20	6,2
Вторая	20	10,4	40	16,6
Третья	20	15,6	60	32,2
Четвертая	20	22,5	80	54,7
Пятая (с наивысшими доходами)	20	45,3	100	100,0
Итого	**100**	**100**	—	—

1. Построить кривую Лоренца для распределения общего объема денежных доходов населения по выделенным группам.
2. Рассчитать коэффициент концентрации доходов — коэффициент Джини.

Решение. 1. Для построения кривой Лоренца и расчета коэффициента Джини (G) в графах 4 и 5 рассчитаем накопленные (**кумулятивные**) итоги (в процентах) доли численности населения (p_i) и доли общего объема денежных доходов (q_i) по группам (нарастающим итогом).

Построив квадрат 100×100 со шкалой p_i по горизонтали и q_i по вертикали, находим точки пересечения для отдельных значений p_i и q_i и, плавно соединив эти точки, получим кривую Лоренца (рис. 2).

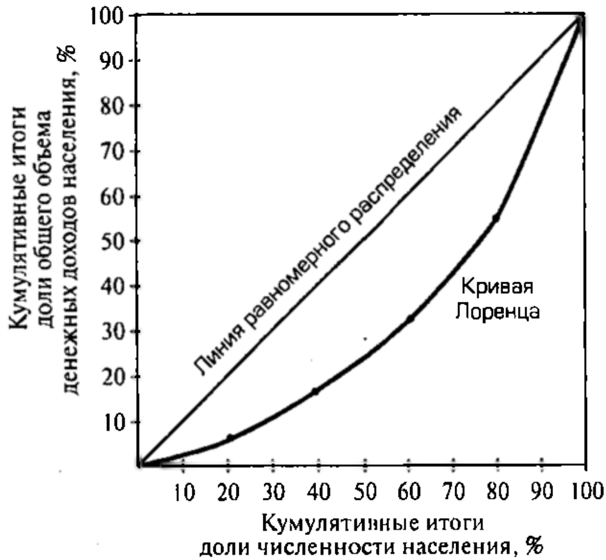

Рис. 2. Кривая Лоренца для распределения денежных доходов населения

2. Для определения коэффициента Джини, характеризующего степень концентрации или неравномерности распределения доходов между отдельными группами населения, используем формулу

$$G = \sum p_i q_{i+1} - \sum p_{i+1} q_i .$$

Так как p_i и q_i выражены в процентах, то результат надо разделить на 10 000. В нашем примере

$G = [(20 \cdot 16{,}6 + 40 \cdot 32{,}2 + 60 \cdot 54{,}7 + 80 \cdot 100) - (40 \cdot 6{,}2 + 60 \cdot 16{,}6 + 80 \cdot 32{,}2 + 100 \cdot 54{,}7)] : 10\,000 = (12\,902 - 9290) : 10\,000 = 0{,}3612.$

Такой же результат получим, используя формулу (2):

$G = 1 - \sum w_i (q_{i-1} + q_i) = 1 - (20 \cdot 6{,}2 + 20 \cdot 22{,}8 + 20 \cdot 48{,}8 + 20 \cdot 86{,}9 + 20 \cdot 154{,}7) : 10\,000 = 1 - 0{,}6388 = 0{,}3612.$

Чем ближе значение коэффициента Джини к единице, тем больше степень концентрации изучаемого суммарного показателя в отдельных группах единиц совокупности или степень неравномерности распределения.

В нашем примере $G = 0{,}36$ характеризует среднюю степень неравномерности распределения доходов.

Следует иметь в виду, что для выпуклой кривой Лоренца, расположенной выше линии равномерного распределения, т.е. характеризующей концентрацию исследуемого суммарного показателя в первых группах единиц совокупности, формулы коэффициента Джини несколько видоизменяются; их составляющие как бы меняются местами. Так, первая формула имеет вид

$$G = \sum p_{i+1} q_i - \sum p_i q_{i+1},$$

а вторая соответственно –

$$G = \sum w_i (q_{i-1} + q_i) - 1.$$

Литература

[1, с. 36–50, 68–77, 107–113], [2, с. 65–95].

Контрольные вопросы

1. В чем суть и каково значение сводки как второго этапа статистического исследования?
2. Что такое централизованная и децентрализованная сводка?
3. Какова роль группировок в статистике?
4. Какие группировки называют простыми и какие комбинационными?
5. Что такое вторичная группировка?
6. Что такое ряды распределения?
7. Что такое вариационный ряд?
8. Как строятся вариационные ряды по дискретному и непрерывному признакам?
9. Что такое полигон и гистограмма?
10. Как строятся кумулята и огива?
11. Что представляет собой кривая Лоренца и каково ее значение в анализе?
12. Как рассчитывается и что характеризует коэффициент Джини?

Тема 3
СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

Результаты обработки и сводки статистических данных, как правило, оформляются в **статистические таблицы**. В каждой таблице есть свое подлежащее и сказуемое. **Подлежащее** — это совокупность или ее части, которые подвергаются характеристике. **Сказуемое** — все показатели, характеризующие подлежащее. От студентов, работающих над обработкой материалов, собранных для различного рода экономических исследований, требуется умение строить простые, групповые и комбинационные таблицы, таблицы с простой и сложной разработкой сказуемого. Примером **простой** таблицы (в подлежащем — перечень отдельных отраслей) может служить следующая:

Показатели забастовок в РФ по отраслям в 1994 г.

Отрасль	Число предприятий, на которых прошли забастовки	Число участников, тыс. чел.	Потери рабочего времени, тыс. чел.-дней
Промышленность	209	115,0	557,0
Образование	279	36,1	165,0
Строительство	8	1,3	5,3
Транспорт и связь	8	1,3	3,0
Прочие	10	1,6	4,8

К числу простых относятся и таблицы, характеризующие динамику тех или иных показателей, как, например, следующая:

Производство молока и яиц на сельскохозяйственных предприятиях РФ в 1994 г.

Месяц	Янв.	Февр.	Март	Апр.	Май	Июнь	Июль	Авг.
Валовой надой молока, млн. т	1,5	1,5	2,0	2,2	2,6	3,3	3,1	2,7
Производство яиц, млрд. шт.	2,3	2,2	2,3	2,4	2,3	2,2	2,2	2,1

Если в подлежащем таблицы имеется группировка единиц изучаемой совокупности по одному признаку, то такая таблица называется **групповой**, например макет таблицы с группировкой хозяйств области по размеру посевных площадей:

Группы хозяйств по размеру посевной площади, га	Число хозяйств	Общая посевная площадь, га	Валовой сбор зерновых, ц	Средняя урожайность, ц/га
До 5000				
5000—10 000				
10 000—25 000				
25 000—50 000				
Свыше 50 000				
Итого				

В подлежащем **комбинационной** таблицы содержится группировка по двум или нескольким признакам, например группировка хозяйств по размеру посевной площади и по числу коров.

Сказуемое в таблицах тоже может быть разработано по-разному. Различают простую и сложную разработки сказуемого. При **простой** разработке сказуемого все его показатели располагаются независимо друг от друга, например дается характеристика студентов-заочников института:

Курс	Численность студентов				
	Всего	мужчин	женщин	москвичей	иногородних
I II III IV V					
Всего					

При **сложной разработке сказуемого** показатели сказуемого сочетаются друг с другом, как, например, в приведенном ниже макете таблицы:

Курс	Численность студентов								
	москвичей			иногородних			Всего		
	Всего	муж.	жен.	Всего	муж.	жен.	Всего	муж.	жен.
I II III IV V									
Всего									

Сложная разработка сказуемого дана и в приводимой ниже таблице.

Оценка экономического состояния хозяйств РФ в 1998 г.
(по материалам выборочного обследования 559 сельскохозяйственных предприятий (2%) и 493 крестьянских (фермерских) хозяйств (0,2%), проведенного в мае 1999 г.; в процентах от числа обследованных хозяйств)

Вид хозяйства	Растениеводство			Животноводство			Хозяйство в целом		
	Удовлетворительное	Неудовлетворительное	На грани банкротства	Удовлетворительное	Неудовлетворительное	На грани банкротства	Удовлетворительное	Неудовлетворительное	На грани банкротства
Сельскохозяйственные предприятия	43	48	6	37	49	9	40	50	10
Крестьянские (фермерские) хозяйства	56	31	5	34	15	3	61	34	5

Следует помнить, что при построении любой статистической таблицы нужно исходить из цели исследования и содержания обрабатываемого материала.

Литература

[1, с. 46—50], [2, с. 103—116], [3, с. 53—57].

Контрольные вопросы

1. Что представляет собой статистическая таблица и каковы ее составные элементы?
2. Что такое подлежащее и сказуемое таблицы?
3. Что такое простая и сложная разработка сказуемого?
4. Назовите виды статистических таблиц (по характеру подлежащего).

Задачи и упражнения для самостоятельной работы (по темам 2 и 3)

Задача 1

Ниже приведены данные о количестве членов семьи в 50 обследованных фермерских хозяйствах:

2	5	5	6	3	2	5	6	5	6
6	6	4	3	3	5	7	3	5	5
5	4	5	6	4	4	4	4	7	4
4	3	5	3	7	4	6	6	4	7
4	4	6	7	6	3	3	5	8	5

1. Построить дискретный вариационный ряд — распределение 50 хозяйств по количеству членов семьи.
2. Изобразить ряд графически с помощью полигона распределения.

Задача 2

Имеются следующие данные об урожайности озимой пшеницы в 40 обследованных хозяйствах:

27,1	18,2	16,3	22	24,3	24,8	33,0	27,3
28,5	15,1	19,5	28,1	25,1	26,7	28,4	29,6
23,7	18,0	31,0	19,8	26,0	23,5	20,2	25,1
22,8	27,0	20,4	24	29,5	22,9	19,9	27,0
25,3	23,9	21,5	23,1	21,1	22,6	25,8	23,8

1. Определить размах вариации урожайности ($x_{max} - x_{min}$).
2. Построить интервальный вариационный ряд с равными интервалами, выделив 6 групп хозяйств по величине урожайности.

3. Изобразить ряд графически с помощью гистограммы распределения.
4. По накопленным частотам построить кумуляту распределения 40 хозяйств по величине урожайности.

Задача 3

Определить (назвать) вид приводимых ниже таблиц, их подлежащее и сказуемое.

а) **Группировка промышленных предприятий РФ по объему продукции в 1996 г.**

Группы предприятий с объемом продукции, млрд. руб.	Число предприятий, % к итогу	Объем продукции, % к итогу	Среднегодовая численность промышленно-производственного персонала, % к итогу
До 1	84,0	3,6	13,1
1,1—50	14,2	16,2	34,4
50,1—100	1,7	37,3	36,9
Свыше 100	0,1	42,9	15,6
Всего	**100,0**	**100**	**100**

б) **Распределение численности безработных в РФ по возрастным группам (на конец октября 1996 г.)**

Возрастные группы, лет	Численность безработных, % к итогу	
	мужчины	женщины
16—19	7,7	11,2
20—24	17,0	18,5
25—29	11,9	11,7
30—49	50,9	49,5
50—54	4,2	4,0
55—59	5,7	3,8
60—72	2,6	1,3
Итого	**100,0**	**100,0**

в) **Основные показатели промышленности РФ за 1990—1996 гг.**

Год	1990	1991	1992	1993	1994	1995	1996
Число предприятий, тыс.	26,9	28,0	61,1	104	138	137	156
Численность промышленно-производственного персонала, млн. чел.	21,0	20,1	20,0	18,8	17,4	16,0	14,9
Уровень рентабельности продукции, %	12,0	27,1	38,3	32,0	19,5	20,1	9,2

Задача 4

1. *По данным таблицы «а» из задачи 3 построить кривую Лоренца для числа предприятий (%) и объема продукции (%), а также рассчитать коэффициент Джини, характеризующий степень концентрации объема продукции по выделенным группам предприятий.*
2. То же самое проделать для числа предприятий (%) и численности промышленно-производственного персонала.

Ответ: 1) 0,922; 2) 0,780.

Задача 5

Имеются следующие данные по предприятиям промышленности РФ за 1996 г.:

Группы предприятий с численностью промышленно-производственного персонала, чел.	Число предприятий, % к итогу	Объем продукции, % к итогу
До 200	93,6	12,7
201—500	3,4	9,6
Свыше 500	3,0	77,7
Итого	**100**	**100**

1. Построить кривую Лоренца для распределения объема продукции по группам предприятий с разной численностью персонала.
2. С помощью коэффициента Джини определить уровень (степень) концентрации объема продукции по выделенным группам предприятий.

Ответ: 2) 0,833.

Задача 6

Имеется следующее распределение регионов РФ по объему валового регионального продукта (ВРП) в 1994 г.:

Группы регионов, упорядоченных по объему ВРП	Число регионов	Общий объем ВРП по группам, % к итогу
Первая (с наибольшими объемами)	10	45,99
Вторая	10	19,61
Третья	10	11,55
Четвертая	10	8,29
Пятая	10	6,16
Шестая	10	4,40
Седьмая (с наименьшими объемами)	19	4,00
Итого	**79**	**100,00**

Построить кривую Лоренца для объема ВРП и с помощью коэффициента Джини охарактеризовать степень концентрации объема ВРП в первых группах регионов.

Задача 7

Имеется следующее распределение общего объема денежных доходов по 20-процентным группам населения РФ в I полугодии 2000 г.:

Группы населения	Численность населения, %	Общий объем денежных доходов, %
Первая (с наименьшими доходами)	20	6,0
Вторая	20	10,1
Третья	20	14,5
Четвертая	20	20,8
Пятая (с наивысшими доходами)	20	48,6
Итого	**100**	**100,0**

1. Построить кривую Лоренца по данным таблицы.
2. С помощью коэффициента Джини измерить концентрацию доходов населения.

Тема 4
ОТНОСИТЕЛЬНЫЕ И СРЕДНИЕ ВЕЛИЧИНЫ

Относительные и средние величины — основные обобщающие показатели, используемые при анализе статистических данных.

Изучение относительных величин обычно не вызывает у студентов затруднений. Надо уметь в каждом отдельном случае правильно выбрать форму относительной величины (простое кратное отношение, процентное, промильное или продецимильное выражение); следует помнить о необходимости соблюдения условия сопоставимости данных, на основе которых рассчитываются относительные величины выполнения плана, динамики, сравнения.

Знакомясь со средними величинами, следует помнить, что средние должны рассчитываться лишь для качественно однородных совокупностей. Кроме того, в зависимости от исходных данных средние значения тех или иных признаков могут рассчитываться по-разному. Очень часто среднее значение какого-либо показателя вычисляется в статистике на основе итоговых показателей, рассчитанных для совокупности. Так, например, зная валовой сбор и посевную площадь под зерновыми по области, легко определить среднюю урожайность зерновых, разделив валовой сбор на посевную площадь. Если же известны значения признака у отдельных единиц совокупности, то осредненный показатель может быть рассчитан как средняя из отдельных вариантов по одной из формул различных видов средних величин, в одних случаях — как средняя арифметическая (простая или взвешенная), в других — как средняя гармоническая (простая или взвешенная), в третьих — как средняя геометрическая и т.д.

Из средних величин наиболее часто встречаются средняя арифметическая простая $\bar{x} = \frac{\sum x_i}{n}$ и средняя арифметическая взвешенная $\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$, где x_i — отдельные значения признака, варианты; f_i — веса каждого варианта. Взвешенная применяется в тех случаях, когда отдельные значения признаков повторяются. Если вместо абсолютных частот в распределении имеются частости (w_i), выступающие в роли весов, то тогда $\bar{x} = \sum x_i w_i$ (если w_i выражены в долях, $\sum w_i = 1$) или $\bar{x} = \frac{\sum x_i w_i}{\sum w_i}$ (если w_i выражены в процентах, $\sum w_i = 100$).

Рассмотрим ряд примеров.

<u>Задача 4.1</u>

Пусть имеются следующие данные о производстве продукта A рабочими бригады за смену:

Номер рабочего	1	2	3	4	5
Произведено продукции A за смену, шт., x_i	21	18	20	22	19

Нужно определить среднюю выработку одного рабочего данной бригады.

Последняя определяется как средняя арифметическая простая:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{21 + 18 + 20 + 22 + 19}{5} = 20 \text{ (шт.)}.$$

Задача 4.2

Имеется следующее распределение 60 рабочих по тарифному разряду:

Тарифный разряд x_i	2	3	4	5	6
Число рабочих f_i	8	16	17	12	7

Нужно определить средний тарифный разряд рабочих. Расчет производим по средней арифметической взвешенной:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{2 \cdot 8 + 3 \cdot 16 + 4 \cdot 17 + 5 \cdot 12 + 6 \cdot 7}{8 + 16 + 17 + 12 + 7} = \frac{234}{60} = 3{,}9.$$

Для интервальных рядов сначала находят центры (середины) интервалов, а затем последние умножают на веса, произведения суммируют и делят на сумму весов.

Задача 4.3

Требуется определить среднемесячную заработную плату одного рабочего по следующим данным (графы 1 и 2 таблицы):

Месячная заработная плата, руб.	Число рабочих f_i	Середина интервала x_i	$x_i f_i$
1	2	3	4
2400—2500	10	2450	24 500
2500—2600	20	2550	51 000
2600—2700	48	2650	127 200
2700—2800	60	2750	165 000
2800—2900	42	2850	119 700
2900—3000	20	2950	59 000
Итого	**200**	—	**546 400**

Среднее значение признака (x_i) в каждом интервале дано в графе 3. Результаты умножения вариантов (x_i) на веса (f_i) показаны в графе 4.

Отсюда средняя заработная плата одного рабочего

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{546\,400}{200} = 2732 \text{ (руб.)}.$$

Примечание. В целях упрощения вычислений при больших значениях x_i последние можно уменьшать, помня, что всякое уменьшение вариантов (x_i) влечет за собой соответствующее уменьшение средней арифметической и что поэтому после вычисления средней арифметической из уменьшенных вариантов следует увеличить ее «во столько раз» и «на столько», «во сколько» и «на сколько» уменьшались варианты. Этот прием расчета покажем на примере дальше (см. задачу 5.4).

Когда весами у отдельных признаков служат показатели, являющиеся произведением этих вариантов на количество единиц, то средняя из всех вариантов рассчитывается как средняя гармоническая взвешенная, формула которой

$$\bar{x} = \frac{\sum M_i}{\sum \frac{M_i}{x_i}}.$$

Задача 4.4

Предположим, по пяти хозяйствам имеются данные об урожайности зерновых и валовом сборе:

Хозяйство	Урожайность зерновых, ц/га, x_i	Валовой сбор зерна, ц, M_i
1	18	18 000
2	20	30 000
3	21	63 000
4	22	44 000
5	25	30 000
$\sum$	—	**185 000**

Нужно рассчитать среднюю урожайность для всех хозяйств. Для решения задачи следует валовой сбор всех хозяйств разделить на общую площадь. Площадь по каждому хозяйству неизвестна, но легко может быть рассчитана путем деления валового сбора на урожайность. Произведя последовательно все расчеты, получим среднюю урожайность:

$$\bar{x} = \frac{18\,000 + 30\,000 + 63\,000 + 44\,000 + 30\,000}{\frac{18\,000}{18} + \frac{30\,000}{20} + \frac{63\,000}{21} + \frac{44\,000}{22} + \frac{30\,000}{25}} =$$

$$= \frac{185\,000}{1000 + 1500 + 3000 + 2000 + 1200} = \frac{185\,000}{8700} = 21{,}26 \text{ (ц/га)}.$$

Если урожайность обозначить через x_i, а валовой сбор через M_i, то нетрудно убедиться, что расчет средней урожайности произведен по средней гармонической взвешенной $\bar{x} = \frac{\sum M_i}{\sum \frac{M_i}{x_i}}$. Если бы валовой сбор был везде одинаков, то значения весов в числителе и знаменателе сократились бы и можно было бы воспользоваться средней гармонической простой:

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

Кроме средней арифметической и средней гармонической в статистике применяют и другие виды средних (геометрическая, квадратическая и др.). Их использование будет показано в дальнейших темах.

Литература

[1, с. 51—67, 77—87], [2, с. 159—176], [3, с. 69—78].

Контрольные вопросы

1. Какова роль относительных величин в статистике?
2. Какие существуют формы выражения относительных величин?
3. Как классифицируются относительные величины по их познавательному значению?
4. Каково значение средних величин в статистике?
5. Какие виды средних величин применяются в статистике?
6. Как исчисляются средние арифметические: простая и взвешенная?
7. В каких случаях применяется средняя гармоническая?

Тема 5
МОДА, МЕДИАНА И ДРУГИЕ ХАРАКТЕРИСТИКИ ВАРИАЦИОННОГО РЯДА

В статистике недостаточно знать лишь среднюю величину того или иного признака у единиц совокупности. Большой интерес при статистическом исследовании различных совокупностей представляет изучение вариации (колеблемости) признака у отдельных единиц и характера распределения единиц по данному признаку.

Вариации признака и характер распределения изучаются прежде всего с помощью некоторых характеристик вариационного ряда, из которых рассмотренная ранее средняя арифметическая является основной, характеризующей центр группирования. Другими характеристиками центра группирования являются мода (Mo) и медиана (Me).

Мода (наиболее часто встречающееся значение признака у единиц совокупности) для дискретного ряда определяется непосредственно как вариант (x), имеющий наибольшую частоту или частость (в вариационном ряду задачи 4.2 из предыдущей темы Mo равняется 4). Для интервального ряда с равными интервалами мода рассчитывается по формуле

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

где $x_{н}$ — начальная (нижняя) граница модального интервала;
h — величина интервала;
f_2 — частота модального интервала;
f_1 — частота интервала, предшествующего модальному;
f_3 — частота интервала, следующего за модальным.

<u>Задача 5.1</u>

Определить моду по данным задачи 4.3. В условии задачи 4.3 наибольшую частоту (60) имеет интервал (2700—2800). Отсюда

$$Mo = 2700 + 100\frac{60-48}{(60-48) + (60-42)} = 2700 + 100\cdot\frac{12}{38} = 2731{,}6 \text{ (руб.)}.$$

т.е. наиболее часто встречается заработная плата в размере 2731,6 руб.

В ряду с неравными интервалами мода определяется в интервале, имеющем наибольшую плотность распределения, и в формуле вместо f_1, f_2, f_3 принимаются соответствующие плотности распределения.

Для нахождения **медианы** (значения признака у средней единицы ранжированного ряда) сначала определяется ее порядковый номер $\left(\frac{\sum f}{2}\right)$, а затем по накопленным частотам определяется либо сама медиана (для дискретных рядов), либо медианный интервал (для интервальных рядов), в котором путем простой интерполяции рассчитывается значение медианы по формуле

$$Me = x_н + h\frac{\frac{\sum f}{2} - S_{Me-1}}{f_{Me}},$$

где $x_н$ — нижняя граница медианного интервала;

$\frac{\sum f}{2}$ — порядковый номер медианы;

S_{Me-1} — накопленная частота до медианного интервала;

f_{Me} — частота медианного интервала.

Задача 5.2

Для приведенного в задаче 4.3 распределения рабочих по размеру заработной платы определить медиану.

Перепишем этот ряд и рассчитаем в нем накопленные частоты:

Месячная заработная плата, руб.	Число рабочих f_i	Накопленные частоты S
2400—2500	10	10
2500—2600	20	30
2600—2700	48	78
2700—2800	60	138
2800—2900	42	180
2900—3000	20	200
Итого	**200**	—

Определяем порядковый номер медианы:

$$N_{Me} = \frac{\sum f}{2} = \frac{200}{2} = 100.$$

По накопленным частотам видно, что сотая единица находится в интервале (2700—2800), ее значение определяем по формуле

$$Me = 2700 + 100\frac{100 - 78}{60} = 2736{,}7 \text{ (руб.)},$$

т.е. делаем вывод по медиане, что половина рабочих получает заработную плату ниже 2736,7 руб., а половина — выше.

Мода и медиана могут быть определены графически: первая — по гистограмме, а вторая — по кумуляте.

Рассмотрим это на примере.

Построим гистограмму распределения 200 рабочих по размеру заработной платы (рис. 3), для чего на оси абсцисс построим ряд сомкнутых прямоугольников, у каждого из которых основанием служит величина интервала признака (размер заработной платы в рублях), а высотой — частота каждого интервала (число рабочих). (Для рядов с неравными интервалами в качестве высоты прямоугольников принимается плотность распределения.)

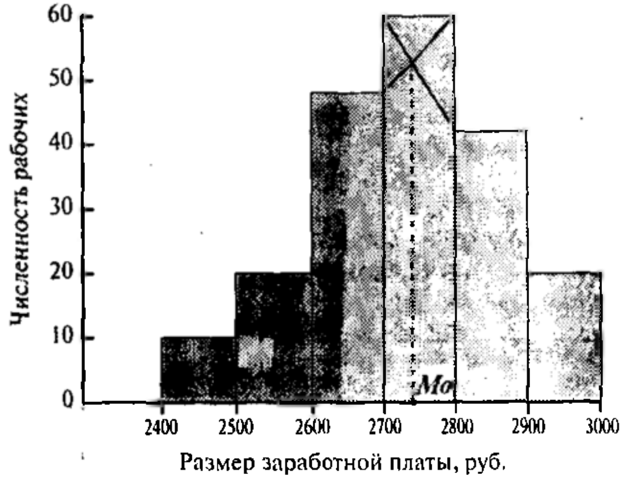

Рис. 3. Гистограмма распределения 200 рабочих по размеру заработной платы (графическое определение моды)

В прямоугольнике, имеющем наибольшую высоту, проводим две линии, как показано на рис. 3, и из точки их пересечения опускаем перпендикуляр на ось абсцисс. Значение *x на оси абсцисс в этой* точке есть мода (*Mo*).

Для графического отыскания медианы по накопленным частотам строим **кумуляту** (рис. 4). Для этого из верхней границы каждого *интервала на оси абсцисс восстанавливаем перпендикуляр*, соответствующий по высоте накопленной частоте с начала ряда по данный *интервал. Соединив последовательно* вершины перпендикуляров, мы и получим кривую, называемую кумулятой. Из точки на оси орди-

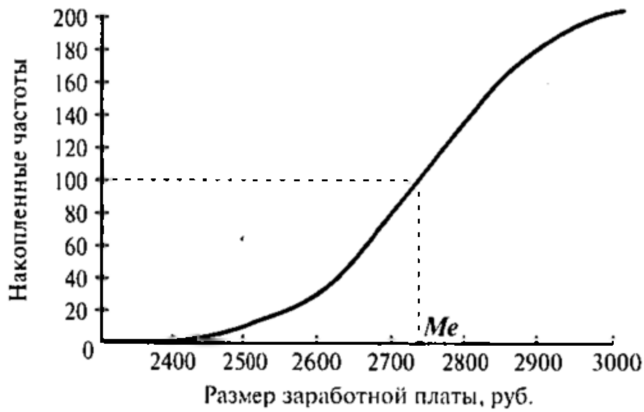

Рис. 4. Кумулята распределения 200 рабочих по размеру заработной платы (графическое определение медианы)

нат, соответствующей половине всех частот (порядковому номеру медианы), проводим прямую, параллельную оси абсцисс, до пересечения ее с кумулятой. Опустив из этой точки перпендикуляр на ось абсцисс, находим значение медианы (*Me*).

Пользуясь кумулятой, можно определить значение признака у любой единицы ранжированного ряда.

Аналогично рассчитываются показатели, именуемые **квартилями**. Первая квартиль (Q_1) — значение признака у единицы, делящей ранжированный ряд в соотношении $^1/_4$ и $^3/_4$, вторая квартиль равна медиане ($Q_2 = Me$), третья квартиль (Q_3) — значение признака у единицы, делящей ранжированный ряд в соотношении $^3/_4$ и $^1/_4$. Порядковый номер Q_1 определяется как $\sum f/4$, для Q_3 — соответственно как $^3/_4 \sum f$.

Представляет интерес и расчет показателей, именуемых **децилями** (значение признака у единицы, делящей ранжированный ряд в соотношении $^1/_{10}$ и $^9/_{10}$ (первая дециль — D_1), $^2/_{10}$ и $^8/_{10}$ (вторая дециль — D_2) и т.д.).

Для симметричных распределений характерно совпадение значений средней арифметической, моды и медианы. Если $Mo > \bar{x}$, то ряд будет иметь левостороннюю асимметрию (вытянутость), а если $Mo < \bar{x}$, то правостороннюю асимметрию. В умеренно асимметричных рядах соотношение между указанными показателями выражается следующим образом:

$$|Mo - \bar{x}| \leqslant 3\,|Me - \bar{x}|.$$

ПОКАЗАТЕЛИ ВАРИАЦИИ

Для суждения о вариации признака в статистике используют ряд показателей.

1. **Среднее линейное отклонение** $(\bar{d})$:

$$\bar{d} = \frac{\sum |x_i - \bar{x}|}{n} \quad \text{(для несгруппированных данных)}$$

или

$$\bar{d} = \frac{\sum |x_i - \bar{x}| f_i}{\sum f_i} \quad \text{(для вариационного ряда),}$$

где $|x_i - \bar{x}|$ — абсолютные значения отклонений отдельных вариантов (x_i) от средней арифметической $(\bar{x})$.

2. **Среднее квадратическое отклонение** (σ) — наиболее распространенный и применяемый показатель вариации, рассчитываемый по формуле

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \quad \text{(для несгруппированных данных)}$$

или

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}} \quad \text{(для вариационного ряда).}$$

Порядок расчета среднего квадратического отклонения по основной формуле таков:

1) находим среднюю арифметическую ряда $(\bar{x})$;
2) находим отклонение каждого варианта от средней арифметической $(x_i - \bar{x})$;
3) возводим каждое отклонение в квадрат: $(x_i - \bar{x})^2$;
4) умножаем каждый квадрат отклонений на соответствующие веса: $(x_i - \bar{x})^2 f_i$;
5) суммируем все произведения: $\sum (x_i - \bar{x})^2 f_i$;

6) делим сумму произведений на сумму весов (частот)

$$\frac{\sum\left(x_i-\overline{x}\right)^2 f_i}{\sum f_i}$$

и получаем дисперсию (σ^2);

7) извлекаем квадратный корень из частного, тогда

$$\sigma=\sqrt{\frac{\sum\left(x_i-\overline{x}\right)^2 f_i}{\sum f_i}}.$$

Полученный результат и представляет собой среднее квадратическое отклонение (σ).

Среднее квадратическое отклонение в квадрате, т.е. σ^2, называется **дисперсией**:

$$\sigma^2=\frac{\sum\left(x_i-\overline{x}\right)^2}{n} \quad \text{или} \quad \sigma^2=\frac{\sum\left(x_i-\overline{x}\right)^2 f_i}{\sum f_i}. \tag{3}$$

Практически для расчета дисперсии иногда удобнее пользоваться другими формулами, непосредственно вытекающими из приведенной выше. Так, если возвести в квадрат правую часть равенства и почленно разделить на $\sum f$, то получим формулу дисперсии в следующем виде:

$$\sigma^2=\overline{x^2}-\left(\overline{x}\right)^2, \tag{4}$$

где $\overline{x^2}$ — средняя величина квадратов вариантов (x);

$\left(\overline{x}\right)^2$ — квадрат средней арифметической $\left(\overline{x}\right)$.

Дисперсия, рассчитанная относительно средней арифметической величины $\left(\overline{x}\right)$, минимальна. Но порой удобнее рассчитывать средний квадрат отклонений от произвольного числа a (т.е. «дисперсию от a») по формуле

$$\sigma_a^2=\frac{\sum\left(x_i-a\right)^2 f_i}{\sum f_i},$$

а затем от нее переходить к искомой σ^2. Для этого введем в числитель записанной выше формулы σ_a^2 величину $\overline{x}$ со знаком «+» и «—» и выполним следующие несложные алгебраические преобразования:

$$\sigma_a^2 = \frac{\sum\left[\left(x_i - \bar{x}\right) + \left(\bar{x} - a\right)\right]^2 f_i}{\sum f_i} = \frac{\sum\left(x_i - \bar{x}\right)^2 f_i}{\sum f_i} +$$

$$+ \frac{2\left(\bar{x} - a\right)\sum\left(x_i - \bar{x}\right) f_i}{\sum f_i} + \frac{\left(\bar{x} - a\right)^2 \sum f_i}{\sum f_i} = \sigma^2 + \left(\bar{x} - a\right)^2$$

(второе слагаемое обращается в нуль, так как $\sum\left(x_i - \bar{x}\right) f_i = 0$).

Пользуясь последней формулой, можно рассчитывать дисперсию от любого постоянного числа a, а затем корректировать эту дисперсию на квадрат разности между средней арифметической и числом a.

Для сравнения вариации в разных совокупностях рассчитывается относительный показатель вариации, именуемый коэффициентом вариации (V), представляющий собой процентное отношение среднего квадратического отклонения к средней арифметической:

$$V = \frac{\sigma}{\bar{x}} 100\%.$$

Следует помнить, что если в совокупности исследуется доля единиц, обладающих тем или иным альтернативным признаком, то дисперсия этой доли определяется по формуле $\sigma^2 = pq$, где p — доля *единиц, обладающих данным признаком*, а q — доля единиц, не обладающих данным признаком. Максимальное значение дисперсии доли равно 0,25 (когда $p = q = 0{,}5$).

Задача 5.3

Воспользовавшись исходными данными задачи 4.2, где $\bar{x} = 3{,}9$, рассчитаем основные показатели вариации. Расчет оформим в следующей таблице:

x_i	f_i	$\left\|x_i - \bar{x}\right\|$	$\left\|x_i - \bar{x}\right\| f_i$	$\left(x_i - \bar{x}\right)^2 f_i$
2	8	1,9	15,2	28,88
3	16	0,9	14,4	12,96
4	17	0,1	1,7	0,17
5	12	1,1	13,2	14,52
6	7	2,1	14,7	30,87
$\sum$	**60**	—	**59,2**	**87,40**

А. Среднее линейное отклонение

$$\bar{d} = \frac{\sum\left|x_i - \bar{x}\right| f_i}{\sum f_i} = \frac{59{,}2}{60} = 0{,}987.$$

Б. Дисперсия

$$\sigma^2 = \frac{\sum\left(x_i - \overline{x}\right)^2 f_i}{\sum f_i} = \frac{87,4}{60} = 1,46.$$

В. Среднее квадратическое отклонение

$$\sigma = \sqrt{\frac{\sum\left(x_i - \overline{x}\right)^2 f_i}{\sum f_i}} = \sqrt{1,46} = 1,21.$$

Г. Коэффициент вариации

$$V = \frac{\sigma}{\overline{x}}\,100\% = \frac{1,21}{3,9}\,100\% = 31,0\%.$$

Задача 5.4

Воспользовавшись исходными данными задачи 4.3, рассчитаем среднюю арифметическую, дисперсию и коэффициент вариации, предварительно уменьшив до предела варианты (т.е. способом расчета от условного нуля); вычтем из всех x_i значение $a = 2650$ и разделим разности на 100:

Середина интервала x_i	Частоты f_i	$x' = \frac{x_i - 2650}{100}$	$\frac{x_i - 2650}{100} f_i$	$\left(\frac{x_i - 2650}{100}\right)^2 f_i$
2450	10	−2	−20	40
2550	20	−1	−20	20
2650	48	0	0	0
2750	60	1	60	60
2850	42	2	84	168
2950	20	3	60	180
$\sum$	**200**	—	**164**	**468**

1. Средняя арифметическая

$$\overline{x} = \frac{\sum\left(\frac{x - 2650}{100}\right) f}{\sum f}\,100 + 2650 = \frac{164}{200}\,100 + 2650 = 2732 \text{ (руб.)}.$$

2. Рассчитаем дисперсию. Сначала найдем средний квадрат отклонений от произвольного числа — в нашем примере от 2650, а затем скорректируем его на квадрат разности между средней арифметической и этим произвольным числом, т.е. применим формулу

$$\sigma^2 = \sigma^2_{2650} - \left(\bar{x} - 2650\right)^2;$$

$$\sigma^2_{2650} = \frac{\sum\left(\frac{x-2650}{100}\right)^2 f}{\sum f}\,100^2 = \frac{468}{200}\,10\,000 = 23\,400;$$

$$\sigma^2 = 23\,400 - (2732 - 2650)^2 = 16\,676.$$

3. Отсюда среднее квадратическое отклонение

$$\sigma = \sqrt{\sigma^2} = \sqrt{16\,676} = 129{,}1.$$

4. Коэффициент вариации

$$V = \frac{129{,}1}{2732}\,100\% = 4{,}7\%.$$

Задача 5.5

Пусть имеются следующие данные о результатах экзаменационной сессии на I и II курсах одного из вузов: на I курсе 85% студентов сдали сессию без двоек, а на II курсе — 90%. Определить дисперсию доли студентов, успешно сдавших сессию (или, что то же самое, доли студентов, получивших двойки на сессии), на каждом курсе.

Так как $p_1 = 0{,}85$ и $q_1 = 0{,}15$, а $p_2 = 0{,}9$ и $q_2 = 0{,}1$, то:

на I курсе $\sigma^2_1 = p_1\,q_1 = 0{,}85 \cdot 0{,}15 = 0{,}1275 \Rightarrow \sigma_1 = 0{,}35$;

на II курсе $\sigma^2_2 = p_2\,q_2 = 0{,}9 \cdot 0{,}1 = 0{,}09 \Rightarrow \sigma_2 = 0{,}3$.

Следовательно, на II курсе дисперсия и среднее квадратическое отклонение доли студентов, успешно сдавших сессию, меньше, чем на I курсе.

Задача 5.6

Имеются следующие данные о распределении населения РФ по размеру среднедушевого денежного дохода в месяц в июле 1997 г. (графы 1, 2, 3 табл. 1).

Определить:

1) *среднедушевой доход с помощью средней арифметической, моды и медианы;*
2) *среднее квадратическое отклонение душевого дохода и коэффициент вариации;*
3) *децильный коэффициент дифференциации доходов;*
4) *степень концентрации (неравномерности) доходов у отдельных групп населения (коэффициент Джини).*

Решение. Необходимые для расчета суммы показаны в табл. 1.

Таблица 1

Среднедушевой доход в месяц, тыс. руб.	Численность населения		Середина интервала x_i	Суммарный доход по группам $x_i f_i$	Накопленные частости (доли населения в %) P_i	$x_i^2 f_i$
	млн. чел. f_i	% к итогу w_i				
1	2	3	4	5	6	7
До 400	25,8	17,5	300	7740	17,5	2 322 000
400—600	28,7	19,5	500	14 350	37,0	7 175 000
600—800	24,7	16,7	700	17 290	53,7	12 103 000
800—1000	19,2	13,0	900	17 280	66,7	15 552 000
1000—1200	13,8	9,4	1100	15 180	76,1	16 698 000
1200—1600	16,6	11,2	1400	23 240	87,3	32 536 000
1600—2000	8,4	5,7	1800	15 120	93,0	27 216 000
Свыше 2000	10,3	7,0	2200	22 660	100	49 852 000
Итого	**147,5**	**100,0**	—	**132 860**	—	**163 454 000**

$x_i \cdot w_i = 90075$

1. а) Среднедушевой доход

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{132\,860}{147{,}5} = 900{,}75 \text{ (тыс. руб.)}.$$

(В качестве весов можно принимать и w_i — относительный показатель численности населения в % к итогу. Результат будет тот же.)

б) Мода (наиболее часто встречающийся размер душевого дохода) находится в интервале 400—600. Определим ее как

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

$$= 400 + 200\frac{28{,}7 - 25{,}8}{(28{,}7 - 25{,}8) + (28{,}7 - 24{,}7)} =$$

$$= 400 + 200\frac{2{,}9}{2{,}9 + 4} = 484 \text{ (тыс. руб.)},$$

где $x_н$ — нижняя граница модального интервала.

в) Медиану удобнее находить по данным распределения в %, т.е. по w_i. Тогда порядковый номер медианы равен $\frac{\sum w_i}{2} = \frac{100}{2} = 50\%$. По накопленным частостям определяем, что медиана (50-й процент) находится в интервале 600—800.

Отсюда

$$Me = x_н + h\frac{\frac{\sum w_i}{2} - S_{Me-1}}{w_{Me}} = 600 + 200\,\frac{50 - 37}{16{,}7} = 755{,}7 \text{ (тыс. руб.)},$$

т.е. *в июле 1997 г. половина населения имела среднедушевой денежный доход ниже 755,7 тыс. руб., а половина — выше 755,7 тыс. руб.*

2. а) Среднее квадратическое отклонение найдем по формуле

$$\sigma = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{\frac{\sum x_i^2 f_i}{\sum f_i} - \left(\frac{\sum x_i f_i}{\sum f_i}\right)^2} =$$

$$= \sqrt{\frac{163\,454\,000}{147.5} - (900{,}75)^2} = \sqrt{1\,108\,162{,}7 - 811\,350{,}56} =$$

$$= \sqrt{296\,812{,}14} = 544{,}8 \text{ (тыс. руб.)}.$$

(Можно было использовать и формулу $\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}}$.)

б) Коэффициент вариации

$$V = \frac{\sigma}{\bar{x}}\,100\% = \frac{544{,}8}{900{,}75}\,100\% = 60\%.$$

Такая вариация весьма значительна, т.е. население далеко не однородно по размеру душевого дохода.

3. Для оценки степени дифференциации доходов по накопленным частостям (графа 6) определяем, что первая дециль (D_1) находится в первом интервале (200—400), а девятая дециль (D_9) — в предпоследнем интервале (1600—2000):

а) $$D_1 = x_н + h\,\frac{10\% - S_{D_1-1}}{w_{D_1}} = 200 + 200\,\frac{10 - 0}{17{,}5} = 314{,}3 \text{ (тыс. руб.)}.$$

D_1 — это *максимальный* размер душевого дохода у *десяти процентов населения с наименьшими доходами.*

Учитывая, что частость первого интервала (17,5%) больше 10% (порядкового номера D_1), расчет D_1 удобнее вести от значения верхней границы интервала, вычитая из него величину интервала, приходящуюся на излишние 7,5% единиц, т.е.

$$D_1 = x_в - h\,\frac{S_{D_1} - 10\%}{w_{D_1}} = 400 - 200\,\frac{17{,}5 - 10}{17{,}5} = 314{,}3 \text{ (тыс. руб.)};$$

б) $D_9 = x_n + h\dfrac{90\% - S_{D_9-1}}{w_{D_9}} = 1600 + 400\dfrac{90 - 87{,}3}{5{,}7} =$

$= 1600 + 189{,}5 = 1789{,}5$ (тыс. руб.).

D_9 — это *минимальный* размер душевого дохода у *десяти процентов населения с наиболее высокими доходами*;

в) децильный коэффициент дифференциации (*ДКД*) доходов

$$ДКД = \frac{D_9}{D_1} = \frac{1789{,}5}{314{,}3} = 5{,}7 \text{ (раза)}.$$

Примечание. Поскольку в первом интервале не указана нижняя граница, а последующий интервал равен 200 (400—600), то мы вправе и в первом интервале принимать такую же величину, т.е. (200—400).

4. Для расчета коэффициента Джини, G, сначала определяем долю (в %) суммарного дохода по каждой группе как $(x_i f_i : \sum x_i f_i) \times$ $\times$ 100% (графа 8 в продолжении табл. 1), а затем в графе 9 находим кумулятивные (накопленные) итоги суммарного дохода в % (q_i).

Продолжение табл. 1

Суммарный доход по группам, % к итогу $(x_i f_i : \sum x_i f_i)$ 100%	**Накопленные итоги суммарного дохода, % q_i**
8	**9**
5,8	5,8
10,8	16,6
13,0	29,6
13,0	42,6
11,4	54,0
17,5	71,5
11,4	82,9
17,1	100,0
100,0	—

По данным граф 6 и 9,

$G = \sum p_i q_{i+1} - \sum p_{i+1} q_i = [(17{,}5 \cdot 16{,}6 + 37 \cdot 29{,}6 + 53{,}7 \cdot 42{,}6 +$ $+ 66{,}7 \cdot 54 + 76{,}1 \cdot 71{,}5 + 87{,}3 \cdot 82{,}9 + 93 \cdot 100) - (37{,}0 \cdot 5{,}8 +$ $+ 53{,}7 \cdot 16{,}6 + 66{,}7 \cdot 29{,}6 + 76{,}1 \cdot 42{,}6 + 87{,}3 \cdot 54 + 93 \cdot 71{,}5 +$ $+ 100 \cdot 82{,}9)] : 10\,000 = (29253{,}44 - 25975{,}9) : 10\,000 = 0{,}327 \approx 0{,}33.$

По $G = 0{,}33$ делаем вывод, что степень концентрации суммарных денежных доходов у населения с более высокими доходами средняя, не очень высокая.

ПРАВИЛО СЛОЖЕНИЯ ДИСПЕРСИЙ

В совокупности, разбитой на группы по какому-либо признаку, общая вариация определенного показателя складывается из вариации внутригрупповой и межгрупповой. Это находит отражение в правиле сложения дисперсий. Другими словами, если совокупность разбита на группы по какому-либо факторному признаку и по каждой группе рассчитаны групповые средние и дисперсии определенного результативного показателя, то общая дисперсия последнего для всей совокупности может быть рассчитана по формуле

$$\sigma^2 = \overline{\sigma^2} + \delta^2,$$

где $\overline{\sigma^2} = \dfrac{\sum_1^m \sigma_i^2 n_i}{\sum_1^m n_i}$ — средняя из групповых дисперсий (m — число выделенных групп, n_i — число единиц в каждой группе);

$\delta^2 = \dfrac{\sum_1^m \left(\bar{y}_i - \bar{y}\right)^2 n_i}{\sum_1^m n_i}$ — межгрупповая дисперсия ($\bar{y}_i$ — групповые средние, y — общая средняя изучаемого показателя).

Отношение межгрупповой дисперсии к общей именуется **эмпирическим коэффициентом детерминации** ($\eta^2_{эмп} = \dfrac{\delta^2}{\sigma^2}$), а корень квадратный из него — **эмпирическое корреляционное отношение** ($\eta_{эмп} = \sqrt{\dfrac{\delta^2}{\sigma^2}}$) используется для оценки тесноты зависимости вариации результативного показателя от признака, положенного в основу группировки.

Задача 5.7

Имеются следующие условные данные по трем группам рабочих с разным стажем работы:

Стаж работы, лет	Число рабочих n_i	Средняя заработная плата, руб. $\bar{y}_i$	Среднее квадратическое отклонение заработной платы, руб. σ_i
До 3	10	2500	120
3–10	15	2600	100
Более 10	25	2700	200

1. Рассчитать: а) среднюю заработную плату для всей совокупности рабочих; б) общую дисперсию и среднее квадратическое отклонение заработной платы.
2. С помощью эмпирического корреляционного отношения определить степень влияния стажа работы на вариацию заработной платы.

Решение: 1. Общая средняя

$$\bar{y} = \frac{\sum \bar{y}_i n_i}{\sum n_i} = \frac{2500 \cdot 10 + 2600 \cdot 15 + 2700 \cdot 25}{10 + 15 + 25} = 2630 \text{ (руб.).}$$

Общая дисперсия находится по правилу сложения дисперсий; сначала находим среднюю из групповых дисперсий:

$$\overline{\sigma^2} = \frac{\sum \sigma_i^2 n_i}{\sum n_i} = \frac{14400 \cdot 10 + 10000 \cdot 15 + 40000 \cdot 25}{50} = \frac{1\,294\,000}{50} = 25880,$$

а затем межгрупповую дисперсию

$$\delta^2 = \frac{\sum (\bar{y}_i - \bar{y})^2 n_i}{\sum n_i} =$$

$$= \frac{(2500 - 2630)^2 \cdot 10 + (2600 - 2630)^2 \cdot 15 + (2700 - 2630)^2 \cdot 25}{50} = 6100.$$

Тогда общая дисперсия заработной платы

$$\sigma^2 = \overline{\sigma^2} + \delta^2 = 25880 + 6100 = 31980.$$

Отсюда среднее квадратическое отклонение заработной платы во всей совокупности рабочих

$$\sigma = \sqrt{31980} \approx 178{,}8 \text{ (руб.).}$$

2. Коэффициент детерминации

$$\eta^2_{\text{эмп}} = \frac{\delta^2}{\sigma^2} = \frac{6100}{31\,980} = 0{,}1907.$$

Эмпирическое корреляционное отношение

$$\eta_{\text{эмп}} = \sqrt{0{,}1907} = 0{,}44,$$

что означает не очень сильную зависимость вариации заработной платы от стажа работы.

Задача 5.8

Предположим, имеются следующие данные по трем факультетам одного из вузов:

Факультет	Численность преподавателей n	Доля лиц старше 60 лет p_i	Дисперсия доли по группам $\sigma_i^2 = p_i q_i$
1	2	3	4
I	160	0,28	0,2016
II	220	0,15	0,1275
III	120	0,10	0,0900

Определить долю преподавателей старше 60 лет в целом по трем факультетам, а также общую дисперсию доли (двумя способами).

Решение: Средняя доля (общая)

$$\bar{p} = \frac{\sum p_i n_i}{\sum n_i} = \frac{0{,}28 \cdot 160 + 0{,}15 \cdot 220 + 0{,}1 \cdot 120}{160 + 220 + 120} = \frac{898}{500} = 0{,}1796,$$

или 17,96%.

Способ 1. Находим общую дисперсию доли по формуле $\sigma^2 = \bar{p}\,\bar{q} = \bar{p}(1 - \bar{p})$:

$$\sigma^2 = 0{,}1796 \cdot 0{,}8204 = 0{,}1473.$$

Способ 2. Общую дисперсию доли можно определить и по правилу сложения дисперсий: $\sigma^2 = \overline{\sigma^2} + \delta^2$.

а) Сначала находим групповые дисперсии $\sigma_i^2 = p_i q_i$ (см. графу 4 таблицы):

$$\sigma_1^2 = 0{,}28 \cdot 0{,}72 = 0{,}2016; \quad \sigma_2^2 = 0{,}15 \cdot 0{,}85 = 0{,}1275;$$
$$\sigma_3^2 = 0{,}1 \cdot 0{,}9 = 0{,}09.$$

б) Средняя из групповых дисперсий

$$\overline{\sigma_i^2} = \frac{\sum \sigma_i^2 n_i}{\sum n_i} = \frac{0{,}2016 \cdot 160 + 0{,}1275 \cdot 220 + 0{,}09 \cdot 120}{500} = 0{,}1422.$$

в) Рассчитываем межгрупповую дисперсию:

$$\delta^2 = \frac{\sum \left(p_i - \bar{p}\right)^2 n_i}{\sum n_i} =$$

$$= \frac{(0{,}28 - 0{,}1796)^2 \cdot 160 + (0{,}15 - 0{,}1796)^2 \cdot 220 + (0{,}1 - 0{,}1796)^2 \cdot 120}{500} =$$

$$= \frac{2{,}56592}{500} = 0{,}0051.$$

Общая дисперсия доли $\sigma^2 = \overline{\sigma^2} + \delta^2 = 0{,}1422 + 0{,}0051 = 0{,}1473.$

ПОНЯТИЕ О МОМЕНТАХ РАСПРЕДЕЛЕНИЯ

Изучая данную тему, следует обратить внимание на такие характеристики вариационных рядов, как моменты распределения. В математической статистике под моментом k-го порядка (M_k) понимается средняя арифметическая k-й степени отклонений отдельных вариантов от какой-то постоянной величины A. Если A — любое произвольное число, то момент k-го порядка можно записать в виде

$$M_k = \frac{\sum (x_i - A)^k f_i}{\sum f_i}.$$

Если принять $A = 0$, то моменты в этом случае называются **начальными**. Обозначив их через m, можно записать

$$m_k = \frac{\sum (x_i)^k f_i}{\sum f_i}.$$

Тогда начальный момент первого порядка

$$m_1 = \frac{\sum x_i f_i}{\sum f_i} = \bar{x} \quad \text{(средняя арифметическая);}$$

начальный момент второго порядка

$$m_2 = \frac{\sum x_i^2 f_i}{\sum f_i} = \overline{x^2};$$

начальный момент третьего порядка

$$m_3 = \frac{\sum x_i^3 f_i}{\sum f_i} = \overline{x^3} \quad \text{и т.д.}$$

Практически при анализе вариационных рядов ограничиваются расчетом моментов первых четырех порядков.

Если принять $A = \bar{x}$, то моменты называются **центральными**; обозначив их через μ_k, можно записать в общем виде:

$$\mu_k = \frac{\sum (x_i - \bar{x})^k f_i}{\sum f_i},$$

а моменты первых четырех порядков в виде:

$$\mu_1 = \frac{\sum (x_i - \bar{x}) f_i}{\sum f_i} = 0; \quad \mu_2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i} = \sigma^2;$$

$$\mu_3 = \frac{\sum (x_i - \bar{x})^3 f_i}{\sum f_i}; \quad \mu_4 = \frac{\sum (x_i - \bar{x})^4 f_i}{\sum f_i}.$$

Как видно из записанных формул, **центральный момент второго порядка** представляет собой не что иное, как **дисперсию**. Центральный момент третьего порядка (μ_3) используется для характеристики асимметричности распределения, ибо для симметричных рядов всегда

$$\sum\left(x_i - \overline{x}\right)^3 = 0.$$

Чтобы можно было сравнивать асимметричность в разных рядах, μ_3 сопоставляют со средним квадратическим отклонением в кубе и найденное отношение, называемое **нормированным** моментом третьего порядка $\left(r_3 = \frac{\mu_3}{\sigma^3}\right)$, принимается в качестве показателя асимметрии (A_s). При $r_3 > 0$ говорят о правосторонней асимметрии распределения (вытянутость вправо), при $r_3 < 0$ — о левосторонней асимметрии.

(Для характеристики асимметрии (A_s) распределения также используется показатель $A_s = \frac{\overline{x} - M_0}{\sigma}$, именуемый **коэффициентом асимметрии Пирсона.**)

Центральный момент четвертого порядка (μ_4) используется в *качестве характеристики «крутости» ряда* (эксцесса).

Для нормального распределения характерно следующее соотношение между μ_4 и μ_2: $\mu_4 = 3\ (\mu_2^2)$; но $\mu_2 = \sigma^2$, отсюда $\frac{\mu_4}{\sigma^4} = 3$. Поэтому в качестве показателя эксцесса (Ex) принимается показатель

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

Если $Ex > 0$, то ряд островершинен, если $Ex < 0$, то ряд низковершинен.

Все перечисленные характеристики играют большую роль при анализе вариационных рядов и определении типа кривой распределения и при выравнивании вариационных рядов.

Задачи и упражнения для самостоятельной работы (по темам 4 и 5)

Задача 1

На конец октября 1995 г. в РФ имелось следующее распределение безработных по возрастным группам:

Возрастные группы, лет	Численность безработных, % к итогу
16—19	10,2
20—24	17,9
25—29	12,7
30—49	48,8
50—54	4,0
55—59	4,5
60—72	1,9
Всего	**100,0**

По данным распределения определить:

1) средний (арифметический) возраст безработных;
2) модальный возраст (моду);
3) медианный возраст (медиану);
4) среднее квадратическое отклонение;
5) коэффициент вариации.

Задача 2

Имеются следующие данные за 1995 и 1996 гг. по РФ об урожайности, посевных площадях и валовом сборе пшеницы (озимой и яровой):

Культура	1995		1996	
	Урожайность, ц/га	Посевная площадь, млн. га	Урожайность, ц/га	Валовой сбор, млн. ц
Пшеница озимая	16,9	8,2	17,9	167
Пшеница яровая	10,3	15,8	11,0	182

Определить среднюю урожайность пшеницы: а) в 1995 г. и б) в 1996 г.

Ответ: а) 12,555 ц/га; б) 13,48 ц/га.

Задача 3

По данным таблицы «б» из задачи 3 темы 2 определить средний возраст безработных в 1996 г.: а) мужчин; б) женщин.

Ответ: а) 35,5 года; б) 33,8 года.

Задача 4

По тем же данным (таблица «б» из задачи 3 темы 2) определить:

1) модальный возраст (моду) для мужчин и женщин;
2) медианный возраст (медиану) для мужчин и женщин.

Ответ: 1) $Mo_{\text{мужчин}} = 22{,}6$ года и $Mo_{\text{женщин}} = 22$ года;
2) $Me_{\text{мужчин}} = 35$ лет и $Me_{\text{женщин}} = 33{,}3$ года.

Задача 5

Имеются следующие данные о распределении населения Московской области по уровню среднемесячного душевого дохода в 1995 г.:

Среднемесячный душевой доход, тыс. руб.	Численность населения, % к итогу
До 200	15,3
200—400	50,6
400—600	23,5
600—800	7,3
800—1000	2,2
Свыше 1000	1,1

Определить в данном распределении:

1) среднемесячный душевой доход по области в целом;
2) моду;
3) медиану;
4) среднее квадратическое отклонение доходов;
5) децильный коэффициент дифференциации (*ДКД*) доходов;
6) коэффициент Джини (G).

Ответ: 1) $\bar{x}$ = 367,6 тыс. руб.; 2) Mo = 313,1 тыс. руб.; 3) Me = 337,2 тыс. руб.; 4) σ = 193,8 тыс. руб.; 5) *ДКД* = 4,7 раза; 6) G = 0,265.

Задача 6

Распределение населения Москвы по уровню среднемесячного душевого дохода в 1995 г. характеризовалось следующими данными:

Среднемесячный душевой доход, руб.	Численность населения, % к итогу
До 200	5,6
200—400	21,1
400—600	17,0
600—800	10,1
800—1000	6,1
Свыше 1000	40,1
Всего	**100,0**

Определить:

1) среднедушевой денежный доход на основе средней арифметической, моды и медианы;
2) дисперсию, среднее квадратическое отклонение, коэффициент вариации;
3) децильный коэффициент дифференциации доходов, *ДКД*;
4) коэффициент Джини (показатель концентрации доходов), G.

Ответ: 1) $\bar{x}$ = 720,6 тыс. руб., Mo = 1091,8 тыс. руб., Me = 724,7 тыс. руб.;
2) σ^2 = 126405,6, σ = 356 тыс. руб., V = 49,4%;
3) *ДКД* = 4,8 раза; 4) G = 0,272.

Задача 7

Имеются следующие данные о распределении населения Санкт-Петербурга по уровню среднемесячного душевого дохода в 1995 г.:

Среднемесячный душевой доход, тыс. руб.	Численность населения, % к итогу
До 200	10,9
200—400	30,0
400—600	22,9
600—800	14,0
800—1000	8,3
Свыше 1000	13,9

Определить:

1) среднемесячный душевой доход в целом по городу;
2) моду и медиану (по формулам и графически);
3) коэффициент вариации доходов;
4) *ДКД* населения по доходам;
5) степень концентрации доходов по группам (с помощью G).

Построить кривую Лоренца.

По результатам анализа сделать выводы.

Ответ: 1) $\bar{x} = 541$ тыс. руб.;
2) $Mo = 345{,}8$ тыс. руб., $Me = 479{,}5$ тыс. руб.;
3) $V = 57{,}5\%$; 4) $ДКД = 5{,}75$ раза; 5) $G = 0{,}265$.

Задача 8

По переписи населения 1926 г. в России доля грамотных среди женщин составляла 46%, а среди мужчин 77%. Определить общий (средний) процент грамотности всего населения и дисперсию этого показателя, если женщины составляли 53% в общей численности населения.

Ответ: 1) $\bar{p} = 60{,}57\%$ (0,6057); 2) $\sigma^2 = \bar{p}\ \bar{q} = 0{,}2388$.

Задача 9

В коллективных хозяйствах района средняя урожайность зерновых составила 19 ц/га при среднем квадратическом отклонении 3 ц/га, а в фермерских хозяйствах — соответственно 26 ц/га и 4 ц/га.

Определить:

1) среднюю урожайность зерновых по району, если известно, что посевная площадь под зерновыми в коллективных хозяйствах в 9 раз превышает площадь фермерских хозяйств;
2) общую дисперсию и среднее квадратическое отклонение урожайности зерновых в районе (по правилу сложения дисперсии).

Ответ: 1) $\bar{x} = 19{,}7$ ц/га;
2) $\sigma^2 = \overline{\sigma^2} + \delta^2 = 9{,}7 + 4{,}41 = 14{,}11$, $\sigma = 3{,}76$ ц/га.

Задача 10

Для изучения уровня заработной платы рабочих на предприятии выборочно обследовано 500 мужчин и 300 женщин. Результаты исследования показали, что у мужчин средняя заработная плата составила 1200 руб. при среднем квадратическом отклонении 200 руб., *а у женщин — соответственно 800 руб. и 150 руб.*

Определить:

1) общую среднюю заработную плату рабочих на заводе;
2) среднюю из групповых дисперсий;
3) межгрупповую дисперсию;
4) общую дисперсию заработной платы;
5) коэффициент вариации заработной платы на предприятии.

О т в е т: 1) $\bar{x}$ = 1050 руб.; 2) $\overline{\sigma^2}$ = 33 437,5; 3) δ^2 = 37 500; 4) σ^2 = 70 937,5; 5) V = 25,4%.

Задача 11

В I полугодии 2000 г. распределение населения России по среднедушевому денежному доходу в месяц характеризовалось следующими данными:

Среднедушевой денежный доход в месяц, руб.	Численность населения, % к итогу
До 400	2,7
400—600	6,6
600—800	9,3
800—1000	10,1
1000—1200	9,9
1200—1600	16,9
1600—2000	12,6
Свыше 2000	31,9
Итого	**100,0**

Определить:

1) среднедушевой денежный доход в месяц в целом по России;
2) модальный и медианный доходы;
3) среднее квадратическое отклонение и коэффициент вариации доходов;
4) децильный коэффициент дифференциации доходов;
5) коэффициент асимметрии Пирсона;
6) коэффициент Джини (индекс концентрации доходов).

Задача 12

Ниже приведены показатели работы предприятий лесной промышленности за январь—август 1997 г. по вывозке древесины:

Годовой объем вывозки древесины, тыс. плотных м³	Количество предприятий	Произведено за январь–август 1997 г.	
		тыс. плотных м³	% к производству за январь–август 1996 г.
Менее 500	22	5157	81,0
500–1000	14	4076	79,5
1000–3000	11	8784	79,5
Свыше 3000	9	12 429	80,8
Итого	**56**	**30 446**	**?**

Определить:

1) среднюю годовую мощность предприятий лесной промышленности по вывозке древесины;
2) моду и медиану в ряду распределения предприятий по мощности;
3) средний объем производства древесины на одно предприятие за период январь–август 1997 г.;
4) коэффициент Джини – индекс концентрации производства продукции в январе–августе 1997 г.;
5) средний процент изменения объема производства продукции по всем предприятиям в январе–августе 1997 г. по сравнению с аналогичным периодом предыдущего года.

Задача 13

Имеются следующие данные по предприятиям оптовой торговли РФ (за апрель 1995 г.):

Группы предприятий по объему товарооборота, млн. руб.	Число предприятий, % к итогу	Численность работников по группам, % к итогу
Менее 1	37,5	10,5
1–25	22,5	10,8
25–50	8,7	5,6
50–100	8,7	7,9
100–200	7,8	9,3
200–500	7,3	13,4
Более 500	7,5	42,5
Итого	**100,0**	**100,0**

1. Рассчитать: а) средний объем товарооборота на одно предприятие; б) моду; в) медиану.

2. Определить степень дифференциации предприятий по объему товарооборота.
3. Построить на одном графике две кривые Лоренца: по общей численности работников и по общему объему товарооборота.
4. С помощью коэффициентов Джини сравнить степень концентрации оптового товарооборота и численности работников.

Задача 14

По данным анкетного обследования рабочих Санкт-Петербурга, имеющих вторичную занятость, получены следующие результаты за 1997 г., характеризующие влияние дополнительной занятости на доходы работников:

Размер дополнительного заработка, % к основному	Количество ответов		
	Всего	мужчин	женщин
До 20	114	52	62
21–40	108	88	20
41–60	137	96	41
61–80	298	221	77
81–100	40	29	11
Свыше 100	56	52	4
Всего	**753**	**538**	**215**

1. Определить средний размер дополнительного заработка у респондентов, имеющих вторичную занятость: а) у мужчин; б) у женщин; в) в целом.
2. Определить моду и медиану дополнительного заработка: а) у мужчин; б) у женщин; в) в целом.
3. Рассчитать групповые дисперсии исследуемого показателя (у мужчин и у женщин), а также межгрупповую дисперсию.
4. Рассчитать общий коэффициент вариации дополнительного заработка.
5. С помощью эмпирического корреляционного отношения определить, в какой степени вариация дополнительного заработка зависит от признака пола.

Тема 6
ВЫРАВНИВАНИЕ ВАРИАЦИОННЫХ РЯДОВ (ПОСТРОЕНИЕ ТЕОРЕТИЧЕСКИХ РАСПРЕДЕЛЕНИЙ)

Под выравниванием вариационных рядов понимается замена **эмпирического** распределения близким к нему по характеру **теоретическим** (вероятностным) распределением, имеющим определенное аналитическое выражение (параметры последнего определяются по данным эмпирического распределения).

Из многих форм кривых распределения, по которым может выравниваться вариационный ряд, студенты должны ознакомиться с **нормальным распределением**. График нормального распределения имеет форму колоколообразной кривой, симметричной относительно $\bar{x}$, концы которой асимптотически приближаются к оси абсцисс (рис. 5). Она имеет точки перегиба, абсциссы которых находятся на расстоянии σ от центра симметрии. Эта кривая выражается уравнением

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}},$$

где y — ордината кривой нормального распределения;

$t = \dfrac{x - \bar{x}}{\sigma}$ — нормированные отклонения, т.е. отклонения отдельных вариантов от $\bar{x}$, выраженные в единицах среднего квадратического отклонения σ.

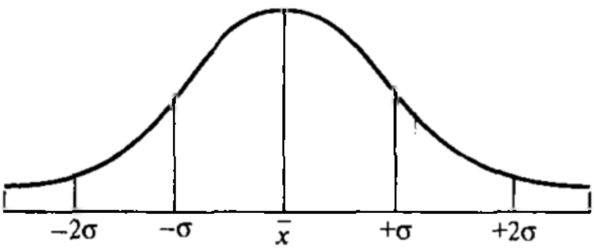

Рис. 5. Кривая нормального распределения

При выравнивании вариационного ряда по кривой нормального распределения теоретические частоты ряда определяются по формуле

$$f' = \frac{Nh}{\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}},$$

где $N = \sum f$ — сумма всех частот вариационного ряда;

h — величина интервала в группах (классах);

σ — среднее квадратическое отклонение;

$t = \frac{x - \bar{x}}{\sigma}$ — нормированное отклонение вариантов от средней арифметической.

Величина $\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ табулирована, ее легко определить по таблице Приложения 1 как функцию t, т.е. $\varphi(t)$.

Как видно из формулы, основными параметрами кривой нормального распределения являются $\bar{x}$ и σ. По этим характеристикам ее и можно построить.

Распределение Пуассона. В целом ряде случаев, если вариационный ряд представляет собой распределение по дискретному признаку, где по мере увеличения значений признака x частоты резко уменьшаются и где средняя арифметическая ряда равна или близка по значению к дисперсии, т.е. $\bar{x} = \sigma^2$, то такой ряд можно выравнять по кривой Пуассона (рис. 6), аналитическое выражение которой

$$P_x = \frac{a^x e^{-a}}{x!},$$

где P_x — вероятность наступления отдельных значений x;

$a = \bar{x}$ — средняя арифметическая ряда.

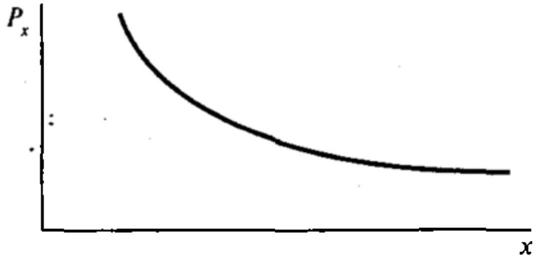

Рис. 6. Кривая Пуассона

Теоретические частоты при выравнивании эмпирических данных по кривой Пуассона определяются по формуле

$$f' = NP_x,$$

где f' — теоретические частоты;

N — общее число единиц ряда.

КРИТЕРИИ СОГЛАСИЯ

После выравнивания ряда, т.е. нахождения теоретических частот, возникает необходимость проверить, случайны или существенны расхождения между эмпирическими и теоретическими частотами, и тем самым проверить правильность выдвинутой при выравнивании ряда гипотезы о наличии того или иного характера распределения в эмпирическом ряду.

Для оценки близости эмпирических (f) и теоретических (f') частот можно применить один из критериев согласия: критерий Пирсона (χ^2 — «хи-квадрат»), критерий Романовского, критерий Колмогорова (λ — «лямбда»).

Критерий Пирсона (χ^2) представляет собой сумму отношений квадратов расхождений между f и f' к теоретическим частотам:

$$\chi^2 = \sum \frac{(f-f')^2}{f'}.$$

Фактическое значение χ^2 сравнивают с критическим, определяемым по специальным таблицам в зависимости от принимаемого уровня значимости и числа степеней свободы.

Уровень значимости (α) — вероятность допуска ошибки первого рода, т.е. отвергнуть правильную гипотезу о законе распределения — обычно принимается равным 5% или 1% ($\alpha = 0{,}05$ или $\alpha = 0{,}01$).

Число степеней свободы (v) рассчитывается как число групп (m) в ряду распределения минус единица и минус число параметров эмпирического распределения, использованных для нахождения теоретических частот. Так, при выравнивании по кривой нормального распределения число степеней свободы $v = m - 1 - 2$, поскольку при расчете теоретических частот используется два параметра эмпирического распределения: $\bar{x}$ и σ, т.е. $v = m - 3$.

Если фактическое χ^2 оказывается меньше табличного (критического), то расхождения между эмпирическими и теоретическими частотами можно считать случайными.

При отсутствии таблиц для оценки случайности расхождений теоретических и эмпирических частот можно воспользоваться **критерием Романовского**

$$\frac{|\chi^2 - v|}{\sqrt{2v}}.$$

Если указанное отношение меньше 3, то расхождения считают случайными, если больше 3, то они существенны.

Критерий Колмогорова (λ) основан на определении максимального расхождения между накопленными частостями или частотами эмпирического и теоретического распределений:

$$\lambda = d\sqrt{N},$$

где d — максимальная величина расхождений между накопленными частостями;

N — число наблюдений, или сумма всех частот.

Если пользоваться не накопленными частостями, а частотами (абсолютными показателями), то формула примет вид

$$\lambda = \frac{D}{\sqrt{N}},$$

где D — максимальная разность между накопленными частотами;

N — сумма всех частот.

Рассмотрим решение некоторых задач к этой теме.

Задача 6.1

Пусть имеется следующее распределение 200 проб нити по крепости (графы 1 и 2 таблицы):

Крепость нити, г	Число проб	Середина интервала	$x-\bar{x}$	$\frac{x-\bar{x}}{\sigma}=t$	$\varphi(t)$	$154\,\varphi(t)\approx f'$
1	2	3	4	5	6	7
120—130	1	125	−36,4	−2,80	0,008	1
130—140	8	135	−26,4	−2,03	0,051	8
140—150	27	145	−16,4	−1,26	0,180	28
150—160	58	155	−6,4	−0,49	0,354	55
160—170	56	165	3,6	0,28	0,384	59
170—180	34	175	13,6	1,05	0,230	35
180—190	14	185	23,6	1,82	0,076	12
190—200	2	195	33,6	2,58	0,014	2
Итого	**200**	—	—	—	—	**200**

Исходя из гипотезы о нормальном распределении результатов испытаний необходимо выравнять ряд по кривой нормального распределения (т.е. рассчитать теоретические частоты) и оценить близость эмпирических и теоретических частот с помощью критериев согласия: Пирсона (χ^2), Романовского и Колмогорова (λ).

Для нахождения теоретических частот используем формулу

$$f' = \frac{Nh}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}, \text{ или } f' = \frac{Nh}{\sigma}\varphi(t),$$

где $t = \frac{x-\bar{x}}{\sigma}$ — нормированные отклонения от средней, т.е. $\bar{x}$ и σ — основные параметры кривой нормального распределения.

С них и начинаем свои расчеты. Опуская вычисления, запишем результаты:

1) $\bar{x} = 161{,}4$;
2) $\sigma = 13$.

Дальнейшие расчеты таковы:

3) находим отклонения отдельных вариантов от средней (графа 4);
4) делим каждое отклонение на σ, т.е. находим нормированные отклонения $t = \frac{x-\bar{x}}{\sigma}$ (графа 5);
5) зная t, по таблице Приложения 1 находим $\varphi(t)$ (графа 6);
6) так как ряд с равными интервалами, рассчитываем постоянный множитель $\text{const} = \frac{Nh}{\sigma}$. В нашем примере $\text{const} = \frac{200 \cdot 10}{13} = 154$;
7) умножая последовательно 154 на $\varphi(t)$ и округляя результаты до целых чисел, находим теоретические частоты (графа 7).

Как видно из таблицы, теоретические частоты (f') близки к эмпирическим (f), хотя отдельные расхождения имеют место.

Для суждения о случайности или существенности этих расхождений используем ряд критериев согласия:

1. Критерий Пирсона: $\chi^2 = \sum \frac{(f-f')^2}{f'}$.

Расчет этого критерия показан в следующей таблице:

f	f'	$f-f'$	$(f-f')^2$	$\frac{(f-f')^2}{f'}$
1	1	0	0	0
8	8	0	0	0
27	28	−1	1	0,04
58	55	3	9	0,16
56	59	−3	9	0,15
34	35	−1	1	0,03
14	12	2	4	0,33
2	2	0	0	0
200	**200**	—	—	$\chi^2 = 0{,}71$

В рассматриваемом примере ряд имеет 8 групп (классов) вариантов, следовательно, и 8 групп частот. Поэтому число степеней свободы для последних (при выравнивании по кривой нормального распределения) $\nu = 8 - 3 = 5$. Примем наиболее часто используемый уровень значимости $\alpha = 0{,}05$ и обратимся к таблице Приложения 4.

По таблице значений χ^2-критерия Пирсона для степеней свободы $\nu = 5$ и уровня значимости $\alpha = 0{,}05$ определяем, что $\chi^2_{\text{табл}} = 11{,}07$. Так как полученное в задаче фактическое значение $\chi^2_{\text{факт}} = 0{,}71$, т.е. меньше табличного, то, следовательно, можно считать случайными расхождения между эмпирическими и теоретическими частотами и выдвинутая гипотеза о близости эмпирического распределения к нормальному не опровергается.

2. Применим критерий Романовского:

$$\frac{|\chi^2 - \nu|}{\sqrt{2\nu}} = \frac{|0{,}71 - 5|}{\sqrt{10}} = 1{,}4.$$

Поскольку $1{,}4 < 3$, то можно считать расхождения между эмпирическими и теоретическими частотами случайными.

3. Попробуем проверить нашу гипотезу с помощью критерия Колмогорова $\left(\lambda = \frac{D}{\sqrt{N}}\right)$. Для этого запишем накопленные частоты эмпирического и теоретического распределений и найдем максимальный разрыв между ними:

f	f'	Накопленные частоты		$\lvert S - S' \rvert$
		эмпирические (S)	теоретические (S')	
1	1	1	1	0
8	8	9	9	0
27	28	36	37	1
58	55	94	92	2
56	59	150	151	1
34	35	184	186	2
14	12	198	198	0
2	2	200	200	0

Максимальный разрыв $D = 2$, поэтому $\lambda = \frac{D}{\sqrt{N}} = \frac{2}{\sqrt{100}} = 0{,}2.$

По таблицам $P(\lambda)$ (см. таблицу Приложения 6) находим для $\lambda = 0{,}2$, что $P = 1{,}000$. Следовательно, вполне можно полагать, что расхождения между f и f' носят случайный характер.

Задача 6.2

В течение рабочей недели производилось наблюдение за работой 50 станков и регистрировались неисправности, требовавшие остановки станков для их регулировки. Результаты наблюдений следующие:

Число неисправностей (x)	0	1	2	3	4	5
Число станков (f)	14	16	10	7	2	1

Требуется: 1) вычислить вероятности и теоретические частоты числа неисправностей, считая, что распределение последних подчиняется закону Пуассона; 2) оценить близость эмпирических и теоретических частот с помощью критериев Пирсона, Романовского и Колмогорова.

Решение.

А. Рассчитаем среднее число неисправностей:

$$\bar{x} = a = \frac{0 \cdot 14 + 1 \cdot 16 + 2 \cdot 10 + 3 \cdot 7 + 4 \cdot 2 + 5 \cdot 1}{14 + 16 + 10 + 7 + 2 + 1} = \frac{70}{50} = 1{,}4.$$

Б. Находим значение $e^{-1{,}4} = 0{,}2466$.

В. Подставляя в формулу $P_x = \dfrac{a^x e^{-a}}{x!} = \dfrac{1{,}4^x \cdot 0{,}2466}{x!}$ значения $x = 0, 1, 2, 3, 4, 5$, получаем вероятности наступления случаев неисправностей от 0 до 5.

Г. Умножив последние на 50 (общее число единиц распределения), получим теоретические частоты числа неисправностей, т.е.

$$f' = NP_x.$$

Значения P_x и f' (округленные до целого числа) показаны в приводимой ниже таблице:

P_x	f' (теоретические частоты) = $50P_x$
0,2466	12
0,3452	17
0,2417	12
0,1128	6
0,0395	2
0,0111	1
Итого	**50**

Для оценки близости эмпирических и теоретических частот воспользуемся критериями Пирсона, Романовского и Колмогорова.

А. Критерий Пирсона: $\chi^2 = \sum \dfrac{(f - f')^2}{f'}$.

Все расчеты показаны в следующей таблице:

x	f	f'	$f-f'$	$(f-f')^2$	$\frac{(f-f')^2}{f'}$
0	14	12	2	4	0,33
1	16	17	−1	1	0,06
2	10	12	−2	4	0,33
3	7	6	1	1	0,17
4	2	2	0	0	0
5	1	1	0	0	0

Фактическое значение $\chi^2_{\text{факт}} = 0{,}89$. По Приложению 4 находим критическое (табличное) значение χ^2 при $\nu = 6 - 2 = 4$ и $\alpha = 0{,}05$, $\chi^2_{\text{табл}} = 9{,}49$. Так как $\chi^2_{\text{факт}} < \chi^2_{\text{табл}}$, т.е. $0{,}89 < 9{,}49$, то имеем все основания считать расхождения между эмпирическими и теоретическими частотами случайными, а следовательно, не опровергнутой гипотезу о том, что распределение числа неисправностей подчиняется закону Пуассона.

Б. Применим критерий Романовского: $\frac{|\chi^2 - \nu|}{\sqrt{2\nu}} = \frac{|1-4|}{\sqrt{8}} < 3$. *Следовательно, расхождения случайны.*

В. И, наконец, по критерию Колмогорова: $\lambda = \frac{D}{\sqrt{N}}$.

Накопленные частоты		$\lvert S - S'\rvert$
эмпирические (S)	**теоретические (S')**	
14	12	2 (D)
30	29	1
40	41	1
47	47	0
49	49	0
50	50	0

Таким образом, $\lambda = \frac{2}{\sqrt{50}} = \frac{2}{7{,}07} \approx 0{,}3$. По таблице Приложения 6 находим $P(\lambda \approx 0{,}3) \approx 1$.

Итак, все три критерия оценивают расхождения между эмпирическими и теоретическими частотами как случайные, не опровергая тем самым выдвинутую гипотезу о том, что распределение станков по числу неисправностей подчиняется закону Пуассона.

Л и т е р а т у р а (к темам 5 и 6)

[1, с. 77—135], [2, с. 181—224], [3, с. 89—100, 164—169].

Контрольные вопросы (к темам 5 и 6)

1. Как определяются мода и медиана в дискретных и интервальных вариационных рядах?
2. Как графически отыскать моду и медиану?
3. Что такое квартили и как они рассчитываются в вариационном ряду?
4. Что характеризуют и как рассчитываются децили?
5. Как рассчитывается и что характеризует децильный коэффициент дифференциации?
6. Чем вызвана необходимость изучения вариации признака?
7. Укажите основные показатели вариации.
8. Какие вам известны способы расчета дисперсии и среднего квадратического отклонения?
9. Как определяется дисперсия альтернативного признака?
10. Что такое коэффициент вариации?
11. Что вам известно о правиле сложения дисперсий?
12. Дайте определения понятия моментов распределения. Какие вы знаете моменты распределения?
13. В чем суть выравнивания вариационных рядов (по кривой нормального распределения и кривой Пуассона)?
14. Какие критерии согласия вам известны? Что они характеризуют?

Задачи и упражнения для самостоятельной работы

Задача 1

Распределение населения РФ по размеру среднемесячного душевого денежного дохода в апреле 1993 г. характеризовалось данными, приведенными в таблице.

Среднемесячный душевой доход, тыс. руб.	Численность населения, млн. чел.
До 5	6,6
5—10	39,8
10—15	45,6
15—20	33,4
20—25	21,6
Свыше 25	1,7
Итого	**148,7**

1. Определить среднемесячный душевой доход в апреле 1993 г.
2. Определить среднее квадратическое отклонение.
3. Исходя из гипотезы о нормальном распределении, рассчитать теоретические частоты в данном ряду.
4. С помощью критериев Пирсона (χ^2), Романовского и Колмогорова (λ) проверить, согласуется ли эмпирическое распределение с гипотетическим нормальным.

О т в е т: 1) $\bar{x}$ = 13,465 тыс. руб.; 2) σ = 5,74 тыс. руб.;
3) $f_{теор}$: 8,4; 30,2; 51,1; 40,6; 15,1; 2,6;
4) χ^2 = 8,41; $K_{Ром}$ = 2,2; λ = 0,64.

Задача 2

Предположим, имеется следующее распределение 100 выборочно обследованных на торфяных участках проб по глубине залегания торфа:

Глубина залегания торфа, см	Число проб
70—80	2
80—90	6
90—100	19
100—110	30
110—120	22
120—130	13
130—140	5
140—150	3

1. Рассчитать теоретические частоты, исходя из гипотезы о нормальном распределении.
2. С помощью критериев согласия Пирсона, Романовского, Колмогорова проверить, согласуется ли эмпирическое распределение с гипотетическим нормальным распределением.

О т в е т: 1) $f_{теор}$: 2, 7, 18, 26, 25, 15, 6, 1 (с округлением до целых);
2) χ^2 = 5,6.

Задача 3

Для контроля качества продукции у 200 рабочих, изготовляющих однотипную продукцию, проверено по 50 изделий у каждого. Результаты следующие:

Количество бракованных изделий из 50 проверенных	Число рабочих
0	110
1	59
2	26
3	4
4	1

1. Рассчитать среднее количество бракованных изделий на одного рабочего.
2. Рассчитать теоретические частоты, исходя из гипотезы о распределении Пуассона.
3. Проверить, случайны или нет расхождения между эмпирическими и теоретическими частотами, используя для этого критерии Пирсона, Романовского, Колмогорова.

О т в е т: 1) $\bar{x}$ = 0,635, $e^{-0,635}$ = 0,5326;
2) $f_{теор}$: 106, 67, 21, 5, 1.

Задача 4

Имеется следующее распределение 200 фермерских хозяйств одной из областей по урожайности зерновых:

Урожайность, ц/га	Число хозяйств
До 20	7
20—24	18
24—28	34
28—32	56
32—36	38
36—40	21
40—44	10
Свыше 44	6
Итого	**200**

1. Рассчитать теоретические частоты (число хозяйств), исходя из гипотезы о нормальном распределении.
2. С помощью критериев Пирсона (χ^2), Романовского и Колмогорова проверить, согласуется ли эмпирическое распределение с гипотетическим нормальным распределением.

Задача 5

Ниже приведено распределение 80 рабочих по показателю выполнения дневной нормы выработки продукции:

Процент выполнения нормы	Число сотрудников
94—96	2
96—98	7
98—99,9	13
100—102	26
102—104	19
104—106	10
106—108	2
108—110	1

Проверить, можно ли считать данное распределение близким к нормальному.

Задача 6

По данным секции женской обуви одного из универмагов получено следующее распределение проданной за день обуви по размерам:

Размер	33	34	35	36	37	38	39	40	41
Количество пар	1	3	34	41	84	52	35	8	2

Проверить гипотезу о соответствии данного распределения нормальному.

Задача 7

По результатам выборочного обследования 400 семей города обеспеченность последних жилой площадью характеризуется следующими данными:

Размер жилой площади на одного члена семьи, м^2	Число семей
До 5	3
5—7	11
7—9	32
9—11	66
11—13	120
13—15	95
15—17	40
17—19	21
Более 19	12
Итого	**400**

1. Рассчитать теоретические частоты (число семей), исходя из гипотезы о нормальном распределении.
2. С помощью критериев согласия (Пирсона, Романовского, Колмогорова) *проверить*, согласуется ли эмпирическое распределение с гипотетическим нормальным.

Тема 7
ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ

В целом ряде случаев средние и относительные величины для какой-либо совокупности рассчитываются на основе данных выборочного наблюдения, суть которого заключается в том, что из генеральной совокупности, наудачу, чисто случайно, отбирается n единиц, составляющих выборочную совокупность; для отобранных единиц рассчитываются обобщенные характеристики (средние или относительные показатели), а затем результаты выборочного обследования распространяются на всю генеральную совокупность. Основной задачей при этом является определение ошибок выборки, т.е. возможных расхождений между выборочной средней ($\tilde{x}$) и генеральной ($\bar{x}$) или выборочной долей единиц (w), обладающих изучаемым *признаком, и генеральной долей* (p).

Различают среднюю и предельную ошибки выборки.

Средняя ошибка выборки (μ) характеризует среднюю величину возможных расхождений выборочной и генеральной средней (или доли) и представляет собой по форме и содержанию среднее квадратическое отклонение возможных значений выборочной средней от генеральной.

В математической статистике доказывается, что μ^2 — дисперсия возможных значений выборочной средней — в n раз меньше дисперсии изучаемого признака в генеральной совокупности, т.е. $\mu^2 = \frac{\sigma_г^2}{n}$.

Исходя из этого *средняя ошибка выборочной средней* при повторном отборе определяется по формуле

$$\mu = \sqrt{\frac{\sigma^2}{n}},$$

где σ^2 — дисперсия изучаемого показателя в генеральной совокупности*, а n — объем (численность) выборки.

Как видно из формулы, средняя ошибка выборки (μ) при повторном отборе зависит от показателя вариации (σ) и от объема выборки (n).

Средняя ошибка выборочной доли определяется по формуле $\mu = \sqrt{\frac{w(1-w)}{n}}$, где w — выборочная доля единиц, обладающих изучаемым признаком, а $w(1-w)$ — дисперсия доли (альтернативного признака).

При бесповторном отборе в формулах под знаком радикала появляется множитель $\left(1 - \frac{n}{N}\right)$, где N — численность генеральной совокупности.

Говоря об ошибках выборки, следует иметь в виду, что в каждой конкретной выборке разность $|\tilde{x} - \bar{x}|$ может быть меньше, больше или равна μ. И вероятность каждой такой ошибки различна.

Отклонение выборочной характеристики от генеральной называется **предельной ошибкой выборки**.

Предельная ошибка выборки, обозначаемая через Δ, рассчитывается как

$$\Delta = t\mu,$$

---

* Так как дисперсия изучаемого показателя в генеральной совокупности неизвестна, то фактически в формулу подставляется дисперсия выборочная, которая при большом числе наблюдений близка к генеральной.

где μ — средняя ошибка выборки, а t — коэффициент доверия, т.е. показатель, зависящий от вероятности (P), с которой предельная ошибка определяется.

Общая формула *предельной ошибки выборки* $\Delta = t\mu$ *для средней* приобретает вид $\Delta = t\sqrt{\frac{\sigma^2}{n}}$ (для п о в т о р н о г о отбора) или $\Delta = t\sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}$ (для б е с п о в т о р н о г о отбора), а *для доли* соответственно

$$\Delta = t\sqrt{\frac{w(1-w)}{n}} \quad \text{и} \quad \Delta = t\sqrt{\frac{w(1-w)}{n}\left(1 - \frac{n}{N}\right)}.$$

Вероятность появления определенной ошибки выборки находят с помощью теорем теории вероятностей.

Согласно так называемой центральной предельной теореме А.М. Ляпунова, при большом объеме выборки вероятность появления того или иного значения выборочной средней, а следовательно, и отклонения последней от генеральной подчиняется закону нормального распределения и может быть найдена как функция от t с помощью интеграла вероятностей Лапласа:

$$F(t) = P\left(\left|\tilde{x} - \bar{x}\right| \leqslant t\mu\right) = \frac{1}{\sqrt{2\pi}}\int_{-t}^{+t} e^{-\frac{t^2}{2}}\,dt,$$

где $t = \frac{\tilde{x} - \bar{x}}{\mu}$ — нормированное отклонение выборочной средней от генеральной.

Значения интеграла Лапласа (P) для разных t рассчитаны и приведены в специальных таблицах (см. Приложение 2). Так, при $t = 1$ вероятность $P = 0{,}683$. Это означает, что с вероятностью 0,683 (или 68,3%) можно гарантировать, что отклонение генеральной средней от выборочной не превысит однократной средней ошибки, т.е. что в генеральной совокупности среднее значение признака ($\bar{x}$) будет находиться в пределах $\tilde{x} - \mu \leqslant \bar{x} \leqslant \tilde{x} + \mu$.

Аналогично при $t = 2$ вероятность $P = 0{,}954$ (точнее, 0,9545) означает, что предельная ошибка не выйдет за размер 2μ, т.е. $\tilde{x} - 2\mu \leqslant \bar{x} \leqslant \tilde{x} + 2\mu$.

Сказанное выше относится и к расхождениям между выборочной долей единиц (w), обладающих определенным признаком, и генеральной долей (p). Так:

при $t = 1$

$$P(|w - p| \leqslant \mu) = 0{,}683, \text{ или } w - \mu \leqslant p \leqslant w + \mu;$$

при $t = 2$

$$P(|w - p| \leqslant 2\mu) = 0{,}954, \text{ или } w - 2\mu \leqslant p \leqslant w + 2\mu \text{ и т.д.}$$

Наряду с абсолютной величиной предельной ошибки в статистической практике рассчитывается *относительная ошибка* – процентное отношение абсолютной ошибки к исследуемому параметру:

$\Delta_{отн} = \frac{\Delta_{абс}}{\tilde{x}} 100\%$ или $\Delta_{отн} = \frac{\Delta_{абс}}{w} 100\%$. Можно и сразу определить относительную ошибку по формуле $\Delta_{отн} = t\sqrt{\frac{V^2}{n}}$, где V – коэффициент вариации.

Выборка считается репрезентативной, если $\Delta_{отн} \leqslant 5\%$.

Формулы предельной ошибки несколько конкретизируются и в зависимости от применяемого вида выборки. Так, указанные выше формулы применимы для собственно случайной и механической выборок. Для *типической* (районированной) выборки, т.е. когда генеральная совокупность делится на группы по какому-либо существенному признаку, а затем из каждой группы производится случайный отбор и общая средняя величина признака (или доля) определяется по групповым выборочным показателям, в формуле предельной ошибки выборки учитывается средняя из групповых дисперсий $\left(\overline{\sigma^2}\right)$, т.е.

$$\Delta = t\sqrt{\frac{\overline{\sigma^2}}{n}} \quad \text{или} \quad \Delta = t\sqrt{\frac{\overline{w(1-w)}}{n}}.$$

В этом случае ошибка выборки зависит от внутригрупповой вариации.

При *серийной* (гнездовой) выборке, когда из генеральной совокупности, разбитой на определенные равновеликие серии (гнезда), случайно отбираются серии, внутри которых проводится сплошное наблюдение, величина ошибки выборки зависит не от числа обследованных единиц, а от числа обследованных серий (s) и от величины межгрупповой дисперсии $\delta^2 = \frac{\sum\left(\bar{x}_i - \bar{x}\right)^2}{s}$. Серийная выборка в основном проводится как бесповторная, и формула ошибки выборки в этом случае имеет следующий вид:

$$\Delta = t\sqrt{\frac{\delta^2}{s}\left(1-\frac{s}{S}\right)},$$

где δ^2 — межсерийная дисперсия;

s — число отобранных серий;

S — число серий в генеральной совокупности.

Все рассмотренные выше формулы используются при так называемой *большой* выборке.

Если $n < 20$, то выборка именуется *малой** и при расчете ошибок выборки необходимо учитывать следующие моменты. Во-первых, в формуле средней ошибки в знаменателе принимается $n - 1$, т.е.

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

И, во-вторых, при нахождении вероятности допуска той или иной ошибки или определении доверительных интервалов исследуемого показателя в генеральной совокупности пользуются таблицами вероятности Стьюдента (см. Приложение 3), где $P = S(t, n)$ определяется в зависимости от объема выборки и t.

Формулы предельной ошибки выборки позволяют решить следующие три задачи:

1. Определить доверительные пределы:

для генеральной средней

$$\tilde{x} - t\mu \leqslant \bar{x} \leqslant \tilde{x} + t\mu;$$

для доли

$$w - t\mu \leqslant p \leqslant w + t\mu.$$

2. Определить вероятность допуска той или иной заданной ошибки Δ.

В этом случае определяется $t = \dfrac{\Delta}{\mu}$ и по таблице Приложения 2 (при $n > 20$) находится вероятность (P).

3. Определить необходимую численность выборки (n), обеспечивающую с определенной вероятностью заданную точность (Δ).

Формулы для n определяются из соответствующих формул предельной ошибки.

---

* Понятие малой выборки некоторыми авторами распространяется до $n < 30$.

Так, для определения средней ($\tilde{x}$) из формулы $\Delta = t\sqrt{\frac{\sigma^2}{n}}$ при повторном отборе имеем

$$n = \frac{t^2\sigma^2}{\Delta^2}.$$

Для доли аналогично из $\Delta = t\sqrt{\frac{w(1-w)}{n}}$ получаем

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

При бесповторном отборе из $\Delta = t\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$ и $\Delta = t\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}$ имеем:

$$n = \frac{Nt^2\sigma^2}{N\Delta^2 + t^2\sigma^2} \text{ — для средней } (\tilde{x});$$

$$n = \frac{Nt^2 w(1-w)}{N\Delta^2 + t^2 w(1-w)} \text{ — для доли } (w).$$

Как видно, в формулах для определения необходимой численности выборки, получаемых из формул случайной ошибки выборки, предполагается обязательное знание величины дисперсии признака (σ^2) или $[w(1-w)]$.

Обычно в этих формулах используется значение дисперсии признака в аналогичных предшествующих исследованиях или же проводится пробное обследование небольшого числа единиц, для которых определяется значение σ^2. В случае изучения доли определенных единиц в совокупности при отсутствии каких-либо сведений о дисперсии принимается максимальное значение $[w(1-w)]$, равное 0,25.

Рассмотрим решение некоторых задач к этой теме с применением формул предельной ошибки выборки.

Задача 7.1

Методом собственно случайной выборки обследована жирность молока у 100 коров. По данным выборки средняя жирность молока оказалась равной 3 64%, а дисперсия составила 2,56.

Определить: а) среднюю ошибку выборки; б) с вероятностью, равной 0,954, предельные значения генеральной средней.

Решение.

А. Формула средней ошибки выборки: $\mu = \sqrt{\frac{\sigma^2}{n}}$.

По условию $n = 100$, $\sigma^2 = 2{,}56$. Отсюда $\mu = \sqrt{\frac{2{,}56}{100}} = \frac{1{,}6}{10} = 0{,}16\%$.

Б. Формула предельной ошибки выборки: $\Delta = t\,\mu$.

По таблице значений $F(t)$ (см. Приложение 2) при $P = 0{,}954$ находим, что $t = 2$. Отсюда $\Delta = 2 \cdot 0{,}16 = 0{,}32$, или $\bar{x} = \tilde{x} \pm \Delta = 3{,}64 \pm 0{,}32$, т.е. предельные значения жирности молока (или доверительный интервал генеральной средней) определяются как $3{,}32\% \leqslant \bar{x} \leqslant 3{,}96\%$.

Задача 7.2

На основе выборочного обследования 600 рабочих ($n = 600$) одной из отраслей промышленности установлено, что удельный вес численности женщин составил 0,4 ($w = 0{,}4$).

С какой вероятностью можно утверждать, что при определении доли женщин, занятых в этой отрасли, допущена ошибка (Δ), не превышающая 5% (0,05)?

Решение.

Чтобы определить вероятность допуска той или иной ошибки, из формулы $\Delta = t\,\mu$ находим показатель t, связанный с вероятностью:

$$t = \frac{\Delta}{\mu} = \frac{\Delta}{\sqrt{\frac{w(1-w)}{n}}} = \frac{0{,}05}{\sqrt{\frac{0{,}4 \cdot 0{,}6}{600}}} = 2{,}5.$$

По таблице значений $F(t)$ (см. Приложение 2) для $t = 2{,}5$ находим, что $P = 0{,}988$, т.е. с вероятностью 0,988 можно утверждать, что при определении доли женщин (0,4) в общем числе рабочих допущена ошибка не более 0,05 (5%).

Задача 7.3

Сколько рабочих завода нужно обследовать в порядке случайной выборки для определения средней заработной платы, чтобы с вероятностью (P), равной 0,954, можно было бы гарантировать ошибку не более 50 руб.? Предполагаемое среднее квадратическое отклонение заработной платы $\sigma = 200$ руб.

Решение.

Из формулы $\Delta = t\sqrt{\frac{\sigma^2}{n}}$ находим n:

$$n = \frac{t^2\sigma^2}{\Delta^2} = \frac{4 \cdot 200^2}{50^2} = 64 \text{ (человека).}$$

Задача 7.4

Средняя продолжительность горения, установленная путем испытания 10 случайно отобранных электрических лампочек, оказалась равной 1280 ч при среднем квадратическом отклонении 18 ч.

С какой вероятностью можно утверждать, что допущенная при этом предельная ошибка выборки (т.е. расхождение между выборочной и генеральной средней) не превысит 12 ч?

Решение.

Поскольку $n < 20$, имеем дело с малой выборкой. Определяем среднюю ошибку малой выборки:

$$\mu_{\text{м.в}} = \frac{\sigma}{\sqrt{n-1}} = \frac{18}{\sqrt{9}} = 6.$$

Из формулы предельной ошибки выборки находим:

$$t = \frac{\Delta}{\mu} = \frac{12}{6} = 2.$$

Поскольку при малой выборке вероятность наступления той или иной ошибки выборки подчиняется распределению Стьюдента и, в частности, вероятность того, что генеральная средняя находится в определенных границах, определяется по формуле

$$P\,[\tilde{x} - t\mu_{\text{м.в}} < \bar{x} < \tilde{x} + t\mu_{\text{м.в}}] = 2S\,(t,\ n) - 1,$$

обращаемся к соответствующей таблице, где рассчитаны вероятности $S(t)$ (см. таблицу Приложения 3), и находим для заданных n и t (на пересечении) значение $S(t)$, а затем уже рассчитываем $2S(t) - 1$. Так, в нашем примере по таблице Приложения 3 для $n = 10 - 1$ и $t = 2$ получаем $S(t) = 0{,}962$. Отсюда искомая вероятность допуска ошибки не более 12 ч равняется $2 \cdot 0{,}962 - 1 = 0{,}924$.

(Значение n в таблице Приложения 3 принимается на единицу меньше числа наблюдений, т.е. как число степеней свободы. В нашем примере число наблюдений 10, следовательно, в таблице ищем графу с $n = 9$.)

Задача 7.5

Для определения средней заработной платы рабочих завода была произведена 20%-ная бесповторная выборка (по цехам) с отбором единиц пропорционально численности групп. Результаты выборки представлены в приводимой ниже таблице:

Цех	Объем выборки, чел., n_i	Средняя заработная плата, руб., $\tilde{x}_i$	Среднее квадратическое отклонение, руб., σ_i
1	120	873	30
2	100	886	80
3	180	900	60
Всего	400	—	—

С вероятностью 0,997 (т.е. $t = 3$) определить пределы, в которых находится средняя заработная плата всех рабочих завода.

Решение.

А. Находим общую выборочную среднюю заработную плату:

$$x = \frac{\sum \tilde{x}_i n_i}{\sum n_i} = \frac{873 \cdot 120 + 886 \cdot 100 + 900 \cdot 180}{400} = \frac{355\,360}{400} = 888{,}4 \text{ (руб.)}.$$

Б. Находим среднюю из групповых дисперсий:

$$\overline{\sigma^2} = \frac{\sum \sigma_i^2 n_i}{\sum n_i} = \frac{900 \cdot 120 + 6400 \cdot 100 + 3600 \cdot 180}{400} = \frac{1\,396\,000}{400} = 3490.$$

В. Определяем предельную ошибку выборочной средней заработной платы. Для типической бесповторной выборки

$$\Delta = t\sqrt{\frac{\overline{\sigma^2}}{n}\left(1 - \frac{n}{N}\right)} = 3\sqrt{\frac{3490}{400}(1 - 0{,}2)} =$$

$$= \frac{3}{20}\sqrt{3490 \cdot 0{,}8} = \frac{3 \cdot 52{,}84}{20} = 7{,}9.$$

Отсюда генеральная средняя

$$\bar{x} = \tilde{x} \pm \Delta = 888{,}4 \pm 7{,}9 \quad \text{или} \quad 880{,}5 \leqslant \bar{x} \leqslant 896{,}3,$$

т.е. средняя заработная плата всех рабочих находится в пределах от 880,5 руб. до 896,3 руб.

В статистике часто приходится сравнивать результаты двух (или более) выборок. И на основании сравнения двух выборочных средних (или долей) делается вывод о случайности или существенности их расхождений. Для этого абсолютная разность показателей $\left|\tilde{x}_1 - \tilde{x}_2\right|$ сопоставляется со средней ошибкой разности $\mu_{разн} = \sqrt{\mu_1^2 + \mu_2^2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$. Если при $n > 20$ результат этого соотношения $t < 3$, то делается вывод о случайности расхождений. Если же объем выборки мал, т.е. $n < 20$, то полученное значение t (фактическое) сравнивают с табличным, определяемым по таблицам t-распределения Стьюдента при заданном числе степеней свободы и уровне значимости. И если $t_{факт} < t_{табл}$, расхождения можно считать случайными. (Число степеней свободы при этом определяется как $n_1 + n_2 - 2$.)

При $n < 20$ средняя ошибка разности более точно определяется по формуле

$$\mu_{разн} = \sqrt{\frac{\sigma_1^2 n_1 + \sigma_2^2 n_2}{n_1 + n_2 - 2}} \sqrt{\frac{n_1 + n_2}{n_1 n_2}},$$

а

$$t_{факт} = \frac{\left|\tilde{x}_1 - \tilde{x}_2\right| \sqrt{n_1 + n_2 - 2}\,\sqrt{n_1 n_2}}{\sqrt{\sigma_1^2 n_1 + \sigma_2^2 n_2}\,\sqrt{n_1 + n_2}}.$$

Найденное $t_{факт}$, как указано выше, сопоставляется с $t_{табл}$, определяемым по таблице Приложения 9 для числа степеней свободы $\nu = n_1 + n_2 - 2$ и заданного уровня значимости α.

Задача 7.6

Предположим, на предприятии из коллектива рабочих выборочно обследовано 25 мужчин и 25 женщин. Среднемесячная заработная плата мужчин оказалась равна 2830 руб. при среднем квадратическом отклонении 20 руб., а у женщин 2780 руб. при среднем квадратическом отклонении 30 руб. Определить, можно ли считать расхождение между средней заработной платой мужчин и женщин случайным.

Решение.

А. Находим абсолютную разность средних:

$$\left|\tilde{x}_1 - \tilde{x}_2\right| = |2830 - 2780| = 50 \text{ руб.}$$

Б. Средняя ошибка разности

$$\mu_{разн} = \sqrt{\mu_1^2 + \mu_2^2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{20^2}{25} + \frac{30^2}{25}} = \sqrt{16 + 36} = \sqrt{52} = 7{,}2.$$

В. Находим t:

$$t = \frac{\left|\tilde{x}_1 - \tilde{x}_2\right|}{\mu_{разн}} = \frac{50}{7{,}2} \approx 6{,}94.$$

Поскольку в данной задаче в каждой выборке $n > 20$, то $t_{факт}$ сравниваем с 3. Так как $t > 3$, то расхождение между средней заработной платой мужчин и женщин нельзя считать случайным.

Задача 7.7

На одном из рынков города дважды за день проведено выборочное обследование цен на картофель. При первом обследовании было опрошено 10 продавцов, при втором – 15. Средняя цена картофеля в первой выборке оказалась равной 5 руб. при среднем квадратическом отклонении 0,6 руб., а во второй выборке соответственно 5,5 и 0,8 руб. (т.е. $n_1 = 10$ чел., $\tilde{x}_1 = 5$ руб., $\sigma_1 = 0{,}6$ руб. и $n_2 = 15$ чел., $\tilde{x}_2 = 5{,}5$ руб., $\sigma_2 = 0{,}8$ руб.).

Определить, случайны или нет расхождения между $\tilde{x}_1$ и $\tilde{x}_2$.

Решение.

1. Находим абсолютную разность двух выборочных средних:

$$\left|\tilde{x}_1 - \tilde{x}_2\right| = \left|5 - 5{,}5\right| = 0{,}5 \text{ руб.}$$

2. Так как n в каждой выборке меньше 20, то среднюю ошибку разности определяем по формуле

$$\mu_{разн} = \sqrt{\frac{\sigma_1^2 n_1 + \sigma_2^2 n_2}{n_1 + n_2 - 2}} \sqrt{\frac{n_1 + n_2}{n_1 n_2}} =$$

$$= \sqrt{\frac{0{,}36 \cdot 10 + 0{,}64 \cdot 15}{10 + 15 - 2}} \sqrt{\frac{10 + 15}{10 \cdot 15}} = 0{,}53.$$

3. Рассчитываем $t_{факт}$:

$$t_{факт} = \frac{\left|\tilde{x}_1 - \tilde{x}_2\right|}{\mu_{разн}} = \frac{0{,}5}{0{,}53} = 0{,}94.$$

4. По таблице Приложения 9 находим $t_{табл}$ при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы $\nu = n_1 + n_2 - 2 = 10 + 15 - 2 = 23$: $t_{табл} = 2{,}0687$.

5. Сравниваем $t_{факт}$ с $t_{табл}$. Так как $t_{факт} < t_{табл}$ (0,94 < 2,0687), то делаем вывод о том, что расхождение цен на картофель в двух выборках можно считать случайным.

Литература

[1, с. 136—187], [2, с. 230—270], [3, с. 132—183].

Контрольные вопросы

1. Что такое выборочное наблюдение и в каких случаях к нему прибегают? Каковы теоретические основы выборочного метода?
2. Какие существуют способы отбора (виды выборки)?
3. От чего зависит точность выборки?
4. Что такое повторная и бесповторная выборки?
5. Как рассчитать среднюю и предельную ошибку выборки (для средней и для доли)?
6. Как рассчитывается вероятность той или иной ошибки выборки?
7. Как рассчитать необходимую численность выборки, обеспечивающую ту или иную точность выборки?
8. В чем особенность определения ошибок выборки при так называемой малой выборке?
9. Как оценивается случайность (существенность) расхождений двух выборочных средних или долей?

Задачи и упражнения для самостоятельной работы

Задача 1

Из партии готовой продукции в порядке механической выборки проверено 50 лампочек на продолжительность горения. Последняя оказалась равна 840 ч при среднем квадратическом отклонении 60 ч.

Определить:

1) среднюю ошибку (μ) выборочной средней продолжительности горения лампочки;
2) с вероятностью 0,95 доверительные пределы продолжительности горения лампочки в генеральной совокупности.

Ответ: 1) μ = 8,5 ч; 2) 823,3 ч $\leqslant \bar{x} \leqslant$ 856,7 ч.

Задача 2

На городской телефонной станции в порядке собственно случайной выборки проведено 100 наблюдений и установлено, что средняя продолжительность одного телефонного разговора составляет 10 мин при среднем квадратическом отклонении 5 мин.

1. С вероятностью 0,997 определить доверительные пределы для генеральной средней.
2. Можно ли считать данную выборку репрезентативной?

Ответ: 1) 8,5 мин $\leqslant \bar{x} \leqslant$ 11,5 мин;
2) нет, так как относительная ошибка выборки $\Delta_{отн} > 5\%$.

Задача 3

По данным задачи 2 ответить на вопрос: с какой вероятностью можно утверждать, что при определении средней продолжительности одного телефонного разговора допущена ошибка, не превышающая 1 мин.

Ответ: 0,9545.

Задача 4

Из партии готовой продукции в порядке механической бесповторной выборки проверено 400 изделий и установлено, что 80% из них соответствует первому сорту.

С вероятностью 0,9545 определить долю (процент) продукции первого сорта во всей партии.

Задачу решить в двух вариантах:

1) численность изделий в партии готовой продукции неизвестна;
2) в партии готовой продукции 2000 изделий.

Ответ: 1) $76\% \leqslant p \leqslant 84\%$; 2) $76{,}4\% \leqslant p \leqslant 83{,}6\%$.

Задача 5

В порядке случайной выборки обследован дневной надой молока 50 коров. Результаты обследования приведены в таблице.

Дневная удойность, кг	Количество коров
10—14	5
14—18	15
18—22	20
Свыше 22	10
Итого	**50**

Определить:

1) по выборочным данным средний дневной надой молока от одной коровы;
2) среднюю ошибку выборки;
3) вероятность того, что при определении выборочного среднего надоя молока допущена ошибка, не превышающая 1 кг.

Ответ: 1) $\tilde{x}$ = 18,8 кг; 2) μ = 0,51 кг; 3) P = 0,95.

Задача 6

Намечается провести выборочное обследование покупателей в одном из крупных универмагов города в целях определения доли покупателей из других городов. Каким должен быть объем выборки, чтобы с вероятностью 0,9545 можно было бы гарантировать точность результата до 5%.

О т в е т: 400 чел.

Задача 7

На предприятии выборочно проверен стаж работы у 12 мужчин и 8 женщин. Результаты наблюдения следующие:

Группа рабочих	n_i	Средний стаж работы, лет, $\tilde{x}_i$	Среднее квадратическое отклонение стажа, лет, σ_i
Мужчины	12	14	3
Женщины	8	11	2

1. Рассчитать общий средний стаж работы для рабочих по выборочным данным.
2. С вероятностью 0,954 определить доверительные пределы среднего стажа работы рабочих в генеральной совокупности.

О т в е т: 1) $\tilde{x}_{общ} = 12{,}8$ года; 2) 11,6 года $\leqslant \bar{x} \leqslant$ 14,0 года.

Задача 8

По данным задачи 7 определить, можно ли считать расхождения в значениях выборочной средней стажа работы у мужчин и женщин (14 лет и 11 лет) случайными (на уровне значимости $\alpha = 0{,}05$ и при числе степеней свободы $\nu = n_1 + n_2 - 2 = 18$).

Задача 9

Для определения среднего процента выполнения норм выработки проведена 5%-ная типическая выборка из трех групп рабочих с разным стажем. Результаты следующие:

Группы рабочих со стажем	Объем выборки, чел., n_i	Средний процент выполнения норм $\tilde{x}_i$	Среднее квадратическое отклонение, %, σ_i
1—2 года	15	98	3
3—5 лет	20	102	2
Более 5 лет	65	104	4

Определить:

1) средний процент выполнения норм для всех рабочих в выборочной совокупности;
2) вероятность того, что выборочная средняя (процент выполнения норм) отличается от генеральной не более чем на 1%.

Ответ: 1) $\tilde{x} = 102,7$; 2) $P = 0,996$.

Задача 10

Имеются следующие данные выборочного обследования размера заработной платы у 5 мужчин и 5 женщин в одном из НИИ:

Заработная плата, руб.	
женщин	мужчин
2500	2660
2520	2700
2660	2730
2670	2840
2700	2920

Определить:

1) среднюю заработную плату отдельно для мужчин и для женщин, а также общую среднюю, т.е. для всей выборочной совокупности;
2) среднюю ошибку выборки для общей средней заработной платы;
3) вероятность того, что общая выборочная средняя заработная плата отличается от генеральной не более чем на 50 руб.;
4) можно ли считать случайными расхождения между средней заработной платой мужчин и женщин?

Ответ: 1) $\tilde{x}_{м} = 2770$ руб., $\tilde{x}_{ж} = 2610$ руб., $\tilde{x}_{общ} = 2690$ руб.;
2) $\mu = 29,89$ руб.; 3) $P = 0,876$.

Задача 11

Выборочное обследование 10 электрических лампочек для определения их средней продолжительности горения дало следующее распределение:

Продолжительность горения, ч	Количество лампочек
900—920	2
920—940	3
940—960	4
960—980	1

Определить:

1) среднюю продолжительность горения лампочек по выборочным данным;
2) с вероятностью 0,97 оценить доверительные пределы для генеральной средней.

Ответ: 1) $\tilde{x} = 938$ ч; 2) $922,1 \leqslant \bar{x} \leqslant 953,9$.

Задача 12

Имеются следующие данные по результатам выборочной 5%-ной микропереписи населения РФ в 1994 г.:

Северо-Западный район	Обследовано тыс. человек n_i	Доля лиц старше трудоспособного возраста, % w_i
г. Санкт-Петербург	238	20,6
Ленинградская обл.	82	23,1
Новгородская обл.	38	25,5
Псковская обл.	42	26,8

С вероятностью $P = 0{,}954$ определить долю лиц старше трудоспособного возраста в целом по Северо-Западному району.

О т в е т: $\bar{w} = 22{,}2\%$ (0,222), $22{,}07\% \leqslant p \leqslant 22{,}33\%$.

Задача 13

Какой должна быть необходимая численность выборки (при механическом отборе), проводимой с целью определения доли единиц в генеральной совокупности, обладающих тем или иным признаком, чтобы с вероятностью 0,954 можно было гарантировать ошибку доли не более 3%. Дисперсия доли неизвестна.

Ответ дать в предположении, что объем генеральной совокупности содержит: а) 1000 единиц; б) 10 000 единиц; в) численность неизвестна.

Тема 8
КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ

При исследовании социально-экономических явлений часто приходится иметь дело со взаимосвязанными показателями. При этом часто связь, существующая между двумя или несколькими показателями, затушевывается, усложняется наслоением действия других причин (факторов). Изучить, насколько изменение одного показателя зависит от изменения другого (или нескольких), — одна из важнейших задач статистики.

Следует различать функциональные и корреляционные связи. В отличие от **функциональной** зависимости, при которой каждому *значению одной переменной строго соответствует одно определенное* значение другой переменной, зависимость, при которой одному значению переменной (x) может соответствовать (в силу наслоения действия других причин) множество значений другой переменной (y), называют **корреляционной**. Корреляционная зависимость проявляется лишь на основе массового наблюдения.

Примером корреляционной зависимости может служить зависимость производительности труда от стажа работы рабочих, зависимость урожайности от срока сева, зависимость годового удоя коров от количества отелов и т.п.

Наиболее простым случаем корреляционной зависимости является *парная* корреляция, т.е. зависимость между двумя признаками (результативным и одним из факторных).

Основными задачами при изучении корреляционных зависимостей являются: 1) отыскание математической формулы, которая бы выражала эту зависимость y от x; 2) измерение тесноты такой зависимости.

Решение первой задачи, т.е. определение формы связи с последующим отысканием параметров уравнения, называется нахождением уравнения связи (уравнения *регрессии*). Показатели, рассматриваемые как функция x, обозначают $\overline{y_x}$ (читается: «игрек, выравненный по икс»).

Возможны различные формы связи:

1) прямолинейная: $\overline{y_x} = a_0 + a_1x$;
2) криволинейная в виде:
   а) параболы второго порядка $\overline{y_x} = a_0 + a_1x + a_2x^2$ (или высших порядков);
   б) гиперболы $\overline{y_x} = a_0 + \frac{a_1}{x}$;
   в) показательной функции $\overline{y_x} = a_0a_1^x$ и т.д.

Параметры для всех уравнений связи чаще всего определяют из так называемой **системы нормальных уравнений**, отвечающих требованию «метода наименьших квадратов» (МНК). Это требование можно записать как $\sum(y - \overline{y_x})^2 \to \min$ или, при линейной зависимости, $\overline{y_x} = a_0 + a_1x$, $\sum(y - a_0 - a_1x)^2 \to \min$, т.е. требуется определить, при каких значениях параметров a_0 и a_1 сумма квадратов отклонений y от $\overline{y_x}$ будет минимальной. Найдя частные производные указанной суммы по a_0 и a_1 и приравняв их к нулю, легко записать систему уравнений, решение которой и дает параметры искомой функции, т.е. уравнения регрессии.

Так, система нормальных уравнений при линейной зависимости имеет вид

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

Если связь выражена параболой второго порядка

$$\overline{y_x} = a_0 + a_1x + a_2x^2,$$

то система нормальных уравнений для отыскания параметров a_0, a_1, a_2 выглядит следующим образом:

$$\begin{cases} na_0 + a_1\sum x + a_2\sum x^2 = \sum y \\ a_0\sum x + a_1\sum x^2 + a_2\sum x^3 = \sum xy \\ a_0\sum x^2 + a_1\sum x^3 + a_2\sum x^4 = \sum x^2 y. \end{cases}$$

Вторая задача — измерение тесноты зависимости — для всех форм связи может быть решена с помощью исчисления теоретического корреляционного отношения (η):

$$\eta = \frac{\delta}{\sigma} = \sqrt{\frac{\delta^2}{\sigma^2}},$$

где $\delta^2 = \dfrac{\sum\left(\overline{y_x} - \overline{y}\right)^2}{n}$ — дисперсия в ряду выравненных значений результативного показателя $\overline{y_x}$;

$\sigma^2 = \dfrac{\sum\left(y - \overline{y}\right)^2}{n}$ — дисперсия в ряду фактических значений y.

Так как дисперсия δ^2 отражает вариацию в ряду $\overline{y_x}$ только за счет вариации фактора x, а дисперсия σ^2 отражает вариацию y за счет всех факторов, то их отношение, именуемое **теоретическим коэффициентом детерминации**, показывает, какой удельный вес в общей дисперсии ряда y занимает дисперсия, вызываемая вариацией фактора x. Квадратный корень из отношения этих дисперсий дает нам теоретическое корреляционное отношение. Если $\delta^2 = \sigma^2$, то это означает, что роль других факторов в вариации y сведена на нет, и отношение $\eta = \frac{\delta}{\sigma} = 1$ означает полную зависимость вариации y от x. Если $\delta^2 = 0$, то это означает, что вариация x никак не влияет на вариацию y, и в этом случае $\eta = \frac{\delta}{\sigma} = 0$. Следовательно, максимальное значение, которое может принимать корреляционное отношение, равно 1, минимальное значение — 0.

Математически легко доказывается, что в случае линейной зависимости корреляционное отношение $\eta = \frac{\delta}{\sigma}$ может быть заменено выражением $a_1 \frac{\sigma_x}{\sigma_y}$, которое называют **линейным коэффициентом корреляции** и обозначают r, т.е. $r = a_1 \frac{\sigma_x}{\sigma_y}$, где a_1 — коэффициент регрессии в уравнении связи, σ_x и σ_y — соответственно среднее квадратическое отклонение в ряду x и в ряду y.

Линейный коэффициент корреляции можно выразить и другими формулами, тождественными первой, в частности:

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

или

$$r = \frac{\sum(x - \overline{x})(y - \overline{y})}{\sqrt{\sum(x - \overline{x})^2 \sum(y - \overline{y})^2}},$$

а также

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

Линейный коэффициент корреляции может принимать по модулю значения от 0 до 1 (знак «+» при прямой зависимости и знак «–» при обратной зависимости).

Рассмотрим решение некоторых задач по этой теме.

Задача 8.1

Пусть по 10 однотипным предприятиям имеются следующие данные о выпуске продукции (x) в тыс. ед. и о расходе условного топлива (y) в тоннах (графы 1 и 2 таблицы).

Требуется найти уравнение зависимости расхода топлива от выпуска продукции (или уравнение регрессии y по x) и измерить тесноту зависимости между ними.

Решение.

А. Рассматривая уравнение регрессии в форме линейной функции вида $\overline{y_x} = a_0 + a_1 x$, параметры данного уравнения (a_0 и a_1) найдем из системы нормальных уравнений

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

x	y	x^2	xy	$\overline{y_x} = 1{,}16 + 0{,}547x$	y^2
1	2	3	4	5	6
5	4	25	20	3,9	16
6	4	36	24	4,4	16
8	6	64	48	5,5	36
8	5	64	40	5,5	25
10	7	100	70	6,6	49
10	8	100	80	6,6	64
14	8	196	112	8,8	64
20	10	400	200	12,1	100
20	12	400	240	12,1	144
24	16	576	384	14,3	256
125	**80**	**1961**	**1218**	**80**	**770**

Необходимые для решения суммы $\sum x$, $\sum y$, $\sum x^2$, $\sum xy$ рассчитаны выше в таблице. Подставляем их в уравнения и решаем систему:

$$\begin{cases} 10a_0 + 125a_1 = 80 \\ 125a_0 + 1961a_1 = 1218, \end{cases}$$

$$a_1 = \frac{n\sum xy - \sum x \cdot \sum y}{n\sum x^2 - (\sum x)^2} = \frac{10 \cdot 1218 - 125 \cdot 80}{10 \cdot 1961 - (125)^2} = \frac{2180}{3985} = 0{,}547;$$

$$a_0 = \overline{y} - a_1 \overline{x} = \frac{80}{10} - 0{,}547 \cdot \frac{125}{10} = 8 - 6{,}84 = 1{,}16;$$

$$a_0 = 1{,}16;\ a_1 = 0{,}547.$$

Отсюда $\overline{y_x} = 1{,}16 + 0{,}547x$.

Подставляя в это уравнение последовательно значения $x = 5$, 6, 8, 10 и т.д., получаем выравненные (теоретические) значения результативного показателя $\overline{y_x}$ (графа 5 таблицы).

Поскольку параметры уравнения регрессии являются оценочными, то для каждого из них рассчитывается средняя ошибка, т.е. μ_{a_i}.

Конкретный расчет ошибок для a_0 и a_1 по данным нашего примера приведен далее (см. с. 83).

Б. Для измерения тесноты зависимости между y и x воспользуемся прежде всего линейным коэффициентом корреляции (поскольку зависимость рассматривалась линейной):

а) применяем формулу $r = \dfrac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \sigma_y}$.

Находим $\overline{xy} = 121{,}8$; $\overline{x} = 12{,}5$; $\overline{y} = 8$; $\overline{x^2} = 196{,}1$.

Определяем σ_x и σ_y, предварительно найдя $\sum y^2 = 770$ и $\overline{y^2} = 77$:

$$\sigma_x = \sqrt{\overline{x^2} - \left(\overline{x}\right)^2} = \sqrt{196{,}1 - 12{,}5^2} = \sqrt{196{,}1 - 156{,}25} = \sqrt{39{,}85} = 6{,}31.$$

$$\sigma_y = \sqrt{\overline{y^2} - \left(\overline{y}\right)^2} = \sqrt{77 - 8^2} = \sqrt{13} = 3{,}6.$$

Отсюда $r = \dfrac{121{,}8 - 12{,}5 \cdot 8}{6{,}31 \cdot 3{,}6} = \dfrac{21{,}8}{22{,}716} = 0{,}96.$

Значение линейного коэффициента корреляции $r = 0{,}96$ (т.е. близкое к единице) характеризует не только меру тесноты зависимости вариации y от вариации x, но и степень близости этой зависимости к линейной;

б) воспользуемся еще одной формулой линейного коэффициента корреляции:

$$r = a_1 \frac{\sigma_x}{\sigma_y} = 0{,}547 \cdot \frac{6{,}31}{3{,}6} = 0{,}96,$$

т.е. результат тот же.

При расчете коэффициента корреляции очень важно оценить его значимость. Оценка значимости (существенности) линейного коэффициента корреляции основана на сопоставлении значения r с его средней квадратической ошибкой (σ_r).

Средняя ошибка коэффициента корреляции при $n > 50$ рассчитывается приближенно по формуле

$$\sigma_r = \frac{1 - r^2}{\sqrt{n}}.$$

Если при этом коэффициент корреляции r превышает свою среднюю ошибку σ_r больше чем в 3 раза, т.е. если $\frac{|r|}{\sigma_r} > 3$, то он считается значимым, а связь – реальной.

При $n < 30$ значимость коэффициента корреляции проверяется на основе t-критерия Стьюдента. Для этого рассчитывается фактическое (расчетное) значение критерия:

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

которое сопоставляется с $t_{\text{табл}}$, определяемым по Приложению 9, для числа степеней свободы $\nu = n - 2$ и заданного уровня значимости (обычно $\alpha = 0{,}05$).

Если $t_{\text{факт}} > t_{\text{табл}}$, r считается значимым, а связь – реальной. Если $t_{\text{факт}} < t_{\text{табл}}$, то считается, что связь между x и y отсутствует и значение r, отличное от нуля, получено случайно.

В рассматриваемом примере средняя ошибка коэффициента корреляции

$$\sigma_r = \frac{\sqrt{1-r^2}}{\sqrt{n-2}} = \frac{\sqrt{1-0{,}96^2}}{\sqrt{10-2}} = \frac{\sqrt{1-0{,}9216}}{\sqrt{8}} = \frac{\sqrt{0{,}0784}}{2{,}83} = 0{,}1,$$

а $t_{\text{факт}} = \frac{r}{\sigma_r} = \frac{0{,}96}{0{,}1} = 9{,}6.$

По таблице Приложения 9 находим, что при числе степеней свободы $\nu = 10 - 2 = 8$ и уровне значимости $\alpha = 0{,}05$ табличное (критическое, пороговое) t равно 2,306, т.е. $t_{\text{табл}} = 2{,}306$.

Поскольку фактическое (расчетное) t больше табличного, т.е. $t_{\text{факт}} > t_{\text{табл}}$, то линейный коэффициент корреляции $r = 0{,}96$ считается значимым, а связь между x и y — реальной.

В. Кроме линейного коэффициента корреляции для измерения тесноты зависимости можно воспользоваться теоретическим корреляционным отношением:

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}} = \sqrt{\frac{D_{\bar{y}_x}}{D_y}},$$

где δ^2 и σ^2 — дисперсии соответственно теоретических и эмпирических значений результативного показателя.

Расчет их показан ниже в таблице.

x	y	$\bar{y}_x$	$y-\bar{y}$	$(y-\bar{y})^2$	$\bar{y}_x-\bar{y}$	$(\bar{y}_x-\bar{y})^2$	$y-\bar{y}_x$	$(y-\bar{y}_x)^2$
1	2	3	4	5	6	7	8	9
5	4	3,9	–4	16	–4,1	16,81	0,1	0,01
6	4	4,4	–4	16	–3,6	12,96	–0,4	0,16
8	6	5,5	–2	4	–2,5	6,25	0,5	0,25
8	5	5,5	–3	9	–2,5	6,25	–0,5	0,25
10	7	6,6	–1	1	–1,4	1,96	0,4	0,16
10	8	6,6	0	0	–1,4	1,96	1,4	1,96
14	8	8,8	0	0	0,8	0,64	–0,8	0,64
20	10	12,1	2	4	4,1	16,81	–2,1	4,41
20	12	12,1	4	16	4,1	16,81	–0,1	0,01
24	16	14,3	8	64	6,3	39,69	1,7	2,89
125	**80**	**80**	—	**130**	—	**120,14**	—	**10,74**

Дисперсия выравненных значений результативного показателя, или факторная дисперсия, $\delta^2 = \frac{\sum(\bar{y}_x - \bar{y})^2}{n} = \frac{120,14}{10} = 12,014$; общая дисперсия эмпирических значений результативного показателя $\sigma^2 = \frac{\sum(y-\bar{y})^2}{n} = \frac{130}{10} = 13$; теоретическое корреляционное отношение $\eta = \sqrt{\frac{\delta^2}{\sigma^2}} = \sqrt{\frac{12,014}{13}} = \sqrt{0,924} = 0,96$.

Рассчитанные показатели позволяют сделать вывод о том, что связь между вариацией результативного показателя (y) и факторного (x) весьма высокая.

Вместо дисперсии выравненных значений y, т.е. δ^2, можно воспользоваться ***остаточной дисперсией***. В соответствии с правилом сложения дисперсий можно записать, что $\delta^2 = \sigma^2 - \sigma^2_{ост}$, где $\sigma^2_{ост} = \frac{\sum(y - \bar{y}_x)^2}{n}$, тогда $\eta = \sqrt{\frac{\sigma^2 - \sigma^2_{ост}}{\sigma^2}} = \sqrt{1 - \frac{\sigma^2_{ост}}{\sigma^2}}$.

В нашем примере расчет остаточной дисперсии ($\sigma^2_{ост}$) показан в графах 8 и 9 таблицы: $\sigma^2_{ост} = 1{,}074$. Отсюда

$$\eta = \sqrt{1 - \frac{1{,}074}{13}} = 0{,}96.$$

Остаточная дисперсия, вернее, корень квадратный из нее, т.е. $\sigma_{ост}$, используется для расчета средних ошибок параметров уравнения регрессии. Так, средняя ошибка параметра a_0 равна $\mu_{a_0} = \frac{\sigma_{ост}}{\sqrt{n-2}}$, а для a_1 равна $\mu_{a_1} = \frac{\sigma_{ост}}{\sigma_x\sqrt{n-2}}$.

Сопоставляя значение параметра с его средней ошибкой, по значению $t = \frac{a_i}{\mu_{a_i}}$ судят о значимости данного параметра. Если число наблюдений $n > 20$, то параметр считается значимым при $t > 3$.

Если $n < 20$, то обращаются к специальным таблицам значений t-критерия Стьюдента (см. Приложение 9). И в данном случае параметр считается значимым, если $t_{факт} > t_{табл}$.

В рассмотренном на с. 79, 80 примере для уравнения регрессии $\bar{y}_x = 1{,}16 + 0{,}547x$ ошибки параметров

$$\mu_{a_0} = \frac{\sigma_{ост}}{\sqrt{n-2}} = \frac{\sqrt{1{,}074}}{\sqrt{10-2}} = \frac{1{,}036}{2{,}828} = 0{,}366, \text{ т.е. } a_0 = 1{,}16 \pm 0{,}366,$$

$$\mu_{a_1} = \frac{\sigma_{ост}}{\sigma_x\sqrt{n-2}} = \frac{1{,}036}{6{,}31\sqrt{8}} = 0{,}058, \text{ т.е. } a_1 = 0{,}547 \pm 0{,}058.$$

Чтобы сделать вывод о значимости параметров, находим

$$t_{a_0} = \frac{1{,}16}{0{,}366} = 3{,}2 \quad \text{и} \quad t_{a_1} = \frac{0{,}547}{0{,}058} = 9{,}43.$$

Так как $n < 20$, обращаемся к таблице значений t-критерия (Приложение 9). Для $\alpha = 0{,}05$ и $\nu = 10 - 2 = 8$ находим $t_{табл} = 2{,}306$. Поскольку $t > 2{,}306$ и для a_0, и для a_1, то считаем параметры значимыми.

В приведенных формулах ошибок параметров μ_{a_0} и μ_{a_1} вместо остаточной дисперсии $\sigma^2_{ост}$ можно пользоваться значением линейного коэффициента корреляции, исходя из следующих преобразований:

$$\sigma^2_{ост} = \sigma^2_y - \delta^2;$$

разделив обе части равенства на σ^2_y, получим $\dfrac{\sigma^2_{ост}}{\sigma^2_y} = 1 - r^2$. Отсюда

$\sigma_{ост} = \sigma_y\sqrt{1-r^2}$. Тогда $\mu_{a_0} = \dfrac{\sigma_y\sqrt{1-r^2}}{\sqrt{n-2}}$, а $\mu_{a_1} = \dfrac{\sigma_y\sqrt{1-r^2}}{\sigma_x\sqrt{n-2}}$.

Наряду с проверкой значимости отдельных параметров можно оценить значимость уравнения регрессии в целом.

Эта задача решается на основе расчета F-критерия Фишера и сопоставления его с табличным (критическим). F-критерий представляет собой отношение факторной дисперсии ($\delta^2_ф$) к остаточной ($\sigma^2_{ост}$), каждая из которых рассчитана на одну степень свободы:

$$F = \frac{\delta^2_ф/(m-1)}{\sigma^2_{ост}/(n-m)} = \frac{\delta^2_ф}{\sigma^2_{ост}}\cdot\frac{n-m}{m-1}, \tag{5}$$

где m — число параметров в уравнении регрессии;
$(m-1)$ — число степеней свободы для факторной дисперсии;
n — число наблюдений;
$(n-m)$ — число степеней свободы для остаточной дисперсии.

Часто формула F-критерия записывается в виде

$$F = \frac{\delta^2_ф/k}{\sigma^2_{ост}/(n-k-1)} = \frac{\delta^2_ф}{\sigma^2_{ост}}\cdot\frac{n-k-1}{k}, \tag{6}$$

где k — число коэффициентов регрессии, которое на единицу меньше числа параметров, т.е. $k = m - 1$.

F-критерий можно рассчитать и через коэффициент корреляции (r или η):

$$F = \frac{r^2}{1-r^2}\cdot\frac{n-m}{m-1} \quad \left(\text{или } F = \frac{r^2}{1-r^2}\cdot\frac{n-k-1}{k}\right).$$

Расчетное F сопоставляется с табличным, определяемым по Приложению 8 для числа степеней свободы $\nu_1 = m - 1$ и $\nu_2 = n - m$ при заданном уровне значимости (например, $\alpha = 0{,}05$).

Если $F_{расч} > F_{табл}$, уравнение считается значимым.

Для рассмотренного выше примера $\bar{y}_x = 1{,}16 + 0{,}547x$

$$F_{расч} = \frac{\delta^2_{ф}}{\sigma^2_{ост}} \frac{n-m}{m-1} = \frac{12{,}014}{1{,}074} \frac{10-2}{2-1} = 89{,}5.$$

По Приложению 8 находим $F_{табл}$: при $\alpha = 0{,}05$ и $\nu_1 = 2 - 1 = 1$, $\nu_2 = 10 - 2 = 8$ $F_{табл} = 5{,}32$.

Так как $F_{расч} > F_{табл}$, то уравнение значимо.

Для измерения тесноты зависимости кроме перечисленных выше существуют и другие показатели. В частности, широко используются так называемые **ранговые** коэффициенты корреляции (или коэффициенты корреляции рангов), когда коррелируются не сами значения показателей x и y, а их **ранги**, т.е. номера их мест, занимаемых в каждом ряду значений по возрастанию или убыванию (обозначаются ранги буквой R или N). Коэффициент корреляции рангов Спирмэна (ρ) рассчитывается по формуле

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)},$$

где $d = N_x - N_y$, т.е. разность рангов каждой пары значений x и y, а n — число наблюдений.

Коэффициент корреляции рангов Кендэла (τ) определяется по формуле

$$\tau = \frac{S}{\frac{n(n-1)}{2}} = \frac{2S}{n(n-1)}.$$

Порядок расчета этого показателя следующий.

1. Значения x и y ранжируются, т.е. определяются N_x и N_y.
2. Значения N_x записываются строго в порядке возрастания (или, наоборот, убывания): 1, 2, ..., n.
3. Ранги второго показателя (N_y) располагаются в порядке, соответствующем значению x в исходных данных.
4. Для каждого значения N_y подсчитывается число следующих за ним рангов более высокого порядка. Общая сумма таких случаев «правильного следования» последовательно для всех рангов учитывается как баллы со знаком «+» и обозначается буквой P.
5. Аналогично для каждого значения N_y последовательно подсчитывается число следующих за ним рангов, меньших по

значению. Общая сумма таких случаев (инверсий) учитывается как баллы со знаком «–» и обозначается символом Q.

6. Определяется общая сумма баллов, которая обозначается символом S, т.е. $S = P + Q$.
7. Полученная сумма (S) сопоставляется с максимальной, которая равна $\frac{n(n-1)}{2}$ в случае, если в обоих рядах ранги следуют строго последовательно от 1 до n.

Рассмотрим расчет ранговых коэффициентов корреляции Спирмэна и Кендэла на конкретном примере.

Задача 8.2

По данным 10 предприятий (графы 1 и 2 приводимой ниже таблицы) с помощью коэффициентов корреляции рангов Спирмэна (ρ) и Кендэла (τ) измерить тесноту зависимости между объемом выпуска продукции (y), млн. руб., и стоимостью основных производственных фондов (x), млн. руб.

x	y	N_x	N_y	$d = N_x - N_y$	d^2	Подсчет баллов	
						«+»	«–»
1	2	3	4	5	6	7	8
1,5	3,9	1	3	–2	4	7	2
1,8	4,4	2	5	–3	9	5	3
2,0	3,8	3	2	1	1	6	1
2,2	3,5	4	1	3	9	6	0
2,3	4,8	5	6	–1	1	3	1
2,6	4,3	6	4	2	4	4	0
3,0	7,0	7	9	–2	4	1	2
3,1	6,5	8	8	0	0	1	1
3,5	6,1	9	7	2	4	1	0
3,8	8,2	10	10	0	0	—	—
					$\sum d^2 = 36$	$P = 35$	$Q = -10$

Решение.

А. Для расчета коэффициента корреляции рангов Спирмэна (ρ) вначале ранжируем значения признаков в каждом ряду, т.е. каждому значению x и y в порядке их возрастания присваиваем порядковый номер (ранг) N_x и N_y (графы 3 и 4 таблицы), затем находим разности рангов (d), возводим их в квадрат (графа 6 таблицы) и суммируем. *Полученную сумму $\sum d^2$ подставляем в формулу*

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \cdot 36}{10 \cdot 99} = 0{,}78.$$

Судя по значению полученного коэффициента, связь между x и y довольно большая.

Б. Для расчета коэффициента корреляции рангов Кендэла $\tau = \frac{2S}{n(n-1)}$ определяем S в соответствии с описанным выше порядком как сумму положительных (P) и отрицательных (Q) баллов.

Вспомогательные расчеты этих баллов показаны в графах 7 и 8 таблицы. Так как значения рангов x идут строго в возрастающем порядке, то следим лишь за поведением рангов y. Например, после первой пары значений рангов, где $N_y = 3$, в семи случаях идут значения $N_y > 3$, а в двух случаях значения $N_y < 3$ ($N_y = 2, 1$); после второй пары, где $N_y = 5$, наблюдается пять случаев рангов y выше рассматриваемого, а три ($N_y = 2, 1, 4$) ниже и т. д.

По результатам подсчетов находим общую сумму баллов

$$S = P + Q = 35 - 10 = 25.$$

Подставляя ее в формулу коэффициента корреляции рангов Кендэла (τ), определяем

$$\tau = \frac{2S}{n(n-1)} = \frac{2 \cdot 25}{10 \cdot 9} = \frac{50}{90} = 0{,}44.$$

Коэффициент Кэндэла всегда меньше по значению, чем коэффициент Спирмэна $\left(\tau \approx \frac{2}{3}\rho\right)$.

Интерпретация значений ранговых коэффициентов корреляции аналогична любым другим, т.е. чем ближе значение ρ или τ к 1, тем теснее зависимость, а близость к нулю означает отсутствие связи или весьма малую зависимость.

Рассмотрим еще пример, где ранги повторяются.

<u>Задача 8.3</u>

По данным 10 хозяйств с помощью коэффициентов корреляции рангов Спирмэна и Кендэла измерить тесноту зависимости между урожайностью картофеля и количеством внесенных минеральных удобрений (графы 1 и 2 таблицы).

В данном примере отдельные значения x и y повторяются. При ранжировании повторяющихся значений им присваивается ранг, рассчитанный как средняя арифметическая из суммы мест, которые они занимают по возрастанию.

Удобрения, кг/га, x	Картофель, ц/га, y	N_x	N_y	$d = N_x - N_y$	d^2	Подсчет баллов	
							←→
1	2	3	4	5	6	7	8
140	135	1	1,5	–0,5	0,25	8	0
148	135	2	1,5	0,5	0,25	8	0
150	182	3,5	4	–0,5	0,25	6	0
150	175	3,5	3	0,5	0,25	6	0
185	200	5	6	–1	1,00	3	0
190	200	6	6	0	0	3	0
202	200	7	6	1	1,00	3	0
220	210	8,5	8	0,5	0,25	1	0
220	265	8,5	10	–1,5	2,25	0	1
240	250	10	9	1,0	1,00	—	—
					6,5	$P = 38$	$Q = -1$

Расчет рангов показан в графах 3 и 4.

Для случая повторяющихся рангов есть особые скорректированные формулы и для коэффициента Спирмэна, и для коэффициента Кендэла.

Однако на практике часто пользуются приведенной выше формулой Спирмэна и для случая повторяющихся рангов, поскольку ошибку она дает весьма малую.

В нашем примере $\rho = 1 - \dfrac{6\sum d^2}{n(n^2-1)} = 1 - \dfrac{6 \cdot 6{,}5}{10 \cdot 99} = 0{,}96.$

Формула коэффициента Кендэла для повторяющихся рангов имеет вид

$$\tau = \frac{S}{\sqrt{\left(\dfrac{n(n-1)}{2} - U_x\right)\left(\dfrac{n(n-1)}{2} - U_y\right)}},$$

где $S = P + Q$, как и раньше, а U_x и U_y — показатели, корректирующие максимальную сумму баллов и определяемые по формуле $\dfrac{\sum t(t-1)}{2}$, где t — число повторяющихся рангов в соответствующем ряду x и y.

В нашем примере

$$U_x = \frac{\sum t(t-1)}{2} = \frac{2 \cdot 1 + 2 \cdot 1}{2} = 2, \quad U_y = \frac{2 \cdot 1 + 3 \cdot 2}{2} = 4.$$

Баллы Q и P определяются так же, как и в задаче 8.2 с тем лишь добавлением, что в случае одинакового (повторяющегося) значения ранга, следующего за рассматриваемыми в любом из рядов (x и y), последний при подсчете баллов не учитывается ни со знаком «+», ни со знаком «–».

Расчет P и Q показан в графах 7 и 8 таблицы; по результатам подсчетов $S = P + Q = 38 - 1 = 37$.

Отсюда коэффициент корреляции рангов Кендэла

$$\tau = \frac{37}{\sqrt{\left(\frac{10 \cdot 9}{2} - 2\right)\left(\frac{10 \cdot 9}{2} - 4\right)}} = \frac{37}{\sqrt{43 \cdot 41}} = \frac{37}{42} = 0{,}88.$$

По величине коэффициента ($\tau = 0{,}88$) можно сделать вывод о весьма большой тесноте зависимости между x и y.

КОЭФФИЦИЕНТ КОНКОРДАЦИИ

Корреляция рангов (R) может использоваться не только для двух, но и для большего числа показателей (факторов). Исчисляемый для этой цели показатель именуется **коэффициентом конкордации** (W). Его формула

$$W = \frac{12S}{m^2\left(n^3 - n\right)},$$

где m — количество коррелируемых факторов;
n — число наблюдений;
S — сумма квадратов отклонений суммы рангов по m факторам от их средней арифметической, т.е.

а) $S = \sum\limits_{1}^{n}\left(\sum\limits_{1}^{m} R_i - \overline{\sum R}\right)^2$ или, что по значению то же самое,

б) $S = \sum\limits_{1}^{n}\left(\sum\limits_{1}^{m} R_i\right)^2 - \dfrac{\left(\sum\limits_{1}^{n}\sum\limits_{1}^{m} R\right)^2}{n}$, где R_i — ранг i-го показателя.

Рассмотрим расчет коэффициента конкордации на конкретном примере.

Задача 8.4

Пусть имеются следующие условные данные по 5 предприятиям (графы 1, 2, 3, 4 таблицы).

Предприятие	Прибыль, млн. руб. y	Стоимость основных фондов, млн. руб. x	Затраты на 100 руб. продукции, руб. z	Ранжирование факторов			Сумма рангов $\sum_1^3 R_i$	Квадраты суммы рангов $\left(\sum_1^3 R_i\right)^2$
				R_y	R_x	R_z		
1	2	3	4	5	6	7	8	9
1	300	4,1	80	1	2	5	8	64
2	950	6,6	73	4	5	3	12	144
3	520	3,9	72	3	1	2	6	36
4	480	4,2	75	2	3	4	9	81
5	1000	6,3	67	5	4	1	10	100

По данным таблицы $\sum_1^n \sum_1^m R_i = 45$, $\sum_1^n \left(\sum_1^m R_i\right)^2 = 425$.

Определить (измерить) тесноту зависимости между y, x и z с помощью коэффициента конкордации (W).

Решение.

1. Ранжируем каждый из трех показателей (факторов) (графы 5, 6, 7).
2. Находим сумму рангов по каждой строке (графа 8) и общую сумму пяти строк.
3. Возводим в квадрат сумму рангов в каждой строке и находим общую сумму пяти строк (графа 9).
4. Находим S, используя приведенную выше формулу «*б*»: $S = 425 - (45)^2/5 = 20$.

Этот же результат получим, рассчитывая S по формуле «*а*»: сначала определяем $\overline{\sum R} = \frac{45}{5} = 9$, тогда

$$S = \sum_1^n \left(\sum_1^m R_i - \overline{\sum R}\right)^2 = (8-9)^2 + (12-9)^2 + (6-9)^2 +$$

$$+ (9-9)^2 + (10-9)^2 = 1 + 9 + 9 + 0 + 1 = 20.$$

5. Рассчитываем коэффициент конкордации:

$$W = \frac{12S}{m^2(n^3 - n)} = \frac{12 \cdot 20}{3^2(5^3 - 5)} = \frac{240}{9 \cdot 120} = \frac{2}{9} = 0,22.$$

Учитывая малую величину значения W, можно сказать, что зависимость между рассматриваемыми показателями (факторами) весьма незначительна.

Коэффициент конкордации часто используется в экспертных оценках для определения согласованности мнения m экспертов в распределении мест (рангов) между n исследуемыми факторами или объектами по их приоритетности.

ИЗУЧЕНИЕ ВЗАИМОСВЯЗИ НА ОСНОВЕ АНАЛИЗА ТАБЛИЦ ВЗАИМОСОПРЯЖЕННОСТИ

Особое место в изучении взаимосвязи занимает исследование особенностей распределения единиц совокупности по двум признакам. По характеру распределения можно судить, случайно оно или нет, т.е. есть ли зависимость между признаками, положенными в основу группировки, или нет. Покажем это на примере.

<u>Задача 8.5</u>

По одному из факультетов имеются следующие данные о распределении 600 студентов-вечерников по двум признакам: характеру работы и результатам сдачи экзаменов по специальным предметам:

Характер работы	Сдавшие сессию без неудовлетворительных оценок	Получившие неудовлетворительные оценки	Всего студентов
Работающие по профилю факультета	a $270^{(224)}$	b $50^{(96)}$	320
Работающие не по профилю факультета	c $150^{(196)}$	d $130^{(84)}$	280
Всего студентов	**420**	**180**	**600**

1. Определить, случайно или неслучайно распределение в таблице, т.е. сделать вывод о наличии или отсутствии зависимости успеваемости студентов-вечерников от соответствия профиля работы.
2. Измерить тесноту этой зависимости, если она есть.

Решение.

А. Для ответа на первый вопрос воспользуемся критерием Пирсона (χ^2):

$$\chi^2 = \sum \frac{(m_{ij} - m'_{ij})^2}{m'_{ij}},$$

где m_{ij} — эмпирические частоты таблицы;

m'_{ij} — теоретические частоты.

Последние рассчитываются по каждой строке (или столбцу) пропорционально общим итогам исходя из гипотезы о случайности распределения. В нашем примере теоретические частоты показаны в скобках (например, $m_{11} = 320 \cdot \frac{420}{600} = 224$ и т.д.).

$$\chi^2 = \frac{(270-224)^2}{224} + \frac{(50-96)^2}{96} + \frac{(150-196)^2}{196} + \frac{(130-84)^2}{84} = 67{,}5.$$

Определяем число степеней свободы $\nu = (k_1 - 1)(k_2 - 1) = (2 - 1)(2 - 1) = 1$, где k_1 и k_2 — число строк и столбцов.

Чтобы сделать вывод о случайности или неслучайности распределения, по Приложению 4 определяем табличное (пороговое) значение χ^2, допустимое при случайных расхождениях между эмпирическими (m_{ij}) и теоретическими (m'_{ij}) частотами.

Для уровня значимости $\alpha = 0{,}05$ и числа степеней свободы $\nu = 1$ значение $\chi^2_{\text{табл}} = 3{,}84$, а $\chi^2_{\text{факт}} = 67{,}5$. Так как $\chi^2_{\text{факт}} > \chi^2_{\text{табл}}$, делаем вывод, что распределение неслучайно и скорее связано с зависимостью между признаками, положенными в основу группировки, т.е. можно говорить о зависимости между характером работы студентов-вечерников и результатами сдачи ими экзаменов по специальным предметам.

χ^2 помогает выявить наличие или отсутствие зависимости между признаками, положенными в основу группировки в таблицах сопряженности, но не измеряет тесноту этой связи.

Б. Для измерения тесноты зависимости между указанными признаками используются следующие показатели:

1. Коэффициент ассоциации $k_{\text{асс}} = \frac{ad - bc}{ad + bc}$.

2. Коэффициент контингенции

$$k_{\text{континг}} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}.$$

В обеих формулах a, b, c, d — значения частот в четырехклеточной таблице $\begin{matrix} a & b \\ c & d \end{matrix}$.

3. Коэффициент взаимной сопряженности Пирсона

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}}, \text{ или } C = \sqrt{\frac{\varphi^2}{\varphi^2 + 1}}, \text{ где } \varphi^2 = \frac{\chi^2}{N}.$$

Показатель φ^2 можно рассчитать и самостоятельно: $\varphi^2 = \sum \frac{m_{ij}^2}{M_i M_j} - 1$, где m_{ij} – частоты в клетках таблицы, а M_i и M_j – итоговые частоты по строкам и столбцам.

4. Коэффициент взаимной сопряженности Чупрова

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

где k_1 и k_2 – число строк и столбцов в таблице.

Первые два показателя ($k_{асс}$ и $k_{континг}$) могут использоваться только для четырехклеточных таблиц, а коэффициенты сопряженности Пирсона и Чупрова — для таблиц любой размерности.

В нашем примере (задача 8.5) для четырехклеточной таблицы (или таблицы «четырех полей»):

а) $k_{асс} = \frac{ad - bc}{ad + bc} = \frac{270 \cdot 130 - 50 \cdot 150}{270 \cdot 130 + 50 \cdot 150} = 0{,}648;$

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

$$= \frac{270 \cdot 130 - 50 \cdot 150}{\sqrt{320 \cdot 280 \cdot 420 \cdot 180}} = 0{,}335;$$

в) $C = \sqrt{\frac{\chi^2}{\chi^2 + N}} = \sqrt{\frac{67{,}5}{67{,}5 + 600}} = 0{,}318$ или,

если $\varphi^2 = \frac{\chi^2}{N} = \frac{67{,}5}{600} = 0{,}1125$, то $C = \sqrt{\frac{\varphi^2}{\varphi^2 + 1}} = \sqrt{\frac{0{,}1125}{1{,}1125}} = 0{,}318;$

г) $K_ч = \sqrt{\frac{\varphi^2}{\sqrt{(k_1 - 1)(k_2 - 1)}}} = \sqrt{\frac{0{,}1125}{\sqrt{(2-1)(2-1)}}} = 0{,}335.$

Все рассчитанные выше показатели, кроме коэффициента ассоциации, свидетельствуют о том, что зависимость между характером работы у студентов-вечерников и результатами сдачи экзаменов по специальным предметам ниже средней.

Задача 8.6

Предположим, имеется следующее распределение 100 опытных участков (под овощной культурой) по двум признакам: степени полива (x) и уровню урожайности (y).

Полив x	Урожайность y						
	Высокая		Средняя		Низкая		Итого
Обильный	40	(33)	10	(12,1)	5	(9,9)	**55**
Средний	20	(18)	7	(6,6)	3	(5,4)	**30**
Слабый	—	(9)	5	(3,3)	10	(2,7)	**15**
Итого	**60**		**22**		**18**		**100**

1. Определить, случайно ли данное распределение или есть зависимость между x и y.
2. Измерить тесноту зависимости между степенью полива и уровнем урожайности.

Решение.

1. *Для определения зависимости между x и y:*

а) выдвигаем гипотезу о случайности распределения и рассчитываем теоретические (гипотетические) частоты в каждой строке пропорционально итоговым данным. (Например, если в итоговой строке суммы столбцов составляют 0,6, 0,22 и 0,18 от 100, то в каждой строке ее итога сохраняется такое же соотношение.) Теоретические частоты (m'_{ij}) показаны в таблице в скобках;

б) рассчитываем χ^2 (фактическое):

$$\chi^2_{\text{факт}} = \sum \frac{\left(m_{ij} - m'_{ij}\right)^2}{m'_{ij}} = \frac{(40-33)^2}{33} + \frac{(10-12{,}1)^2}{12{,}1} + \frac{(5-9{,}9)^2}{9{,}9} +$$

$$+ \frac{(20-18)^2}{18} + \frac{(7-6{,}6)^2}{6{,}6} + \frac{(3-5{,}4)^2}{5{,}4} + \frac{(0-9)^2}{9} +$$

$$+ \frac{(5-3{,}3)^2}{3{,}3} + \frac{(10-2{,}7)^2}{2{,}7} = 35{,}2.$$

По таблице Приложения 4 находим χ^2 табличное при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы $\nu = (3-1)(3-1) = 4$ $\chi^2_{\text{табл}} = 9{,}49$; так как $\chi^2_{\text{факт}} > \chi^2_{\text{табл}}$ $(35{,}2 > 9{,}49)$ при $\alpha = 0{,}05$ и $\nu = 4$, то гипотеза о случайности распределения опровергается, т.е. распределение в таблице неслучайно. Следовательно, есть основания предполагать наличие зависимости между x и y.

2. Для измерения тесноты зависимости используем

а) коэффициент взаимной сопряженности Пирсона

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} = \sqrt{\frac{35{,}2}{35{,}2 + 100}} = \sqrt{\frac{35{,}2}{135{,}2}} = 0{,}51$$

или, если рассчитать

$$\varphi^2 = \frac{\chi^2}{N} = \frac{35{,}2}{100} = 0{,}352, \text{ то } C = \sqrt{\frac{\varphi^2}{\varphi^2 + 1}} = \sqrt{\frac{0{,}352}{1{,}352}} = 0{,}51;$$

б) коэффициент Чупрова

$$K_ч = \sqrt{\frac{\varphi^2}{\sqrt{(k_1 - 1)(k_2 - 1)}}} = \sqrt{\frac{0{,}352}{\sqrt{2 \cdot 2}}} = \sqrt{\frac{0{,}352}{2}} = 0{,}42.$$

($K_ч$ всегда по значению несколько меньше C.)

Полученные значения коэффициентов (0,51 и 0,42) позволяют сделать вывод о том, что зависимость между x и y средняя.

Примечание: φ^2 можно было рассчитать независимо от χ^2, по формуле

$$\varphi^2 = \sum \frac{m_{ij}^2}{M_i M_j} - 1 = \left(\frac{40^2}{55 \cdot 60} + \frac{10^2}{55 \cdot 22} + \frac{5^2}{55 \cdot 18} + \frac{20^2}{30 \cdot 60} + \right.$$

$$\left. + \frac{7^2}{30 \cdot 22} + \frac{3^2}{30 \cdot 18} + \frac{5^2}{15 \cdot 22} + \frac{10^2}{15 \cdot 18} \right) - 1 = 1{,}352 - 1 = 0{,}352.$$

Литература

[1, с. 184—262], [2, с. 276—338], [3, с. 190—256].

Контрольные вопросы

1. В чем сущность корреляционной связи (зависимости) между показателями?
2. Как определяются параметры уравнения регрессии при линейной зависимости (на основе способа наименьших квадратов)?
3. Как определяются параметры уравнения регрессии криволинейной зависимости?
4. Как определяются ошибки параметров уравнения регрессии?
5. Какие вы знаете показатели измерения тесноты зависимости?
6. Как измеряется теснота зависимости с помощью корреляционного отношения?
7. Какие формулы линейного коэффициента корреляции вы знаете?
8. Как оценивается значимость коэффициента корреляции, рассчитанного по выборочным данным?
9. Каково содержание корреляционной таблицы и какова ее роль в изучении взаимосвязи между двумя показателями?
10. Определите понятие множественной корреляции.

11. Что такое совокупный и частный коэффициенты корреляции?
12. Что представляют собой коэффициенты корреляции рангов Спирмэна и Кендэла?
13. С помощью каких показателей изучается и измеряется корреляционная зависимость между качественными показателями на основе таблиц взаимосопряженности?
14. Что такое коэффициент конкордации? Как он рассчитывается?
15. Как проверяется значимость уравнения регрессии?

Задачи и упражнения для самостоятельной работы

Задача 1

Имеются следующие данные о росте 8 пар братьев и сестер:

Рост брата, см	Рост сестры, см
170	163
165	162
177	168
180	170
181	164
175	162
172	165
180	168

Определить тесноту зависимости между ростом братьев и сестер на основе:

а) коэффициента Фехнера;

б) коэффициентов корреляции рангов Спирмэна и Кендэла.

Ответ: а) $K_ф = 0{,}5$;
б) $\rho = 0{,}63$, $\tau = 0{,}59$.

Задача 2

Имеются следующие данные по 8 сахарным заводам о стоимости основных производственных фондов, *x* (млн. руб.) и суточной переработке сахарной свеклы, *y* (тыс. т):

x	y
2,0	8,9
2,3	10,0
2,4	9,9
2,9	10,3
2,9	10,0
3,7	13,0
3,7	12,8
4,1	13,1

1. Найти уравнение регрессии *y* по *x* и определить значимость его параметров (с помощью *t*-критерия).
2. Проверить остаточные величины на автокорреляцию.
3. Измерить тесноту зависимости между *x* и *y* с помощью:
   а) корреляционного отношения;
   б) линейного коэффициента корреляции;
   в) коэффициентов корреляции рангов Спирмэна и Кендэла.

4. С помощью F-критерия Фишера проверить уравнение регрессии на значимость (существенность).

О т в е т:

1) $\bar{y}_x = 4{,}64 + 2{,}12x$, параметры значимы, так как $t_{факт} > t_{табл}$;
2) автокорреляция отсутствует;
3) а) $\eta = 0{,}96$, б) $r = 0{,}96$, в) $\rho = 0{,}94$, $\tau = 0{,}87$;
4) уравнение значимо, так как $F_{факт} > F_{табл}$.

Задача 3

Имеются данные по 10 хозяйствам о количестве внесенных минеральных удобрений под зерновыми, x (кг/га) и об урожайности зерновых, y (ц/га):

x	y
15	13,5
18	14,0
19	14,0
19	14,3
21	14,0
30	15,0
30	18,2
35	15,0
38	17,0
40	20,0

С помощью коэффициентов корреляции рангов: а) Спирмэна и б) Кендэла измерить тесноту связи между x и y.

О т в е т: а) $\rho = 0{,}897$; б) $\tau = 0{,}786$.

Задача 4

У восьми учащихся колледжа зафиксировано следующее количество баллов, полученных за самостоятельные работы по математике (x) и по гуманитарным предметам (y):

Студент	x	y
А	90	75
Б	60	69
В	46	45
Г	68	49
Д	82	58
Е	71	54
Ж	66	59
З	78	70

Для характеристики корреляции между успеваемостью по математике и гуманитарным предметам рассчитать:

а) коэффициент Фехнера;
б) коэффициент корреляции рангов Спирмэна;
в) коэффициент корреляции рангов Кендэла.

О т в е т: а) $K_ф = 0{,}25$; б) $\rho = 0{,}548$;
в) $\tau = 0{,}428$.

Задача 5

По данным задачи 7.9 с помощью эмпирического корреляционного отношения ($\eta_{эмп}$) измерить тесноту зависимости процента выполнения норм от стажа работы.

О т в е т: $\eta_{эмп} = \sqrt{\frac{\delta^2}{\sigma^2}} = \sqrt{0,264} = 0,51.$

Задача 6

Совокупность разбита по определенному признаку z на 3 группы, численность которых: $n_1 = 10$, $n_2 = 20$, $n_3 = 20$.

Групповые средние исследуемого показателя x равны: $\bar{x}_1 = 5$, $\bar{x}_2 = 8$, $\bar{x}_3 = 15$.

Определить величину эмпирического корреляционного отношения, если общая дисперсия показателя x равна 18,5.

О т в е т: $\eta_{эмп} = 0,946.$

Задача 7

Записать уравнение регрессии y по x, имея следующие данные:

$$\bar{x} = 20,\ \overline{x^2} = 436,\ \bar{y} = 60,\ \overline{y^2} = 3700,\ r_{xy} = 0,75.$$

О т в е т: $\bar{y}_x = 35 + 1,25x.$

Задача 8

На основе опроса 400 работников коммерческих структур и 400 работников бюджетных организаций получено следующее их распределение по ответам на вопрос, довольны ли они своей заработной платой:

Работающие	Довольные заработной платой	Недовольные заработной платой	Итого
В коммерческих структурах	360	40	**400**
В бюджетных организациях	140	260	**400**
Итого работников	**500**	**300**	**800**

1. С помощью критерия Пирсона χ^2 определить, случайно или неслучайно данное распределение.
2. Рассчитать коэффициенты ассоциации и контингенции.

О т в е т: 1) $\chi^2 = 258,2$;

2) $k_{асс} = 0,887$, $k_{континг} = 0,568.$

Задача 9

Имеются следующие данные о распределении 200 молочных ферм области по производительности труда и себестоимости молока:

Себестоимость	Производительность			
	Высокая	Средняя	Низкая	Итого
Высокая	10	10	30	50
Средняя	30	30	10	70
Низкая	50	20	10	80
Итого	90	60	50	200

1. С помощью критерия χ^2 проверить, случайно ли данное распределение, т.е. существует ли зависимость между производительностью труда и себестоимостью молока.
2. Измерить тесноту зависимости между указанными показателями с помощью коэффициентов взаимной сопряженности Пирсона и Чупрова.

Ответ: 1) $\chi^2 = 51,4$;
2) $C = 0,45$, $K_ч = 0,36$.

Задача 10

Имеются следующие данные по Северо-Западному району РФ за 1993 г.:

Область	Урожайность зерновых, ц/га	Надой молока на одну корову, кг	Урожайность картофеля, ц/га
Ленинградская	16,3	3086	122
Новгородская	10,4	1823	123
Псковская	10,6	1772	126
Вологодская	10,8	2346	104

С помощью коэффициента конкордации определить, согласуется ли «рейтинг» областей по всем показателям.

Ответ: $W = 0,2$.

Задача 11

Имеются следующие данные по областям Центрально-Черноземного района РФ за 1998 г.:

Область	ВРП на душу населения в 1997 г., тыс. руб. x_1	Розничный товарооборот на душу населения, тыс. руб. x_2	Обеспеченность жильем, м^2 общей площади на одного жителя x_3
Белгородская	12,25	5,35	20,9
Воронежская	10,33	4,64	20,6
Курская	11,50	4,78	20,0
Липецкая	12,60	5,83	20,2
Тамбовская	7,27	5,10	19,4

1. Измерить тесноту связи между x_1 и x_2 с помощью коэффициентов корреляции рангов:
   а) Спирмэна; б) Кендэла.
2. С помощью коэффициента конкордации W определить, согласуется ли «рейтинг» областей по показателям x_1, x_2 и x_3.

Задача 12

Пусть имеются следующие условные данные о распределении 20 заводов по объему выпуска и о прибыли отдельных предприятий:

Объем выпуска, тыс. ед.	Число заводов	Прибыль по отдельным заводам, млн. руб.
До 10	3	1, 2, 3
11—20	5	2, 3, 4, 5, 6
21—30	7	4, 5, 5, 6, 7, 7, 8
31—40	3	7, 9, 11
41—50	2	12, 13

С помощью эмпирического корреляционного отношения измерить тесноту зависимости прибыли от объема выпуска.

Задача 13

В каком из следующих точечных распределений x и y независимы?

x	y		
	1	2	Итого
1	10	20	30
2	10	60	70
Итого	20	80	100

x	y			
	0	1	2	Итого
0	10	16	14	40
1	15	24	21	60
Итого	25	40	35	100

Задача 14

Получены следующие результаты анкетного обследования рабочих, имеющих вторичную занятость:

Дополнительно заняты	Количество ответов		
	мужчин	женщин	Всего
на одной работе	400	180	580
на двух работах	150	20	170
на трех работах	50	—	50
Всего	600	200	800

1. С помощью критерия χ^2 проверить, является ли данное распределение случайным.
2. Измерить тесноту зависимости между признаками, положенными в основу группировки, с помощью коэффициентов взаимной сопряженности:
   а) Пирсона; б) Чупрова.

Задача 15

Имеются следующие данные о плотности автомобильных дорог общего пользования с твердым покрытием (x_1), обеспеченности населения автобусами общего пользования (x_2) и собственными легковыми автомобилями (x_3) в Центральном районе РФ на начало 1999 г.:

Область	Протяженность дорог на 1000 км² территории, км x_1	Количество автобусов на 100 тыс. населения x_2	Количество автомобилей на 1000 населения x_3
Брянская	170	83	50,9
Владимирская	190	58	89,6
Ивановская	157	72	80,7
Калужская	158	62	103,4
Костромская	89	81	82,1
Московская	324	66	147,6
Орловская	157	83	101,6
Рязанская	161	57	94,0
Смоленская	177	64	94,4
Тверская	176	68	84,5
Тульская	198	81	96,5
Ярославская	171	89	90,3

1. С помощью коэффициента корреляции рангов Спирмэна измерить тесноту зависимости:
   а) между x_1 и x_2; б) между x_1 и x_3; в) между x_2 и x_3.
2. Измерить тесноту связи между всеми тремя показателями с помощью коэффициента конкордации.

Задача 16

По восьми предприятиям имеются следующие условные данные об энерговооруженности труда (x) и производительности труда (y):

Потребление электроэнергии на одного рабочего, кВт-ч x	Выработка на одного рабочего, тыс. руб. y
12	2,3
14	3,8
16	4,0
16	3,9
18	4,5
20	5,4
22	5,1
22	6,0

Измерить тесноту зависимости между x и y, используя:

1) коэффициент Фехнера;
2) коэффициенты корреляции рангов:
   а) Спирмэна; б) Кендэла;
3) линейный коэффициент корреляции.

Задача 17

Ниже в таблице приведено распределение ответов 4000 женщин, вступивших в первый брак 5 лет назад (до опроса), на вопрос о их семейном положении на момент опроса:

Семейное положение на *момент опроса*	Место проживания			Итого
	Сельская местность	Малые города	Большие города	
Состоящие в этом же браке	480	720	2000	**3200**
Разведенные или овдовевшие	20	280	500	**800**
Итого	**500**	**1000**	**2500**	**4000**

1. С помощью критерия χ^2 проверить, случайно или нет данное распределение.
2. Рассчитать коэффициенты взаимной сопряженности: а) Пирсона; б) Чупрова.
3. Сделать выводы о тесноте зависимости между признаками, положенными в основу группировки.

Задача 18

Имеются следующие данные о стаже работы (x) и выработке продукции за смену (y) у 10 рабочих одной специальности:

Номер рабочего	Производственный стаж, лет x	Выработка продукции за смену, шт. y
1	3	68
2	12	95
3	9	84
4	2	60
5	6	80
6	12	100
7	10	82
8	20	110
9	16	98
10	20	105

Измерить тесноту зависимости между x и y с помощью коэффициентов корреляции рангов:

а) Спирмэна; б) Кендэла.

Задача 19

Записать линейное уравнение регрессии y по x, имея следующие условные данные:

$$\bar{x} = 15,\ \overline{x^2} = 289,\ \bar{y} = 50,\ \sigma_y = 4,\ r_{xy} = 0{,}6.$$

Задача 20

Имеются следующие результаты опроса 100 студентов дневного отделения о совмещении учебы с работой:

Студенты	Численность студентов в возрасте		Итого
	до 20 лет	старше 20 лет	
Неработающие	20	10	**30**
Сочетающие работу с учебой	20	50	**70**
Итого	**40**	**60**	**100**

1. С помощью критерия χ^2 определить, случайно или неслучайно данное распределение.
2. Рассчитать коэффициенты ассоциации и контингенции и сделать выводы о тесноте зависимости между признаками, положенными в основу группировки.

Тема 9
РЯДЫ ДИНАМИКИ

Числовые значения того или иного статистического показателя, составляющие ряд динамики, называют **уровнями** ряда и обозначают через y.

Изучая ряды динамики, надо обратить внимание на необходимость сопоставимости уровней ряда, научиться определять различные виды рядов динамики (моментные, интервальные, ряды средних и относительных величин) и рассчитывать для них простейшие показатели: средний уровень, абсолютные приросты, темпы роста и прироста. Очень важно правильно определять средние темпы роста.

В качестве обобщенной характеристики для определенного периода рассчитывается **средний уровень** ряда ($\bar{y}$).

В интервальном ряду абсолютных величин средний уровень рассчитывается как средняя арифметическая простая:

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

Для м о м е н т н о г о ряда, содержащего n уровней с равными промежутками между датами (моментами), средний уровень определяется по формуле

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

Эта средняя известна в статистике как *средняя хронологическая для моментных рядов.*

Если промежутки между датами не равны, то средний уровень для моментного ряда можно рассчитать как среднюю арифметическую из средних значений уровней на каждую пару моментов, взвешенных по величине расстояний (отрезков времени) между датами, т.е.

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

$$= \frac{\sum (y_i + y_{i+1}) t_i}{2\sum t_i}.$$

Если предполагается, что каждое значение y_i остается неизменным до следующего $(i + 1)$-го момента, т.е. известна точная дата изменения уровней, то расчет $\bar{y}$ можно осуществить по формуле

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

где t_i – время, в течение которого уровень y_i считается неизменным.

При анализе рядов динамики, чтобы проследить за направлением и размером изменений уровней во времени, рассчитывают такие показатели, как:

абсолютные приросты уровней (разность между двумя уровнями): *цепные* ($_{ц}\Delta y = y_i - y_{i-1}$) и *базисные* ($_{б}\Delta y = y_i - y_0$);

темпы роста (изменения) $T_р$ – относительные показатели, рассчитываемые как отношение двух уровней ряда. Темпы роста могут быть *цепными*, если каждый уровень сопоставляется с предыдущим ($_{ц}T_р = y_i/y_{i-1}$), и *базисными*, когда все уровни сопоставляются с уровнем одного какого-то периода (часто это начальный уровень ряда: ${}_{б}T_р = y_i/y_0$). Темпы роста как относительные величины могут выражаться в виде коэффициентов, т.е. простого кратного отношения (база сравнения принимается за единицу), и в процентах (база сравнения принимается за 100 единиц).

Между цепными и базисными коэффициентами роста существует взаимосвязь, позволяющая при необходимости переходить от цепных коэффициентов к базисным и наоборот. В частности: произведение цепных коэффициентов роста равно базисному; результат деления двух базисных коэффициентов равен цепному;

темпы прироста (снижения) уровней $T_{пр}$ – относительные показатели, показывающие, на сколько процентов данный уровень (y_i) больше (или меньше) другого, принимаемого за базу сравнения.

Темп прироста можно рассчитать двояко:

- путем вычитания 100% из темпа роста, т.е. $T_{пр} = T_р - 100\%$;
- как процентное отношение абсолютного прироста к тому уровню, по сравнению с которым рассчитан абсолютный прирост.

Средний темп роста за определенный период наиболее часто рассчитывается как средняя геометрическая из цепных коэффициентов роста. Использование именно средней геометрической обосновано следующими рассуждениями.

Пусть имеется определенный ряд динамики с уровнями

$$y_0,\ y_1,\ y_2,\ y_3,\ \dots,\ y_n.$$

Цепные коэффициенты роста k_i для каждого периода составят:

$$k_1 = \frac{y_1}{y_0},\ k_2 = \frac{y_2}{y_1},\ \dots,\ k_n = \frac{y_n}{y_{n-1}}.$$

Из этих отношений каждый уровень можно выразить через предыдущий или y_0 (базисный):

$$y_1 = y_0k_1;\ y_2 = y_1k_2 = y_0k_1k_2;\ \dots;\ y_n = y_{n-1}k_n = y_0k_1k_2\cdots k_n,$$

т.е. конечный уровень y_n равняется базисному y_0, умноженному на произведение цепных коэффициентов роста.

Рассчитывая средний коэффициент роста, предполагаем, что замена индивидуальных коэффициентов роста k_i средними $\bar{k}$ обеспечивает достижение одинакового значения конечного уровня y_n. Так как

$$y_n = y_0k_1k_2\cdots k_n \quad \text{и} \quad y_n = y_0\bar{k}\,\bar{k}\cdots\bar{k} = y_0\left(\bar{k}\right)^n,$$

то

$$y_0k_1k_2\cdots k_n = y_0\left(\bar{k}\right)^n.$$

Отсюда

$$\left(\bar{k}\right)^n = k_1 k_2 \cdots k_n, \quad \text{а} \ \bar{k} = \sqrt[n]{k_1 k_2 \cdots k_n},$$

т.е. средний коэффициент роста равен корню n-й степени из произведения n цепных коэффициентов роста. Умножая результат на 100, получаем средний темп роста в процентах.

Из выражения $y_n = y_0\left(\bar{k}\right)^n$ получаем другую формулу для расчета *среднего коэффициента роста*:

$$\bar{k} = \sqrt[n]{\frac{y_n}{y_0}}.$$

Обе эти формулы используются при расчете среднего темпа роста в тех случаях, когда предполагается обеспечение (достижение) определенного конечного уровня y_n. Однако при этом сумма фактических уровней $\sum y_i$ не будет совпадать с суммой уровней, рассчитанных на основе y_0 и $\bar{k}$.

Если при расчете среднего темпа роста важно обеспечить не только конечный уровень, но и суммарное значение исследуемого показателя за анализируемый период (например, вложение инвестиций за весь период, ввод в действие жилой площади и т.д.), то тогда должно обеспечиваться равенство

$$y_1 + y_2 + \ldots + y_n = y_0 \bar{k} + y_0 \bar{k}^2 + \ldots + y_0 \bar{k}^n,$$

или

$$\sum_1^n y_i = y_0\left(\bar{k} + \bar{k}^2 + \ldots + \bar{k}^n\right).$$

Отсюда

$$\frac{\sum_1^n y_i}{y_0} = \bar{k} + \bar{k}^2 + \ldots + \bar{k}^n.$$

Эта формула условно названа *средней параболической*, а рассчитанное по ней $\bar{k}$ – *средним параболическим коэффициентом (темпом) роста*.

Особое внимание следует обратить на обработку динамических рядов с целью выявления закономерностей (тенденций) изменения явлений и сглаживания случайных колебаний.

Основные способы обработки рядов динамики:

1) укрупнение интервалов и расчет для них средних показателей;
2) сглаживание уровней способом скользящей (подвижной) средней;
3) выравнивание по аналитическим формулам.

Суть последнего способа заключается в том, что по эмпирическим данным находят так называемые уравнения тренда, по которому определяют теоретические уровни, рассматриваемые как функция времени, т.е. $\bar{y}_t = f(t)$.

Нахождение параметров той или иной гипотетической функции осуществляется аналогично нахождению параметров уравнений регрессии, описанных в предыдущей теме (только в качестве фактора x выступает фактор времени t).

Так, при выравнивании ряда по прямой для нахождения параметров прямой решается система нормальных уравнений вида

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

При ручной обработке для упрощения счета при выравнивании динамических рядов условное обозначение временных точек (t) можно вести так, чтобы $\sum t = 0$. В этом случае системы нормальных уравнений значительно упрощаются. Так, при выравнивании по прямой система будет иметь вид

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

при выравнивании по параболе второго порядка (если $\sum t = 0$) система имеет следующий вид:

$$\begin{cases} na_0 + a_2\sum t^2 = \sum y, \\ a_1\sum t^2 = \sum yt, \\ a_0\sum t^2 + a_2\sum t^4 = \sum yt^2. \end{cases}$$

Выравнивание по аналитическим формулам может быть использовано при прогнозировании отдельных показателей путем экстраполяции ряда (нахождения уровней за пределами данного ряда).

При изучении этой темы необходимо иметь представление об автокорреляции. Ряды, у которых каждый уровень может быть выражен как функция предыдущих, например $y_t = f(y_{t-1})$, называют **авторегрессионными**, а зависимость между соседними уровнями ряда именуют **автокорреляцией** и измеряют ее с помощью коэффициента автокорреляции по формуле

$$r_a = \frac{\overline{y_t y_{t-1}} - \overline{y}_t \cdot \overline{y}_{t-1}}{\sigma_{y_t} \sigma_{y_{t-1}}},$$

или

$$r_a = \frac{\sum y_t y_{t-1} - \frac{\sum y_t \sum y_{t-1}}{n}}{\sqrt{\left[\sum (y_t)^2 - \frac{\left(\sum y_t\right)^2}{n}\right]\left[\sum (y_{t-1})^2 - \frac{\left(\sum y_{t-1}\right)^2}{n}\right]}}.$$

Изучение автокорреляции занимает немаловажное место в анализе рядов динамики. В частности, при параллельном рассмотрении рядов динамики измерять корреляцию между ними можно только после проверки каждого ряда на автокорреляцию и исключения ее, если она есть. Исключение автокорреляции в рядах можно обеспечить, коррелируя не сами уровни, а так называемые **остаточные величины**, получаемые как разность эмпирических и теоретических (выравненных) уровней, т.е. $d_x = x - \overline{x}_t$ и $d_y = y - \overline{y}_t$. В этом случае корреляция между остаточными величинами будет определяться по формуле

$$r = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}}.$$

В свою очередь, остаточные величины (обозначим их символом ε_t) также должны проверяться на автокорреляцию.

Для этой цели могут быть использованы коэффициент автокорреляции Андерсона (r_a) и критерий Дурбина—Ватсона(d):

$$r_a = \frac{\sum_{2}^{n} \varepsilon_t \varepsilon_{t-1}}{\sum_{1}^{n} \varepsilon_t^2}, \qquad d = \frac{\sum_{2}^{n} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{1}^{n} \varepsilon_t^2}.$$

Значение r_a сравнивается с табличным (см. Приложение 7). Если фактическое значение меньше табличного, то считается, что автокорреляция между ε_t отсутствует.

В свою очередь значение d также сопоставляется с табличным (см. Приложение 5). Если $d > d_2$, автокорреляция отсутствует; если $d < d_1$, автокорреляция имеет место; если $d_1 \leqslant d \leqslant d_2$, ничего определенного сказать нельзя.

Важное место в анализе рядов динамики занимает изучение сезонных колебаний. Измерение последних осуществляется с помощью так называемых **индексов сезонности**, которые могут рассчитываться по-разному. Так, по данным одного года индексы сезонности рассчитываются как процентное отношение помесячных (или квартальных) уровней к среднему уровню за год, т.е. $I_{сез} = \frac{y_i}{\bar{y}} 100\%$.

Для большей надежности сезонность изучается чаще по данным за 3 года. В этом случае для каждого месяца рассчитывается средний уровень за 3 года, который и сопоставляется со средним уровнем за весь период, т.е. $I_{сез} = \frac{\bar{y}_i}{\bar{y}} 100\%$.

Возможен и другой подход: для каждого года рассчитываются помесячные индексы сезонности $\left(i = \frac{y_i}{\bar{y}}\right)$, а затем для каждого месяца рассчитывается средняя за 3 года, т.е. $\bar{I}_{сез} = \frac{\sum i_i}{3}$.

Чтобы учесть и имеющую место тенденцию изменения, рекомендуется производить сглаживание ряда способом скользящей средней или по аналитической функции ($\bar{y}_t = f(t)$), а затем находить отношение фактического уровня каждого месяца (y_i) к выравненному (сглаженному), т.е. $I_{сез} = \frac{y_i}{\bar{y}_t} 100\%$. Для учета сезонной волны в практических целях (например, при прогнозировании) из полученных отношений для одноименных месяцев (кварталов) рассчитывается средняя величина (за 2 или 3 года).

Рассмотрим решение некоторых задач по этой теме.

Задача 9.1

Пусть имеются следующие данные о производстве зерна в одном из хозяйств за 5 лет:

Год	1993	1994	1995	1996	1997
Производство зерна, тыс. ц, y_i	50	54	62	70	80

Рассчитать:

1) средний уровень за 5 лет;
2) ежегодные абсолютные приросты;
3) ежегодные темпы роста;
4) среднегодовой темп роста за 4 года, с 1994 по 1997 г.

Решение.

А. Так как это интервальный ряд, то средний уровень ряда (среднегодовое производство зерна) определим как среднюю арифметическую простую:

$$\bar{y} = \frac{\sum y_i}{n} = \frac{50 + 54 + 62 + 70 + 80}{5} = \frac{316}{5} = 63{,}2 \text{ (тыс. ц).}$$

Б. Ежегодные абсолютные приросты находим как разность между двумя уровнями:

для 1994 г. ....54 – 50 = 4 тыс. ц;
для 1995 г. ....62 – 54 = 8 тыс. ц;
для 1996 г. ....70 – 62 = 8 тыс. ц;
для 1997 г. ....80 – 70 = 10 тыс. ц.

В. Ежегодные коэффициенты роста (цепные) находим как отношение уровня каждого года к предыдущему:

для 1994 г. .... $k_1 = \frac{54}{50} = 1{,}08;$

для 1995 г. .... $k_2 = \frac{62}{54} = 1{,}148;$

для 1996 г. .... $k_3 = \frac{70}{62} = 1{,}129;$

для 1997 г. .... $k_4 = \frac{80}{70} = 1{,}143.$

Умножая коэффициенты на 100%, получаем темпы роста.

Г. Среднегодовой темп роста можно рассчитать как среднюю геометрическую из годовых темпов роста:

1) $\overline{T} = \sqrt[n]{T_1 T_2 \cdots T_n}$ $\left(\text{или } \overline{T} = \sqrt[n]{k_1 k_2 \cdots k_n} \cdot 100\%\right)$

либо по формуле:

2) $\overline{T} = \sqrt[n]{\dfrac{y_n}{y_0}} \cdot 100\%,$

которая тождественна первой.

По первой формуле

$$\overline{T} = \sqrt[4]{108 \cdot 114{,}8 \cdot 112{,}9 \cdot 114{,}3} = 112{,}5\%.$$

По второй формуле

$$\overline{T} = \sqrt[4]{\frac{80}{50}} \cdot 100\% = \sqrt[4]{1{,}6} \cdot 100\% = 112{,}5\%,$$

т.е. среднегодовой темп роста за 4 года (с 1994 по 1997 г.) равен 112,5%.

Расчет среднего темпа (коэффициента) роста приведен выше по формулам средней геометрической из цепных отношений $\left(\overline{k} = \sqrt[n]{k_1 k_2 \cdots k_n}\right)$ и тождественной ей $\overline{k} = \sqrt[n]{\dfrac{y_n}{y_0}}$, так как ориентирован на достижение конечного уровня (y_n) в исследуемом периоде.

Если же ориентация берется на достижение суммарного значения (объема) исследуемого показателя за определенный период, то, как указывалось выше, для расчета среднего коэффициента (темпа) роста используется средняя параболическая вида

$$\overline{k} + \overline{k}^2 + \overline{k}^3 + \ldots + \overline{k}^n = \frac{\sum_{1}^{n} y_i}{y_0},$$

где значение $\overline{k}$ определяется по специальной таблице* (см. Приложение 10) на основе отношения $\dfrac{\sum_{1}^{n} y_i}{y_0}$ для соответствующего n (длина периода).

Пример расчета среднего параболического темпа роста приведен ниже.

---

* Данная таблица составлена проф. Л.С. Казинцом для $\overline{k} > 1$ и дополнена проф. Г.П. Громыко расчетами для $\overline{k} < 1$.

Задача 9.2

Имеются следующие данные по РФ о вводе в действие жилой площади:

Год	1985	1986—1991
Введено млн. кв. м общей площади, y_i	62,5	394,7

Определить среднегодовой темп роста ввода в действие жилой площади за 1986—1991 гг. (т.е. за 6 лет), ориентированный на достижение общей суммы введенного жилья за указанный период (т.е. 394,7 млн. кв. м).

Решение.

Используем формулу средней параболической:

$$\bar{k} + \bar{k}^2 + \bar{k}^3 + \ldots + \bar{k}^n = \frac{\sum_1^n y_i}{y_0}.$$

В нашем примере $y_0 = 62,5$, а $\sum_1^6 y_i = 394,7$, т.е. отношение

$$\frac{\sum_1^n y_i}{y_0} = \frac{394,7}{62,5} = 6,305.$$

По таблице Приложения 10 в графе $n = 6$ находим значение, наиболее близкое к полученному отношению (6,305). Это число 6,323, которому соответствует $\bar{k} = 1,015$. Это и есть искомый среднегодовой коэффициент роста ввода жилья за 6 лет (с 1986 по 1991 г.). Итак, среднегодовой темп роста ввода в действие жилой площади за указанный период составлял 101,5%, а среднегодовой темп прироста был равен 101,5% – 100% = 1,5%.

Задача 9.3

Имеются следующие данные о лесовосстановительных работах в РФ за 1990—1995 гг.:

Год	1990	1991	1992	1993	1994	1995
Посадка и посев леса, тыс. га	566	521	447	428	391	367

Рассчитать среднегодовой темп роста (снижения) за 1991—1995 гг. лесовосстановительных работ, ориентированный на:

а) достижение фактического уровня в 1995 г.;

б) достижение общей площади посадки и посева леса за 1991—1995 гг.

Решение.

Если ориентироваться на конечный уровень 1995 г., то:

а) среднегодовой коэффициент изменения лесовосстановительных работ по средней геометрической

$$\bar{k} = \sqrt[n]{\frac{y_n}{y_0}} = \sqrt[5]{\frac{367}{566}} = \sqrt[5]{0{,}6484} = 0{,}9169 \approx 0{,}917,$$

или 91,7%, т.е. ежегодно объем работ уменьшался в среднем на 8,3%;

б) если при расчете $\bar{k}$ ориентироваться на общий объем работ, произведенных за 5 лет, то надо применить формулу параболической средней:

$$\bar{k} + \bar{k}^2 + \bar{k}^3 + \ldots + \bar{k}^n = \frac{\sum_1^n y_i}{y_0}.$$

Находим $\dfrac{\sum_1^5 y_i}{y_0} = \dfrac{521 + 447 + 428 + 391 + 367}{566} = 3{,}805.$

По таблице Приложения 10 в графе $n = 5$ находим значение, близкое к 3,805. Это будет 3,801. Ему соответствует (в первой графе) $\bar{k} = 0{,}91$, т.е. среднегодовое снижение посадок при общем объеме работ за 5 лет составило 9%.

Поскольку конечный (последний) уровень ряда может быть случайным (нехарактерным), то представляется, что расчет по второй формуле, где учитывается сумма уровней за n лет, в данном примере более адекватен.

Задача 9.4

Имеются следующие данные о поголовье коров на молочной ферме в 1999 г.:

на 1 января 1999 г. .... 300 голов;
на 1 апреля 1999 г. .... 330 голов;
на 1 июля 1999 г. .... 338 голов;
на 1 октября 1999 г. .... 320 голов;
на 1 января 2000 г. .... 316 голов.

Определить среднее поголовье коров за год.

Решение.

Так как это моментный ряд, то для нахождения среднего уровня ряда используем формулу средней хронологической вида

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

где y_1, y_2, ..., y_n — уровни ряда;
n — число уровней.

В нашем примере

$$\bar{y} = \frac{\frac{300}{2} + 330 + 338 + 320 + \frac{316}{2}}{4} = \frac{1296}{4} = 324 \text{ (головы)}.$$

Задача 9.5

Имеются следующие данные, характеризующие динамику производства валового выпуска продукции предприятия по месяцам (графы 1 и 2 таблицы):

Месяц	Валовой выпуск продукции, млн. руб.	Скользящая сумма трех членов	Скользящая средняя из трех членов
1	2	3	4
Январь	63	—	—
Февраль	93	258	86
Март	102	312	104
Апрель	117	345	115
Май	126	360	120
Июнь	117	383	128
Июль	140	383	128
Август	126	396	132
Сентябрь	130	399	133
Октябрь	143	408	136
Ноябрь	135	423	141
Декабрь	145	—	—

Требуется произвести сглаживание ряда, применяя трехмесячную скользящую среднюю.

Решение.

Чтобы рассчитать первую скользящую среднюю, находим сумму продукции за январь, февраль, март (графа 3) и делим ее на 3:

$$\frac{63 + 93 + 102}{3} = \frac{258}{3} = 86.$$

Найденную среднюю относим к февралю (т.е. к среднему из трех суммируемых месяцев — графа 4). Для отыскания второй скользящей средней находим сумму продукции за февраль, март, апрель и делим на 3:

$$\frac{93+102+117}{3}=\frac{312}{3}=104.$$

Найденную среднюю относим к марту и т.д.

Результаты подсчета скользящих сумм и средних из них показаны в графах 3 и 4 таблицы.

Задача 9.6

Имеются следующие данные о численности населения города за 5 лет (на начало года):

Год	1993	1994	1995	1996	1997
Численность населения, тыс. чел.	72	78	83	87	90

Найти линию тренда и, используя полученное уравнение, определить численность населения в 2000 г. (прогноз).

Решение.

Предположив, что численность населения изменяется во времени по прямой $\bar{y}_t = a_0 + a_1 t$, для нахождения параметров a_0 и a_1 решаем систему нормальных уравнений, отвечающих требованию способа наименьших квадратов:

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

В таблице на с. 116 рассчитаны необходимые для решения системы уравнений суммы: $\sum y$, $\sum t$, $\sum t^2$, $\sum yt$. Годы последовательно обозначены как 1, 2, 3, 4, 5 ($n = 5$).

Подставляя полученные суммы в систему уравнений

$$\begin{cases} 5a_0 + 15a_1 = 410 \\ 15a_0 + 55a_1 = 1275, \end{cases}$$

получаем $a_1 = 4{,}5$; $a_0 = 68{,}5$.

Отсюда искомое уравнение тренда $\bar{y}_t = 68{,}5 + 4{,}5t$. Подставляя в это уравнение значения t: 1, 2, 3, 4, 5, находим выравненные (теоретические) значения $\bar{y}_t$ (см. графу 6 таблицы). Для 2000 г. $t = 8$.

Год	Численность населения, тыс. чел., y_i	Условное обозначение времени t	t^2	yt	$\bar{y}_t = 68{,}5 + 4{,}5t$
1	2	3	4	5	6
1993	72	1	1	72	73
1994	78	2	4	156	77,5
1995	83	3	9	249	82
1996	87	4	16	348	86,5
1997	90	5	25	450	91
$\sum$	410	15	55	1275	410

Следовательно, по прогнозу численность населения города в 2000 г. составит $68{,}5 + 4{,}5 \cdot 8 = 104{,}5$ (тыс. чел.).

Для решения данной задачи можно использовать и *второй способ*, упрощенный.

Как указывалось выше, если время (t) обозначить так, чтобы $\sum t = 0$ (т.е. счет вести от середины ряда), то система упростится и примет вид

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

Каждое уравнение в этом случае решается самостоятельно: $a_0 = \sum y/n$ и $a_1 = \sum yt/\sum t^2$.

Необходимые для расчета a_0 и a_1 суммы приведены ниже.

Условное обозначение времени, t	Год	Численность населения, тыс. чел., y	t^2	yt	$\bar{y}_t = 82 + 4{,}5t$
1	2	3	4	5	6
–2	1993	72	4	–144	73,0
–1	1994	78	1	–78	77,5
0	1995	83	0	0	82,0
1	1996	87	1	87	86,5
2	1997	90	4	180	91,0
$\sum t = 0$	$n = 5$	$\sum y = 410$	$\sum t^2 = 10$	$\sum yt = 45$	$\sum \bar{y}_t = 410$

Получаем

$$5a_0 = 410; \quad a_0 = 82;$$
$$10a_1 = 45; \quad a_1 = 4{,}5;$$

отсюда уравнение прямой для выравненных уровней (линия тренда)

$$\bar{y}_t = 82 + 4{,}5t.$$

Выравненные значения, рассчитанные по последней формуле путем подстановки в нее значений $t = -2, -1, 0, 1, 2$, показаны в графе 6 таблицы.

Численность населения в 2000 г. ($t = 5$) по формуле будет $\bar{y}_t = 82 + 4{,}5 \cdot 5 = 104{,}5$ (тыс. чел.).

Естественно, эта величина условная, рассчитанная при предположении, что линейная закономерность изменения численности населения, принятая для 1993—1997 гг., сохранится на последующий период до 2000 г.

Задача 9.7

По данным задачи 9.6 проверим остаточные величины ($\varepsilon_t = y - \bar{y}_t$) на автокорреляцию (считается, что линия тренда (уравнение регрессии) подобрана удачно, если в остаточных величинах отсутствует автокорреляция).

Для этого рассчитаем коэффициент автокорреляции для остаточных величин (r_a) и критерий Дурбина–Ватсона (d). Все необходимые для них расчеты приведены ниже в таблице.

Год	y	$\bar{y}_t$	$y - \bar{y}_t = \varepsilon_t$	ε_{t-1}	$\varepsilon_t \varepsilon_{t-1}$	ε_t^2	$\varepsilon_t - \varepsilon_{t-1}$	$(\varepsilon_t - \varepsilon_{t-1})^2$
1993	72	73,0	−1,0	—	—	1,00	—	—
1994	78	77,5	0,5	−1	−0,5	0,25	1,5	2,25
1995	83	82,0	1,0	0,5	0,5	1,00	0,5	0,25
1996	87	86,5	0,5	1,0	0,5	0,25	−0,5	0,25
1997	90	91	−1,0	0,5	−0,5	1,00	−1,5	2,25
$\sum$	410	410	0	—	0	3,5	—	5,0

1. Коэффициент автокорреляции (r_a) для остаточных величин рассчитывается по формуле

$$r_a = \frac{\sum\limits_{2}^{n} \varepsilon_t \varepsilon_{t-1}}{\sum\limits_{1}^{n} \varepsilon_t^2}.$$

В нашем примере $\sum \varepsilon_t \varepsilon_{t-1} = 0$, а $\sum \varepsilon_t^2 = 3{,}5$. Следовательно, $r_a = \frac{0}{3.5} = 0$, что свидетельствует об отсутствии автокорреляции в остаточных величинах, а следовательно, и о том, что линия тренда подобрана удачно. В том случае, если $r_a \neq 0$, его значение надо сравнить с табличным (см. Приложение 7) для данного числа уровней ряда (n) и при заданном уровне значимости ($\alpha = 0{,}05$ или $\alpha = 0{,}01$). Если расчетное значение r_a меньше табличного, то можно говорить об отсутствии автокорреляции в рассматриваемых показателях.

2. Критерий Дурбина–Ватсона (d) рассчитывается по формуле

$$d = \frac{\sum_{2}^{n}\left(\varepsilon_t - \varepsilon_{t-1}\right)^2}{\sum_{1}^{n}\varepsilon_t^2}.$$

В нашем примере $d = \frac{5}{3{,}5} = 1{,}43.$

Сравниваем его со значениями d_1 и d_2 (при $v = 1$, где v — число переменных в уравнении тренда) в таблице Приложения 5 для $n = 15$ (таблица не имеет значений d_1 и d_2 для $n < 15$). Табличные $d_1 = 1{,}08$ и $d_2 = 1{,}36$. Рассчитанное нами $d > d_2$. Следовательно, автокорреляция в остаточных величинах отсутствует.

Рассмотрим несколько способов расчета индексов сезонности.

Задача 9.8

Рассчитать поквартальные индексы сезонности по данным о производстве растительного масла в РФ (тыс. т) за 1992 и 1993 гг..

Кварталы года		Производство масла, тыс. т y_i	Индексы сезонности $I_{сез} = (y_i : \bar{y}) \times 100\%$	Средние индексы сезонности, % $\bar{I}_{сез}$
1		2	3	4
1992	I	298,8	121,7	123,45
	II	228,9	93,2	108,0
	III	118,4	48,2	55,2
	IV	270,4	110,1	113,3
1993	I	307,3	125,2	123,45
	II	301,5	122,8	108,0
	III	152,7	62,2	55,2
	IV	286,2	116,5	113,3
Σ		1964,2	–	–

Решение.

I способ. 1. Рассчитываем средний квартальный уровень производства за 2 года:

$$\bar{y} = \frac{\sum y_i}{n} = \frac{1964,2}{8} = 245,525 \text{ (тыс. т).}$$

2. Находим отношение уровня каждого квартала к среднеквартальному за 2 года, т.е. индексы сезонности: $I_{сез} = \frac{y_i}{\bar{y}} 100\%$. Расчет их дан в графе 3.

3. По данным двух лет рассчитываем средний индекс для каждого квартала.

Так, для I квартала $\bar{I}_{сез} = \frac{121,7 + 125,2}{2} = 123,45$ (%), для II квартала $\bar{I}_{сез} = \frac{93,2 + 122,8}{2} = 108,0$ (%) и т.д.

Средние индексы сезонности (квартальные) приведены в графе 4.

II способ. Если в месячных или квартальных показателях заметна тенденция к изменениям (росту или снижению) по годам, то рекомендуется произвести аналитическое выравнивание ряда, а затем индексы сезонности рассчитывать как отношение y_i к выравненным уровням, а не к общей средней.

Применим этот способ к нашему примеру, для чего сначала, выравняв ряд по прямой, рассчитаем теоретические уровни $\left(\bar{y}_t\right)$. Все расчеты, необходимые для выравнивания ряда, а также расчеты индексов сезонности показаны далее в таблице.

Решение.

1. Для выравнивания примем линейную функцию:

$\bar{y}_t = a_0 + a_1 t$. Параметры a_0 и a_1 определяем из системы нормальных уравнений

$$\begin{cases} 8a_0 + 36a_1 = 1964,2 \\ 36a_0 + 204a_1 = 8897,4, \end{cases}$$

получаем $a_1 = 1,4$; $a_0 = 239,22$.

Отсюда $\bar{y}_t = 239,22 + 1,4t$.

Подставляя в данное уравнение значения t: 1, 2, ..., 8, находим выравненные значения уравнений ряда, т.е. $\bar{y}_t$ (графа 5).

Квар-талы t	Произ-водство масла, тыс. т y_i	t^2	$y_i t$	$\bar{y}_t$	$I_{сез} = \frac{y_i}{\bar{y}_t} 100\%$	Средние индексы сезон-ности, %	Выравненные уровни $\left(\bar{y}_t\right)$ с учетом сезонности
1	2	3	4	5	6	7	8 = = гр. 5 × гр.7
I 1	298,8	1	298,8	240,62	124,2	124,5	299,6
II 2	228,9	4	457,8	242,02	94,6	108,2	261,9
III 3	118,4	9	355,2	243,42	48,6	55,0	133,9
IV 4	270,4	16	1081,6	244,82	110,4	112,4	275,2
I 5	307,3	25	1536,5	246,22	124,8	124,5	306,5
II 6	301,5	36	1809,0	247,62	121,8	108,2	267,9
III 7	152,7	49	1068,9	249,02	61,3	55,0	137,0
IV 8	286,2	64	2289,6	250,42	114,3	112,4	281,5
$\sum$ 36	**1964,2**	**204**	**8897,4**	**1964,2**	—	—	**1963,5**

2. Индексы сезонности (графа 6) рассчитаны как отношение уровня каждого квартала к выравненному за этот же квартал.

3. Так как для двух лет поквартальные индексы сезонности различаются, рассчитываем среднее значение из двухлетних данных (графа 7).

Имея полученные средние квартальные индексы сезонности (графа 7), можно скорректировать выравненные уровни $\left(\bar{y}_t\right)$ с учетом сезонной волны. Для этого каждое значение $\bar{y}_t$ (графа 5) умножаем на средний индекс сезонности соответствующего квартала. Результаты таких расчетов приведены в графе 8.

По такой же схеме, зная уравнение линии тренда и сезонную волну, можно спрогнозировать производство растительного масла по кварталам на следующий год. (В нашем примере на 1994 г.)

Так, если $\bar{y}_t = 239{,}22 + 1{,}4t$ и для IV квартала 1993 г. принималось $t = 8$, то для 1994 г. берем t: 9, 10, 11, 12.

Тогда прогноз производства масла для I квартала 1994 г. — $y_{t=9} = (239{,}22 + 1{,}4 \cdot 9) \cdot 1{,}245 = 313{,}5$ (тыс. т).

Для II квартала $y_{t=10} = (239{,}22 + 1{,}4 \cdot 10) \cdot 1{,}082 = 274$ (тыс. т).

Для III квартала $y_{t=11} = (239{,}22 + 1{,}4 \cdot 11) \cdot 0{,}55 = 140$ (тыс. т.).

Для IV квартала $y_{t=12} = (239{,}22 + 1{,}4 \cdot 12) \cdot 1{,}124 = 287{,}8$ (тыс. т).

Это один из способов прогнозирования с учетом сезонной волны (*мультипликативный*). Возможны и другие. Например, *аддитивный*, когда после нахождения выравненных уровней находится разность между эмпирическими и выравненными уровнями и затем при прогнозировании к значению уровня, полученному по линии тренда, прибавляется среднее значение упомянутых разностей ($y - \bar{y}_t$), рассчитанное для отдельных кварталов (или месяцев) за 2 или 3 года.

Рассмотрим задачу на корреляцию рядов динамики.

__Задача 9.9__

Имеются следующие данные по сельскохозяйственным предприятиям РФ за 5 лет:

Год	**Внесено минеральных удобрений под зерновые, кг/га, x**	**Урожайность зерновых, ц/га, y**
1992	52	17,2
1993	46	16,3
1994	24	14,4
1995	16	11,6
1996	17	12,9

Измерить корреляцию между уровнями двух рядов динамики (x и y), т.е. между урожайностью зерновых и количеством минеральных удобрений, внесенных на 1 га под зерновые.

Решение.

Коррелировать непосредственно уровни двух рядов можно лишь тогда, когда в каждом из них отсутствует автокорреляция, так как наличие последней может существенно повлиять на величину коэффициента, измеряющего зависимость между анализируемыми показателями.

Поэтому прежде чем применять ту или иную формулу для измерения корреляции между x и y, **каждый из этих рядов нужно проверить на автокорреляцию.**

1. **Проверим на автокорреляцию ряд x_t.** Для этого параллельно со значениями x_t запишем x_{t-1}, т.е. сдвинутые на единицу. А чтобы ряд не укорачивался и характеристики этих рядов были одинаковыми ($\bar{y}_t = \bar{y}_{t-1}$ и $\sigma_{y_t} = \sigma_{y_{t-1}}$), последнее значение x_t перенесем в первую строку значений x_{t-1}.

Для измерения автокорреляции между уровнями одного ряда используем следующую модификацию формулы коэффициента автокорреляции (r_a):

$$r_a = \frac{\overline{x_t x_{t-1}} - \overline{x}_t \, \overline{x}_{t-1}}{\sigma_{x_t} \sigma_{x_{t-1}}} = \frac{\overline{x_t x_{t-1}} - \left(\overline{x}_t\right)^2}{\sigma^2_{x_t}} = \frac{\sum x_t x_{t-1} - n\left(\overline{x}_t\right)^2}{\sum x_t^2 - n\left(\overline{x}_t\right)^2}.$$

Необходимые суммы рассчитаны ниже в таблице:

x_t	x_{t-1}	x_t^2	$x_t x_{t-1}$
52	(17)	2704	884
46	52	2116	2392
24	46	576	1104
16	24	256	384
17	16	289	272
$\sum$155	**155**	**5941**	**5036**

$$\overline{x}_t = \frac{155}{5} = 31.$$

Находим r_a (расчетное):

$$r_{a\,расч} = \frac{5036 - 5\cdot(31)^2}{5941 - 5\cdot(31)^2} =$$

$$= \frac{5036 - 4805}{5941 - 4805} = \frac{231}{1136} = 0{,}203.$$

По таблице Приложения 7 находим, что для $n = 5$ и 5%-ном уровне значимости ($\alpha = 0{,}05$) табличное значение $r_a = 0{,}253$. Так как $r_{a\,расч} < r_{a\,табл}$, то делаем вывод **об отсутствии автокорреляции** в ряду x_t.

2. **Проверим на автокорреляцию ряд y_t.**

В этом случае

y_t	y_{t-1}	y_t^2	$y_t y_{t-1}$
17,2	(12,9)	295,84	221,88
16,3	17,2	265,69	280,36
14,4	16,3	207,36	234,72
11,6	14,4	134,56	167,04
12,9	11,6	166,41	149,64
$\sum$ 72,4	**72,4**	**1069,86**	**1053,64**

$$\overline{y}_t = \frac{72{,}4}{5} = 14{,}48;$$

$$r_a = \frac{\sum y_t y_{t-1} - n\left(\overline{y}_t\right)^2}{\sum y_t^2 - n\left(\overline{y}_t\right)^2} =$$

$$= \frac{1053{,}64 - 5\cdot(14{,}48)^2}{1069{,}86 - 5\cdot(14{,}48)^2} = 0{,}246.$$

Так как и в этом случае $r_{a\,расч} < r_{a\,табл}$, то это свидетельствует **об отсутствии автокорреляции** в ряду y_t. Следовательно, приведенные в начале задачи ряды динамики (их уровни) можно коррелировать, используя для этой цели обычный линейный коэффициент корреляции $r_{xy} = \dfrac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x \sigma_y}$. Выпишем все имеющиеся значения и рассчитаем недостающие:

$\bar{x} = 31$, $\bar{y} = 14{,}48$;

$$\sigma_x^2 = \overline{x^2} - (\bar{x})^2 = \frac{5941}{5} - (31)^2 = 227{,}2;$$

$$\sigma_y^2 = \overline{y^2} - (\bar{y})^2 = \frac{1069{,}86}{5} - (14{,}48)^2 = 4{,}3;$$

$$\overline{xy} = \frac{\sum xy}{n} = \frac{52 \cdot 17{,}2 + 46 \cdot 16{,}3 + 24 \cdot 14{,}4 + 16 \cdot 11{,}6 + 17 \cdot 12{,}9}{5} =$$

$$= \frac{2394{,}7}{5} = 478{,}94.$$

Отсюда $r_{xy} = \dfrac{\overline{xy} - \bar{x}\,\bar{y}}{\sigma_x \sigma_y} = \dfrac{478{,}94 - 31 \cdot 14{,}48}{\sqrt{227{,}2 \cdot 4{,}3}} = 0{,}96$, т.е. зависимость между вариацией количества вносимых минеральных удобрений и вариацией урожайности зерновых в 1992—1996 гг. была очень большой.

Литература

[1, с. 263—337], [2, с. 344—407], [3, с. 257—313].

Контрольные вопросы

1. Что такое ряды динамики и какова их роль в статистическом анализе?
2. Как решается вопрос о сопоставимости уровней динамического ряда?
3. Какие существуют виды динамических рядов?
4. Как исчисляется средний уровень для различных рядов?
5. Какие основные показатели рассчитываются для анализа динамических рядов?
6. Как рассчитывается средний темп роста?
7. Чем вызывается необходимость обработки динамических рядов?
8. Какие существуют способы обработки динамических рядов?
9. Как измеряются сезонные колебания в динамических рядах?
10. Определите понятие автокорреляции в рядах динамики. Как она измеряется?
11. В чем состоят особенности корреляции рядов динамики?
12. Как рассчитывается критерий Дурбина—Ватсона и что он характеризует?

Задачи и упражнения для самостоятельной работы

Задача 1

Имеются следующие данные о производстве картофеля в хозяйствах населения РФ за 1991—1996 гг.:

Год	1991	1992	1993	1994	1995	1996
Производство картофеля, млн. т	24,8	29,9	31,1	29,8	35,9	34,9

Определить:

1) абсолютные приросты производства картофеля по годам (цепные);
2) цепные и базисные коэффициенты роста;
3) среднегодовой уровень производства картофеля за 1991—1996 гг.;
4) среднегодовой коэффициент роста производства картофеля в хозяйствах населения за 1992—1996 гг.

Ответ: 3) $\bar{y}$ = 31,07 млн. т; 4) $\bar{K}$ = 1,0707, или ≈ 107,1%.

Задача 2

Имеются следующие данные об остатках наличных денег у населения РФ в первой половине 1997 г.:

Месяц	Январь	Февраль	Март	Апрель	Май	Июнь	Июль
Остаток денег на начало месяца, трлн. руб.	75,8	70,5	74,5	77,1	84,7	88,4	100,5

Рассчитать:

1) средний остаток наличных денег у населения за январь—июнь;
2) среднемесячный темп роста наличных денег у населения за 6 месяцев 1997 г.

Ответ: 1) $\bar{y}$ = 80,6 трлн. руб.; 2) $\bar{K}$ = 1,048, или 104,8%.

Задача 3

Имеются следующие данные по РФ о вводе в действие жилых домов, построенных населением за свой счет и с помощью кредитов:

Год	1990	1991	1992	1993	1994	1995	1996
Введено млн. кв. м общей площади	6,0	5,4	4,9	5,6	7,1	9,0	10

Определить за 1991—1996 гг. среднегодовой коэффициент роста строительства жилья населением РФ, ориентированный на достижение:

а) конечного уровня 1996 г.;

б) общей площади, построенной за 1991—1996 гг.

О т в е т: а) $\overline{K} = 1{,}089$, или 108,9%;

б) $\overline{K} = 1{,}045$, или 104,5%.

Задача 4

Имеются следующие данные об остатках вкладов населения в банках РФ в первой половине 1997 г.:

Месяц	Январь	Февраль	Март	Апрель	Май	Июнь	Июль
Остаток денег на начало месяца, млрд. руб.	127,6	129,7	132,7	133,8	135,3	137,1	139,8

Определить:

1) средний остаток вкладов населения в банках РФ за 6 месяцев 1997 г.;

2) среднемесячный темп прироста вкладов.

О т в е т: 1) $\overline{y} = 133{,}7$ млрд. руб.;

2) $\overline{T}_{\text{пр}} = \overline{T}_{\text{р}} - 100\% = 101{,}5 - 100 = 1{,}5\%$.

Задача 5

Провести аналитическое выравнивание ряда динамики о производстве картофеля в хозяйствах населения РФ за 1992—1996 гг.

Год	1992	1993	1994	1995	1996
Производство картофеля, млн. т	29,9	31,1	29,8	35,9	34,9

О т в е т: уравнение тренда $\hat{y}_t = 27{,}88 + 1{,}48t$ (при отсчете времени от 1992 г., где $t = 1$).

Задача 6

По данным предыдущей задачи спрогнозировать производство картофеля в хозяйствах населения в 1998 и 1999 гг.

О т в е т: 38,24 млн. т; 39,72 млн. т.

Задача 7

По результатам решения задачи 5 найти остаточные величины $\left(y_t - \hat{y}_t = \varepsilon_t\right)$ и проверить их на автокорреляцию с помощью коэффициента автокорреляции Андерсона и критерия Дурбина–Ватсона.

Задача 8

Имеются следующие данные об урожайности зерновых в одном из регионов:

Год	1990	1991	1992	1993	1994	1995
Урожайность, ц/га	17,0	14,4	19,7	23,0	22,1	23,8

1. Проверить ряд на автокорреляцию.
2. Провести аналитическое выравнивание ряда.
3. Рассчитать теоретические значения урожайности, а также остаточную дисперсию.
4. С помощью критерия Дурбина–Ватсона проверить остаточные величины на автокорреляцию.
5. Рассчитать среднегодовой темп роста урожайности зерновых за 1991—1995 гг.

Задача 9

Имеются следующие данные по сельскохозяйственным предприятиям РФ об урожайности зерновых культур и обеспеченности зерноуборочными комбайнами в 1991—1996 гг.:

Год	Количество комбайнов на 1000 га x	Урожайность зерновых, ц/га, y
1991	6,5	14,4
1992	6,2	17,2
1993	6,2	16,3
1994	6,1	14,4
1995	5,8	11,6
1996	5,4	12,9

1. Проверить каждый ряд на автокорреляцию.

2. Провести аналитическое выравнивание каждого ряда и найти остаточные величины $d_x = x - \hat{x}_i$ и $d_y = y - \hat{y}_i$.
3. Измерить корреляцию между уровнями двух рядов.
4. Измерить корреляцию между остаточными величинами двух рядов.
5. Сравнить результаты, полученные в пп. 3 и 4, и сделать выводы о связи между x и y.

Задача 10

Имеются следующие данные по РФ:

Год	Число клубных учреждений, тыс. x	Зарегистрировано преступлений на 1000 чел. y
1990	73,2	12,4
1991	70,6	14,6
1992	66,0	18,6
1993	63,7	18,8
1994	61,3	17,7
1995	59,9	18,6

Измерить корреляцию между указанными показателями на основе корреляции первых разностей уровней в ряду x и y:

$$\Delta x = x_i - x_{i-1} \quad \text{и} \quad \Delta y = y_i - y_{i-1}.$$

Задача 11

Имеются следующие данные о вводе в действие жилых домов (млн. кв. м общей площади) в Москве:

Год	1990	1991	1992	1993	1994	1995
Введено млн. кв. м общей площади	2,26	2,47	2,48	2,28	2,47	2,55

1. Рассчитать среднегодовой темп роста ввода в действие жилых домов в Москве, ориентированный на достижение:

а) фактического уровня в 1995 г.;

б) общего объема жилья, введенного в действие за 1991—1995 гг.

2. Найти уравнение тренда и спрогнозировать ввод жилья в Москве в 1998 г.
3. Проверить остаточные величины на автокорреляцию.

Задача 12

По данным задачи 10:

1) рассчитать среднегодовые темпы роста (снижения) x и y за 1991—1995 гг.;
2) проверить ряд игреков на автокорреляцию;
3) измерить корреляцию между x и y с помощью коэффициентов корреляции рангов:

   а) Спирмэна;

   б) Кендэла.

Задача 13

Имеются следующие данные о мировых ценах на кофе в Бразилии (за 1 кг):

Год	1990	1991	1992	1993	1994	1995	1996
Цена, дол. США	1,3	1,27	0,95	1,1	2,55	2,73	2,21

1. Рассчитать основные аналитические показатели ряда динамики (абсолютные приросты, темпы роста и др.).
2. Определить уравнение тренда мировых цен на кофе.
3. Проверить остаточные величины на автокорреляцию.

Задача 14

Производство шерсти в фермерских хозяйствах РФ характеризуется следующими данными:

Год	1992	1993	1994	1995	1996
Произведено тыс. т	1,4	2,9	3,3	4,2	3,6

1. Определить среднегодовой темп роста производства шерсти за 1993–1996 гг.
2. Провести аналитическое выравнивание уровней ряда и на основе уравнения тренда спрогнозировать производство шерсти в 1999 г.
3. Измерить автокорреляцию в остаточных величинах с помощью коэффициентов автокорреляции Андерсона и критерия Дурбина—Ватсона.

Задача 15

Имеются следующие данные о производстве и импорте мяса и мясопродуктов в РФ (млн. т):

Год	Произ-водство x	Импорт y
1991	9,4	1,5
1992	8,3	1,4
1993	7,5	1,4
1994	6,8	1,7
1995	5,8	2,3
1996	5,3	2,1
1997	4,9	3,0
1998	4,7	2,3

Измерить тесноту зависимости между x и y с помощью:

а) коэффициента корреляции рангов Спирмэна;

б) коэффициента Кендэла;

в) первых разностей (Δx и Δy).

<u>Задача 16</u>

1. По данным задачи 15 проверить каждый ряд (x и y) на автокорреляцию.
2. При отсутствии автокорреляции в каждом ряду измерить корреляцию между двумя рядами непосредственно по значениям уровней.
3. При наличии автокорреляции в каждом ряду осуществить аналитическое выравнивание каждого ряда, рассчитать остаточные величины ($d_x = x - \hat{x}_t$ и $d_y = y - \hat{y}_t$) и измерить корреляцию между ними.

Сделать выводы по результатам расчетов.

<u>Задача 17</u>

Имеются следующие данные о динамике валового внутреннего продукта (ВВП) РФ (в сопоставимых ценах):

Год	1991	1992	1993	1994	1995	1996	1997	1998
ВВП, % к предыдущему году	95	85,5	91,3	87,3	95,9	96,6	100,9	95,4

1. Определить процент изменения ВВП в 1998 г. по сравнению с 1990 г.
2. Рассчитать среднегодовой темп изменения ВВП за 1991–1998 гг.

<u>Задача 18</u>

Имеются следующие данные о производстве шерсти (x) и шерстяных тканей (y) в РФ за 1992–1997 гг.:

Год	Производство шерсти, тыс. т x	Производство шерстяных тканей, млн. м² y
1992	179	276
1993	158	206
1994	122	91,1
1995	93	72,2
1996	77	50,3
1997	61	46,8

1. Рассчитать среднегодовой темп изменения за 1993–1997 гг.:
   а) производства шерсти;
   б) производства шерстяной ткани.
2. Проверить каждый ряд (x и y) на автокорреляцию уровней.
3. В случае отсутствия автокорреляции измерить корреляцию между уровнями двух рядов.

Задача 19

По данным задачи 18 осуществить аналитическое выравнивание ряда x и ряда y.

Проверить уравнения тренда на значимость (с помощью F-критерия Фишера).

Задача 20

Ниже приведены показатели производства чугуна в РФ за 1993–1998 гг.:

Год	1993	1994	1995	1996	1997	1998
Произведено млн. т	40,9	36,5	39,8	37,1	37,3	34,7

1. Рассчитать среднегодовой темп роста (изменения) производства чугуна за 1994–1998 гг.
2. Провести аналитическое выравнивание уровней ряда и спрогнозировать производство чугуна в 2000 г.

Задача 21

Имеются следующие данные о количестве проданных легковых автомобилей в РФ за 1990–1998 гг.:

Год	1990	1991	1992	1993	1994	1995	1996	1997	1998
Продано тыс. шт.	971	788	810	867	1176	1279	969	1550	1156

Определить среднегодовой темп роста продаж легковых автомобилей, ориентированный на обеспечение показателя:

а) конечного уровня (1998 г.);

б) общего количества автомобилей, проданных за 1991–1998 гг.

Тема 10
ЭКОНОМИЧЕСКИЕ ИНДЕКСЫ

Приступая к изучению темы, следует прежде всего отчетливо представлять, что **индексы** — это относительные показатели, характеризующие соотношение явлений во времени, в пространстве и по сравнению с планом. Индексы делятся на индивидуальные и общие.

Индивидуальные индексы (i) характеризуют относительное изменение единичного элемента сложной совокупности:

$$i_q = \frac{q_1}{q_0}$$ — индекс объема определенного продукта (товара);

$$i_p = \frac{p_1}{p_0}$$ — индекс цены определенного продукта;

$$i_y = \frac{y_1}{y_0}$$ — индекс урожайности отдельной культуры

и т.д.

Индивидуальные индексы показывают, каково соотношение между отчетным (со знаком «1») и базисным (со знаком «0») показателями или во сколько раз увеличилась (уменьшилась) индексируемая величина.

Общие индексы (I) характеризуют изменение индексируемого показателя в целом по сложной совокупности, отдельные элементы которой несоизмеримы в физических единицах (например, индекс объема промышленной продукции, индекс цен на продовольственные товары и т.п.).

Общие индексы могут быть построены как агрегатные или как средние из индивидуальных.

Построение **агрегатных** индексов сводится к тому, что с помощью определенных соизмерителей выражаются итоговые величины сложной совокупности в отчетном и базисном периодах, а затем первая сопоставляется со второй. Например, нужно показать изменение объема выпускаемой продукции на мебельной фабрике в 1998 г. (отчетный период) по сравнению с 1997 г. (базисный период). Фабрика выпускает столы, шкафы, диваны. Ясно, что сложить эту различную несоизмеримую продукцию в физических единицах нельзя. Но если представить всю продукцию в стоимостном выражении (приняв цены в качестве соизмерителя), то тогда можно сравнивать стоимость продукции одного года со стоимостью продукции другого года. А чтобы изменение цен не влияло на величину стоимостного показателя, продукцию двух лет надо оценить в одних и тех же ценах. Если выпуск продукции условно обозначить через q, а цены —

через p, то формула агрегатного индекса физического объема выразится следующим образом:

$$I_{\text{ф.об}} = \frac{\sum q_1 p}{\sum q_0 p} \quad \text{или} \quad I_{\text{ф.об}} = \frac{\sum q_1 p_0}{\sum q_0 p_0},$$

где q_1 и q_0 — количество продукции соответственно в отчетном и базисном периодах;

p и p_0 — цены соответственно сопоставимые и базисного периода.

Аналогично для построения агрегатного индекса цен определенный набор продуктов (q) оценивается в ценах двух периодов и сопоставляются стоимости набора в разных ценах. При этом количество этого набора (q) может приниматься на уровне базисного периода (q_0) или отчетного (q_1).

В первом случае индекс цен именуется **индексом Ласпейреса** (по имени немецкого ученого, предложившего этот метод) и записывается в виде формулы

$$I_{\text{цен Ласпейреса}} = \frac{\sum q_0 p_1}{\sum q_0 p_0}.$$

Если же принимается продукция в объеме (количестве) отчетного периода (q_1), то индекс цен носит имя своего автора — **индекс Пааше** и записывается

$$I_{\text{цен Пааше}} = \frac{\sum q_1 p_1}{\sum q_1 p_0}.$$

Общий индекс может быть построен и как **средний взвешенный из индивидуальных**. При этом надо помнить, что веса для индивидуальных индексов должны быть подобраны так, чтобы было обеспечено тождество среднего арифметического или гармонического индекса агрегатному.

Так, для индекса физического объема средний арифметический индекс будет иметь вид

$$I = \frac{\sum i\, q_0 p_0}{\sum q_0 p_0},$$

где $i = \frac{q_1}{q_0}$ — индивидуальные индексы объема;

$q_0 p_0$ — стоимость продукции базисного периода в базисных ценах (веса).

Нетрудно заметить, что этот индекс тождествен агрегатному:

$$\frac{\sum \frac{q_1}{q_0} q_0 p_0}{\sum q_0 p_0} = \frac{\sum q_1 p_0}{\sum q_0 p_0}.$$

Поскольку агрегатные индексы цен могут быть построены по формуле Ласпейреса или Пааше, то и средние из индивидуальных строятся, соответственно, по-разному. Так, средний арифметический индекс цен, тождественный агрегатному индексу Ласпейреса, исчисляется по формуле

$$\overline{I}_{\text{ар.(Л)}} = \frac{\sum i\, q_0 p_0}{\sum q_0 p_0},$$

где $i = \frac{p_1}{p_0}$ — индивидуальные индексы цен;

$q_0 p_0$ — веса.

А средний арифметический индекс цен, тождественный агрегатному индексу Пааше,

$$\overline{I}_{\text{ар.(П)}} = \frac{\sum i\, q_1 p_0}{\sum q_1 p_0},$$

т.е. веса здесь иные ($q_1 p_0$).

Соответственно, разную форму будут иметь и средние гармонические индексы цен.

Тождественный индексу Ласпейреса $\overline{I}_{\text{гарм.(Л)}} = \frac{\sum q_0 p_1}{\sum \frac{q_0 p_1}{i}}$, а тождественный индексу Пааше $\overline{I}_{\text{гарм.(П)}} = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{i}}$.

На практике всегда оговаривается, по какой методике рассчитывается тот или иной индекс цен.

Надо иметь в виду, что для средних индексов в качестве весов могут приниматься не только абсолютные показатели стоимости продукции (например, $q_0 p_0$ или $q_1 p_1$), но и относительные величины в виде долей или процентов отдельных групп товаров в структуре производства, потребления, товарооборота и пр.

Общий индекс цен может быть построен и как **средняя геометрическая из агрегатных индексов цен Ласпейреса и Пааше**, т.е. по формуле

$$I_{\text{цен}} = \sqrt{\frac{\sum q_0 p_1}{\sum q_0 p_0} \frac{\sum q_1 p_1}{\sum q_1 p_0}}\,.$$

Это так называемый **индекс Фишера**, рекомендуемый его автором в тех случаях, когда трудно отдать предпочтение весам q_0 или q_1. Поскольку в этой формуле учтены веса обоих периодов, Фишер считал этот индекс идеальным.

Следует также обратить внимание на то, что если строится р я д индексов, то они могут быть построены или как цепные (ряд индексов, каждый из которых построен по отношению к предыдущему периоду), или как базисные (ряд индексов, построенных в сравнении с одной и той же базой). Произведение цепных индексов дает базисный индекс. Путем деления двух базисных индексов легко получить цепной.

Особое место в статистике занимают так называемые индексы переменного и фиксированного состава, используемые при анализе динамики средних показателей.

Индексом переменного состава ($I_{\text{п с}}$) называют отношение двух средних уровней.

Если индексируемую величину обозначить через x, а веса через f, то в общем виде индекс переменного состава можно записать как

$$\overline{I}_{\text{п.с}} = \overline{x}_1 : \overline{x}_0 = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}.$$

Очевидно, что средняя величина показателя ($\overline{x}$) может меняться как за счет изменения значений осредняемого признака (x) у отдельных единиц, так и за счет изменения их весов (f), т.е. за счет изменения состава (структуры) совокупности. Это и является основанием для именования данного отношения средних величин индексом переменного состава.

Если при расчете средних величин за два периода зафиксировать веса одного и того же периода, то при сравнении таких средних влияние изменения структурного фактора будет устранено, и этот индекс называют **индексом фиксированного (или постоянного) состава** ($I_{\text{ф с}}$).

Веса при этом фиксируются, как правило, на уровне текущего периода (f_1), т.е.

$$I_{ф.с}=\frac{\sum x_1 f_1}{\sum f_1}:\frac{\sum x_0 f_1}{\sum f_1}.$$

Нетрудно заметить, что при сокращении на $\sum f_1$, этот индекс можно записать как $I_{ф.с}=\dfrac{\sum x_1 f_1}{\sum x_0 f_1}$, т.е. в агрегатном виде.

Индекс фиксированного состава характеризует среднее изменение самого осредняемого признака при постоянстве структуры совокупности.

При сравнении средних показателей можно принять неизменными значения x, тогда на динамику средних будет оказывать влияние только изменение весов, т.е. структуры совокупности. Этот индекс условно называют **индексом структуры** (или **индексом структурных сдвигов**) ($I_{стр}$):

$$I_{стр}=\frac{\sum x_0 f_1}{\sum f_1}:\frac{\sum x_0 f_0}{\sum f_0}, \text{ или } \quad I_{стр}=\frac{I_{п.с}}{I_{ф.с}},$$

т.е. индекс структуры можно получить, разделив индекс переменного состава на индекс фиксированного состава.

Индекс структуры показывает, в какой степени изменение средней величины индексируемого показателя произошло за счет изменения структуры (состава) совокупности.

Записанные выше в общем виде формулы индексов переменного и фиксированного состава, а также индекс структуры принимают тот или иной конкретный вид в зависимости от символики, используемой для отдельных показателей. Так, при изучении динамики средней урожайности зерновых, если обозначить урожайность отдельных культур через y, а посевную площадь под ними через $П$, индекс урожайности переменного состава будет иметь вид

$$I_{ур.п.с}=\bar{y}_1:\bar{y}_0=\frac{\sum y_1 П_1}{\sum П_1}:\frac{\sum y_0 П_0}{\sum П_0},$$

индекс урожайности фиксированного состава —

$$I_{ур.ф.с}=\frac{\sum y_1 П_1}{\sum П_1}:\frac{\sum y_0 П_1}{\sum П_1}=\frac{\sum y_1 П_1}{\sum y_0 П_1}$$

и индекс структуры —

$$I_{\text{стр}} = \frac{\sum y_0 П_1}{\sum П_1} : \frac{\sum y_0 П_0}{\sum П_0}, \text{ или } I_{\text{стр}} = I_{\text{п.с}} : I_{\text{ф.с}}.$$

Далее при решении задач будут приведены формулы индексов и для других показателей.

При изучении различных взаимосвязанных показателей следует иметь в виду, что индексы этих показателей находятся точно в такой же зависимости, как и сами показатели (индивидуальные индексы всегда, общие — при определенном построении). Например, если валовой сбор какой-либо культуры можно представить в виде показателя, зависящего от посевной площади и урожайности (валовой сбор = урожайность × посевная площадь), то и индекс валового сбора можно представить в виде произведения индекса посевных площадей на индекс урожайности. На основе такого рода взаимосвязанных индексов, зная два из трех, легко рассчитать третий.

Рассмотрим решение некоторых задач к данной теме.

Задача 10.1

Имеются следующие данные о продаже и ценах на продукты на одном из рынков города:

Продукт	Единица измерения	Продано, тыс. ед.		Цена единицы, руб.	
		в базисном периоде q_0	в отчетном периоде q_1	в базисном периоде p_0	в отчетном периоде p_1
А	л	50	60	3	2,5
Б	кг	40	50	2	1,5
В	кг	1,5	2	20	18

Определить:

1) общее изменение физического объема продаж;
2) общее изменение цен на указанные продукты;
3) абсолютную экономию населения от снижения цен.

Решение.

1. Общее (в среднем) изменение объема продаж определим по агрегатной формуле индекса физического объема:

$$I_{\text{ф.об}} = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{60 \cdot 3 + 50 \cdot 2 + 2 \cdot 20}{50 \cdot 3 + 40 \cdot 2 + 1{,}5 \cdot 20} = \frac{320}{260} = 1{,}23 \text{ (или } 123\%\text{)},$$

т.е. в отчетном периоде было продано продуктов на 23% больше (123 – 100 = 23), чем в базисном периоде.

2. Общий индекс цен, характеризующий среднее изменение цен на все продукты, определяем по формуле Пааше:

$$I_{ц} = \frac{\sum q_1 p_1}{\sum q_1 p_0} = \frac{60 \cdot 2,5 + 50 \cdot 1,5 + 2 \cdot 18}{60 \cdot 3 + 50 \cdot 2 + 2 \cdot 20} = \frac{261}{320} = 0,8156 \text{ (или } 81,56\%),$$

т.е. цены на все продукты снизились в среднем на 18,44% (81,56 – – 100 = –18,44).

3. Для ответа на третий вопрос вычтем из числителя агрегатной формулы индекса цен знаменатель:

$$\sum q_1 p_1 - \sum q_1 p_0 = 261 - 320 = -59 \text{ (тыс. руб.)},$$

т.е. абсолютная экономия населения от снижения цен составила 59 тыс. руб.

Задача 10.2

Определить среднее снижение цен на швейные изделия в отчетном периоде по сравнению с базисным по следующим данным:

Наименование швейных изделий	Снижение цен в отчетном периоде по сравнению с базисным, %	$i_{ц} = \frac{p_1}{p_0}$	Продано в отчетном периоде, млн. руб. $q_1 p_1$
Хлопчатобумажные	–20	0,80	10
Капроновые	–15	0,85	17

Решение.

В данном случае общий индекс цен может быть рассчитан из индивидуальных по формуле среднего гармонического индекса, тождественного агрегатному индексу Пааше:

$$I_{ц} = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{i_{ц}}} = \frac{10 + 17}{\frac{10}{0,8} + \frac{17}{0,85}} = \frac{27}{12,5 + 20} = \frac{27}{32,5} = 0,83 \text{ (или } 83\%),$$

т.е. цены в среднем снизились на 17% (83 – 100 = –17).

Задача 10.3

Имеются следующие данные о выпуске продукции мебельной фабрики:

Наименование изделий	Изменение выпуска в мае по сравнению с апрелем, %	Выпуск продукции в апреле, млн. руб., $q_0 p_0$
Столы	+12	20
Диваны	+10	50
Стулья	+15	30

Определить увеличение выпуска всей продукции в мае по сравнению с апрелем (в %), т.е. рассчитать общий индекс физического объема.

Решение.

Общий индекс физического объема может быть рассчитан как средний арифметический:

$$I_{ф.об} = \frac{\sum i\, q_0 p_0}{\sum q_0 p_0} = \frac{1{,}12 \cdot 20 + 1{,}1 \cdot 50 + 1{,}15 \cdot 30}{20 + 50 + 30} = 1{,}119 \text{ (или } 111{,}9\%\text{)},$$

т.е. в целом по предприятию выпуск продукции в мае по сравнению с апрелем увеличен на 11,9%.

Задача 10.4

Имеются следующие данные о динамике потребительских цен в РФ за 1994 г.:

Группа товаров	Индекс потребительских цен (1994 г. по отношению к 1993 г.), в разах i	Структура потребительских расходов в 1993 г. (по данным обследований семейных бюджетов), % d_0
Продовольственные товары	3,1	49,4
Непродовольственные товары	2,9	43,1
Платные услуги	7,6	7,5

Рассчитать сводный (общий) индекс потребительских цен.

Решение.

Находим общий индекс потребительских цен как средний арифметический из групповых, приняв в качестве весов долю отдельных товаров (в %) в общих расходах за 1993 г., т.е. по формуле, тождественной агрегатному индексу цен Ласпейреса:

$$\bar{I} = \frac{\sum i\, d_0}{\sum d_0} = \frac{3{,}1 \cdot 49{,}4 + 2{,}9 \cdot 43{,}1 + 7{,}6 \cdot 7{,}5}{100} = 3{,}35 \text{ раза},$$

т.е. в целом индекс потребительских цен составил 3,35 раза (рост в 3,35 раза).

Задача 10.5

Имеются следующие данные об изменении численности рабочих на заводе, в % к предыдущему году:

1991	1992	1993	1994	1995
+5	+4	+7	+5	+6

Определить, на сколько процентов увеличилось число рабочих на заводе за 5 лет, т.е. в 1995 г. по сравнению с 1990 г.

Решение.

Зная, что базисный индекс можно получить путем перемножения цепных, находим:

$$I_{\text{ч р. }95/90} = 1{,}05 \cdot 1{,}04 \cdot 1{,}07 \cdot 1{,}05 \cdot 1{,}06 = 1{,}30 \text{ (или } 130\%),$$

т.е. за 5 лет число рабочих на заводе возросло на 30%.

Задача 10.6

Имеются следующие данные о производстве и себестоимости продукта A по двум фабрикам за два периода:

Фабрика	Произведено, тыс. ед.		Себестоимость единицы продукта, руб.	
	в базисном периоде q_0	в отчетном периоде q_1	в базисном периоде c_0	в отчетном периоде c_1
№ 1	50	80	150	135
№ 2	60	40	250	230
$\sum$	**110**	**120**	—	—

Определить:

1) изменение себестоимости продукта A по каждой фабрике;
2) изменение себестоимости в целом по обеим фабрикам с помощью индексов переменного и фиксированного составов;
3) индекс структуры.

Решение.

1. Изменение себестоимости единицы продукта A по каждой фабрике определяем с помощью индивидуальных индексов:

а) по фабрике № 1: $i_1 = 135 : 150 = 0{,}9$ (или 90%), т.е. себестоимость снизилась на 10%;

б) по фабрике № 2: $i_2 = 230 : 250 = 0{,}92$ (или 92%), т.е. себестоимость снизилась на 8%.

2. Общий индекс себестоимости в данном случае может быть рассчитан как индекс переменного состава (сравнение средней себе-

стоимости по двум фабрикам за два периода) и как индекс фиксированного состава (характеризует среднее изменение себестоимости продукта *А* по двум фабрикам без учета влияния структурного фактора).

Чтобы рассчитать индекс себестоимости переменного состава, определяем среднюю по двум фабрикам себестоимость продукта *А* в отчетном и базисном периодах, а затем их сопоставляем. Средняя себестоимость в отчетном периоде $\left(\bar{c}_1\right)$:

$$\bar{c}_1 = \frac{\sum c_1 q_1}{\sum q_1} = \frac{135 \cdot 80 + 230 \cdot 40}{80 + 40} = 166{,}7 \text{ (руб.)}.$$

Средняя себестоимость в базисном периоде $\left(\bar{c}_0\right)$:

$$\bar{c}_0 = \frac{\sum c_0 q_0}{\sum q_0} = \frac{150 \cdot 50 + 250 \cdot 60}{50 + 60} = 204{,}5 \text{ (руб.)}.$$

Тогда индекс себестоимости переменного состава

$$I_{\text{с/ст. п.с}} = \frac{\sum c_1 q_1}{\sum q_1} : \frac{\sum c_0 q_0}{\sum q_0} = \bar{c}_1 : \bar{c}_0 = 166{,}7 : 204{,}5 =$$
$$= 0{,}815 \text{ (или } 81{,}5\%),$$

т.е. средняя по двум фабрикам себестоимость продукта *А* снизилась на 18,5%. Очевидно, что это снижение произошло как за счет снижения себестоимости на каждой фабрике, так и за счет влияния структурного фактора — увеличения выпуска более дешевого продукта на фабрике № 1.

Для устранения влияния структурного фактора рассчитываем индекс себестоимости фиксированного состава ($I_{\text{с/ст. ф с}}$):

$$I_{\text{с/ст. ф с}} = \frac{\sum c_1 q_1}{\sum q_1} : \frac{\sum c_0 q_1}{\sum q_1} = \frac{135 \cdot 80 + 230 \cdot 40}{120} : \frac{150 \cdot 80 + 250 \cdot 40}{120} =$$
$$= 166{,}7 : 183{,}3 = 0{,}909, \text{ или } 90{,}9\%,$$

т.е. себестоимость продукта *А* в среднем по двум фабрикам снизилась на 9,1%.

Этот же результат получим, сократив обе дроби на $\sum q_1$, т.е. применив формулу агрегатного индекса себестоимости:

$$I_{\text{с/ст. ф с}} = \frac{\sum c_1 q_1}{\sum c_0 q_1} = \frac{135 \cdot 80 + 230 \cdot 40}{150 \cdot 80 + 250 \cdot 40} = \frac{20\,000}{22\,000} = 0{,}909, \text{ или } 90{,}9\%.$$

3. Индекс структуры ($I_{\text{стр}}$) получим, разделив индекс себестоимости переменного состава на индекс фиксированного состава:

$$I_{\text{стр}} = I_{\text{п с}} : I_{\text{ф с}} = 0{,}815 : 0{,}909 = 0{,}896, \text{ или } 89{,}6\%.$$

Этот индекс показывает, как изменилась средняя себестоимость продукта *А* за счет структурного фактора, т.е. средняя себестоимость продукта *А* снизилась на 10,4% (89,6 – 100 = –10,4) за счет увеличения выпуска (доли) продукта *А* на фабрике № 1.

Индекс структуры (или структурных сдвигов) можно рассчитать и самостоятельно по формуле

$$I_{\text{стр}} = \frac{\sum c_0 q_1}{\sum q_1} : \frac{\sum c_0 q_0}{\sum q_0} = 183{,}3 : 204{,}5 = 0{,}896, \text{ или } 89{,}6\%.$$

Рассмотрим задачу на индексы урожайности.

Задача 10.7

Имеются следующие данные по РФ об урожайности и посевных площадях озимых зерновых культур в 1991 и 1995 гг. (первые пять граф таблицы).

Определить:

1. Общий индекс урожайности озимых зерновых культур:

 а) переменного состава;

 б) фиксированного состава.

2. Индекс структурных сдвигов.

Культура	Урожайность, ц/га		Посевная площадь, млн. га		Валовой сбор, млн. ц		$у_0П_1$
	1991 $у_0$	1995 $у_1$	1991 $П_0$	1995 $П_1$	1991 $у_0П_0$	1995 $у_1П_1$	
Пшеница	28,1	16,9	9,2	8,2	258,52	138,58	230,42
Рожь	16,4	12,6	6,5	3,2	106,60	40,32	52,48
Ячмень	35,1	28,3	0,78	0,47	27,38	13,30	16,50
$\sum$	–	–	**16,48**	**11,87**	**392,50**	**192,20**	**299,40**

Решение.

1. Определяем индекс урожайности переменного состава по формуле

$$I_{\text{ур. п.с}} = \frac{\sum y_1 П_1}{\sum П_1} : \frac{\sum y_0 П_0}{\sum П_0} = \bar{y}_1 : \bar{y}_0,$$

т.е. как отношение средней урожайности зерновых в 1995 г. к средней урожайности в 1991 г. Все необходимые данные приведены в таблице. Подставляя их в формулу, получаем:

$$I_{\text{ур. п.с}} = \frac{192,2}{11,87} : \frac{392,5}{16,48} = 16,2 : 23,8 = 0,68, \text{ или } 68\%,$$

т.е. средняя урожайность зерновых (озимых) снизилась за указанный период на 32%, или на 7,6 ц/га ($16,2 - 23,8 = -7,6$).

2. Находим индекс урожайности фиксированного состава, устраняющего влияние изменения структуры посевных площадей на динамику средней урожайности:

$$\text{а) } I_{\text{ур. ф.с}} = \frac{\sum y_1 П_1}{\sum П_1} : \frac{\sum y_0 П_1}{\sum П_1} \quad \text{или б) } I_{\text{ур. ф.с}} = \frac{\sum y_1 П_1}{\sum y_0 П_1},$$

где $\sum y_0 П_1$ — *валовой сбор*, который был бы получен с площади 1995 г. при урожайности 1991 г. (последняя графа таблицы).

Подставляя все значения в формулу «а», получаем:

$$I_{\text{ур. ф.с}} = \frac{192,2}{11,87} : \frac{299,4}{11,87} = 16,2 : 25,2 = 0,64, \text{ или } 64\%.$$

Это означает, что за счет изменения урожайности отдельных культур средняя урожайность зерновых снизилась на 36%, или на 9 ц/га ($16,2 - 25,2 = -9$).

Такой же относительный результат получим и по формуле «б»:

$$I_{\text{ур. ф.с}} = \frac{\sum y_1 П_1}{\sum y_0 П_1} = \frac{192,2}{299,4} = 0,64.$$

Однако пользуясь этой формулой, нельзя показать изменение средней урожайности зерновых в абсолютном выражении.

3. Находим индекс структуры, отражающий влияние изменения структуры посевных площадей на динамику средней урожайности зерновых, по двум формулам:

а) $I_{стр} = I_{п.с} : I_{ф.с} = 0{,}68 : 0{,}64 = 1{,}06$
или

$$б)\ I_{стр} = \frac{\sum y_0 П_1}{\sum П_1} : \frac{\sum y_0 П_0}{\sum П_0} = 25{,}2 : 23{,}8 = 1{,}06,\ \text{или } 106\%.$$

Это означает, что средняя урожайность зерновых возросла на 6%, или на 1,4 ц/га (25,2 – 23,8 = 1,4) за счет изменения структуры посевных площадей, в частности за счет увеличения в 1995 г. доли посева пшеницы и ячменя и уменьшения доли посева ржи (менее урожайной).

Таким образом, общее снижение средней урожайности (–7,6 ц/га) обусловлено двумя факторами: снижением урожайности отдельных культур и изменением структуры посевных площадей, т.е. –7,6 = –9 + 1,4.

Задача 10.8

По данным задачи 10.7 определить абсолютное изменение валового сбора озимых зерновых культур в 1995 г. по сравнению с 1991 г. и разложить его по факторам, т.е. показать, какая часть этого прироста (убыли) получена за счет изменения:

1) размера посевных площадей;
2) урожайности отдельных культур;
3) структуры посевных площадей.

Решение.

Определяем в абсолютном выражении изменение валового сбора озимых зерновых ($\Delta_{в\,сб}$) в 1995 г. по сравнению с 1991 г.:

$$\Delta_{в\,сб} = \sum y_1 П_1 - \sum y_0 П_0 = 192{,}2 - 392{,}5 = -200{,}3\ (\text{млн. ц}),$$

т.е. имело место уменьшение валового сбора на 200,3 млн. ц, в том числе:

1) за счет изменения размера посевных площадей (находим разность их сумм и умножаем на величину средней урожайности 1991 г.):

$$\Delta_{в\,сб\,(\sum П)} = \left(\sum П_1 - \sum П_0\right) \bar{y}_0 = (11{,}87 - 16{,}48) \cdot 23{,}8 = -109{,}7\ (\text{млн. ц});$$

2) за счет изменения урожайности отдельных культур (эту величину легко получить как разность между числителем и знаменателем индекса урожайности фиксированного состава в агрегатном виде):

$$\Delta_{в\,сб\,(y)} = \sum y_1 П_1 - \sum y_0 П_1 = 192{,}2 - 299{,}4 = -107{,}2\ (\text{млн. ц});$$

3) за счет изменения структуры посевных площадей:

$$\Delta_{\text{в.сб (стр. }П)} = \left(\frac{\sum y_0 П_1}{\sum П_1} : \frac{\sum y_0 П_0}{\sum П_0} \right) \cdot \sum П_1 = (25{,}2 - 23{,}8) \cdot 11{,}87 =$$

$$= 16{,}6 \text{ (млн. ц).}$$

Таким образом, делаем вывод, что общее уменьшение валового сбора озимых зерновых в 1995 г. по сравнению с 1991 г. (на 200,3 млн. ц) обусловлено тремя факторами:

1) на 109,7 млн. ц **уменьшился за счет сокращения размера посевных площадей** под озимыми зерновыми культурами;
2) на 107,2 млн. ц **уменьшился за счет снижения урожайности** отдельных культур;
3) на 16,6 млн. ц **увеличился за счет изменения структуры** посевных площадей,

т.е. в целом $-200{,}3 = -109{,}7 + (-107{,}2) + 16{,}6$.

Сумму двух последних слагаемых можно рассматривать как изменение валового сбора за счет изменения средней урожайности озимых зерновых ($-107{,}2 + 16{,}6 = -90{,}6$ млн. ц).

Эту же величину, т.е. изменение валового сбора за счет *динамики средней урожайности*, можно получить следующим расчетом:

$$\Delta_{\text{в.сб}(\bar{y})} = \left(\bar{y}_1 - \bar{y}_0\right) \cdot \sum П_1 = \left(\frac{\sum y_1 П_1}{\sum П_1} - \frac{\sum y_0 П_0}{\sum П_0} \right) \cdot \sum П_1 =$$

$$= \sum y_1 П_1 - \sum y_0 П_0 \cdot \frac{\sum П_1}{\sum П_0} = 192{,}2 - 392{,}5 \cdot \frac{11{,}87}{16{,}48} = -90{,}5 \text{ (млн. ц).}$$

(Небольшое расхождение ($-90{,}6$ и $-90{,}5$) вызвано округлениями в расчетах.)

<u>Задача 10.9</u>

Определить изменение производительности труда на фабрике, если известно, что за отчетный период объем выпускаемой продукции увеличился в 1,2 раза, а численность работающих возросла на 12%.

Решение.

На основе взаимосвязанных индексов можем записать

$$I_{\text{произв. труда}} = I_{\text{объема продукции}} : I_{\text{числа работающих}} =$$

$$= 1{,}2 : 1{,}12 = 1{,}071, \text{ или } 107{,}1\%,$$

т.е. производительность труда возросла за отчетный период на 7,1% ($107{,}1 - 100 = 7{,}1$).

Литература

[1, с. 338—397], [2, с. 439—481], [3, с. 314—353].

Контрольные вопросы

1. Какие индексы называются общими, а какие индивидуальными?
2. Какие существуют способы построения общих индексов?
3. В чем суть построения агрегатных индексов?
4. Запишите формулы агрегатных индексов цен (Ласпейреса и Пааше), а также индекса физического объема.
5. Как строятся средние индексы из индивидуальных? Запишите формулы среднего арифметического и среднего гармонического индексов цен (и других).
6. Какие индексы называют цепными и какие базисными?
7. Какая существует связь между цепными и базисными индексами?
8. В чем суть индексов переменного и фиксированного составов?
9. Приведите примеры взаимосвязанных индексов.
10. Что характеризует индекс структурных сдвигов и как он рассчитывается?
11. Как определить роль отдельных факторов в динамике сложных показателей (относительно и абсолютно)?
12. Какие проблемы возникают при исчислении территориальных индексов и как они решаются?

Задачи и упражнения для самостоятельной работы

Задача 1

Имеются следующие данные за два периода о ценах и объемах реализации трех видов товаров по одному из торговых предприятий:

Вид товара	Базисный период		Текущий период	
	Цена за единицу, руб. p_0	Продано товаров, шт. q_0	Цена за единицу, руб. p_1	Продано товаров, шт. q_1
А	45	2500	87	1700
Б	27	830	35	2300
В	12	610	14	1000

Рассчитать:

1) индивидуальные индексы цен (по каждому виду товаров);
2) индивидуальные индексы физического объема реализации товаров;
3) общий индекс цен:
   а) Ласпейреса;
   б) Пааше;
   в) Фишера;
4) общий индекс физического объема (по методу Ласпейреса);
5) индекс товарооборота (стоимость товаров).

Ответ: 1) 1,93 раза, 1,29, 1,17; 2) 0,68, 2,77, 1,64; 3) а) 179%, б) 161%, в) 169%; 4) 105,9%; 5) 170,4%.

Задача 2

Имеются следующие данные об изменении физического объема розничного товарооборота в РФ в 1995 г.:

Группа товаров	Индекс физического объема товарооборота 1995 г., % к 1994 г.	Структура товарооборота в 1994 г., %
Продовольственные	90,5	42
Непродовольственные	94,8	58

Рассчитать общий индекс физического объема розничного товарооборота РФ в 1995 г. по сравнению с 1994 г.

Ответ: 93%.

Задача 3

Имеются следующие данные по РФ об изменении физического объема розничного товарооборота ($I_{\text{ф.об. т/об}}$) по группам товаров в 1995 и 1996 гг. (% к 1990 г.):

Группа товаров	Структура розничного товарооборота, % к итогу		$I_{\text{ф.об. т/об}}$, % к 1990 г.		$I_{\text{ф.об. т/об 1996}}$, % к 1995 г.
	1990	1996	1995	1996	
Продовольственные	47	48	80	78	?
Непродовольственные	53	52	96	90	?
Все товары	**100**	**100**	**?**	**?**	**?**

1. Определить по отношению к 1990 г. общий индекс физического объема товарооборота:

   а) в 1995 г.; б) в 1996 г.
2. Определить индекс физического объема товарооборота в 1996 г. по сравнению с 1995 г.:

   а) по каждой группе товаров;

   б) в целом по всем товарам.

О т в е т: 1. а) 88,48%; б) 84,36%.
2. а) 97,5%, 93,75%; б) 95,34%.

Задача 4

Динамика промышленного производства РФ за 1991—1996 гг. характеризуется следующими данными:

Год	1991	1992	1993	1994	1995	1996
Индекс физического объема, % к предыдущему периоду	92	82	86	79	97	96

Определить:

1) индекс физического объема промышленного производства в 1996 г. по сравнению с 1990 г.;
2) среднегодовой темп снижения производства за указанный период.

О т в е т: а) 47,7%; б) –11,6%.

Задача 5

Имеются данные по РФ об индексе потребительских цен (ИПЦ) в 1995 г.:

Товары и услуги	ИПЦ к предыдущему периоду (1994 г.), в разах	Структура потребительских расходов в базисном периоде (1994 г.)
Продовольственные товары	2,0	49,8
Непродовольственные товары	2,4	?
Платные услуги	2,8	?
Сводный индекс	**2,3**	**100,0**

Определить в структуре потребительских расходов долю:

1) непродовольственных товаров;
2) платных услуг.

О т в е т: 1) 25,4%; 2) 24,8%.

Задача 6

Рассчитать по приводимым ниже данным общие индексы:

1) физического объема товарооборота;
2) цен;
3) стоимости продукции.

Товар	Индивидуальный индекс цен, %	Стоимость проданной продукции, тыс. руб.	
		июль	август
Картофель	104	118	99
Молоко	102	26	28
Яйца	96	142	155

Ответ: 1) 99,3%; 2) 99,26%; 3) 98,6%.

Задача 7

Имеются следующие данные по РФ об индексах потребительских цен на товары и услуги населению (% к предыдущему месяцу):

1995	ИПЦ на все товары и услуги	В том числе		
		продовольственные товары	непродовольственные товары	платные услуги
Январь	118	121	112	129
Март	109	108	109	110

Определить структуру расходов населения по указанным группам товаров и услуг.

Ответ: 1) 23,1%; 2) 53,8%; 3) 23,1%.

Задача 8

Определить долю товаров народного потребления и долю платных услуг в общем объеме товарооборота в 1990 г. по РФ, имея следующие данные:

Товары и услуги	Изменение цен и тарифов (1991 г. в % к 1990 г.)	Объем товарооборота и услуг в 1990 г. (% к итогу)
Товары народного потребления	195	?
Платные услуги населению	170,6	?
Итого	**192,7**	**100**

Ответ: 1) 90,5%; 2) 9,5%.

Задача 9

Имеются следующие данные по предприятию за два года:

Вид продукции	1994		1995	
	Себестоимость единицы продукции, руб.	Количество произведенной продукции, тыс. шт.	Себестоимость единицы продукции, руб.	Количество произведенной продукции, тыс. шт.
А	300	22	400	25
Б	700	13	820	10
В	80	9	100	18

1. Рассчитать сводный (общий) индекс себестоимости продукции.
2. Определить изменение общих затрат на производство (относительное и абсолютное) и разложить абсолютный прирост (уменьшение) по факторам: а) за счет изменения себестоимости единицы продукции отдельных видов и б) за счет изменения количества произведенной продукции.

Ответ: 1) 125,5%; 2) 121,8%, + 3580 тыс. руб., в том числе:
а) 4060 тыс. руб.; б) –480 тыс. руб.

Задача 10

Имеются следующие данные по РФ об урожайности и валовом сборе пшеницы в 1995 г. и в 1996 г.

Культура	Урожайность, ц/га		Валовой сбор, млн. ц	
	1995	1996	1995	1996
Пшеница озимая	16,9	17,9	138	167
Пшеница яровая	10,3	11,0	163	182

Рассчитать:

1) индексы урожайности пшеницы: а) переменного; б) фиксированного состава;
2) индекс структурных сдвигов (влияния изменения структуры посевных площадей на динамику средней урожайности);
3) изменение (в абсолютном выражении) валового сбора пшеницы в 1996 г. по сравнению с 1995 г. — всего и в том числе за счет изменения: а) урожайности; б) посевных площадей (размера); в) структуры посевных площадей.

Ответ: 1) а) 107,5%; б) 106,3%; 2) 101%;
3) 48 млн. ц:
а) 20,85 млн. ц; б) 23,58 млн. ц; в) 3,57 млн. ц.

Задача 11

Имеются следующие данные за 1996 и 1997 гг. о средней заработной плате и численности занятых в трех отраслях экономики РФ:

Отрасль	Численность занятых, млн. чел.		Среднемесячная заработная плата, руб.	
	1996	1997	1996	1997
Промышленность	16,4	15,5	869	1137
Строительство	5,9	5,6	967	1363
Сельское и лесное хозяйство	9,5	9,3	382	423

1. Определить:
   а) изменение заработной платы в 1997 г. по сравнению с 1996 г. в каждой отрасли;
   б) индекс заработной платы (переменного состава и фиксированного состава) в целом по трем отраслям, а также индекс структурных сдвигов;
   в) изменение фонда заработной платы в целом по трем отраслям: абсолютно и относительно.
2. Разложить абсолютное изменение фонда заработной платы по факторам:
   а) за счет изменения общей численности занятых;
   б) за счет изменения уровня оплаты труда;
   в) за счет изменения доли занятых в отдельных отраслях.

Задача 12

Имеются следующие данные по промышленному предприятию:

Изделие	Общие затраты на производство в 2000 г., млн. руб.	Изменение себестоимости изделия в 2000 г. по сравнению с 1999 г., %
А	69,12	+8
Б	126,72	+5,6
В	221,4	+2,5

Определить:
1) общий индекс себестоимости продукции в 2000 г. по сравнению с 1999 г.;
2) размер экономии или дополнительных затрат, вызванных изменением себестоимости продукции.

Задача 13

Валовой сбор зерна (в хозяйствах всех категорий) в РФ в 1998 г. составил 47,9 млн. т, что на 24,4% меньше, чем в 1995 г. Посевные площади под зерновыми составили: в 1995 г. — 54,7 млн. га, а в 1998 г. — на 7,3% меньше.

Определить:
1) среднюю урожайность зерновых в 1995 г. и 1998 г. (в ц/га);
2) индекс средней урожайности зерновых в 1998 г. по сравнению с 1995 г.;
3) изменение за указанный период валового сбора (абсолютное и относительное), в том числе за счет изменения:
   а) средней урожайности;
   б) размера посевных площадей под зерновыми.

ПРИЛОЖЕНИЯ
(математико-статистические таблицы)

ПРИЛОЖЕНИЕ 1

Значения функции $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$

t	0	1	2	3	4	5	6	7	8	9
0,0	3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3725	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0203	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
4,0	0001	0001	0001	0000	0000	0000	0000	0000	0000	0000

ПРИЛОЖЕНИЕ 2

Значение интеграла вероятностей $F(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{-t}^{+t} e^{-\frac{t^2}{2}} dt$

t	Сотые доли									
	0	1	2	3	4	5	6	7	8	9
0,0	0000	0080	0160	0239	0319	0399	0478	0558	0638	0717
0,1	0797	0876	0955	1034	1114	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2961	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6679	6729	6778
1,0	6827	6875	6923	6970	7017	7063	7109	7154	7199	7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7995	8030
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9089
1,7	9109	9127	9146	9164	9182	9199	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9425	9439	9451	9464	9476	9488	9500	9512	9523	9534
2,0	9545	9556	9566	9576	9586	9596	9606	9615	9625	9634
2,1	9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9755	9762	9768	9774	9780
2,3	9785	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9840	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9909	9912	9915	9917	9920	9924	9926	9927	9929
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9950	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	99730	99739	99747	99755	99763	99771	99779	99786	99793	99800
3,1	99807	99813	99819	99825	99831	99837	99842	99847	99853	99858
3,2	99863	99867	99872	99876	99880	99884	99889	99892	99896	99900
3,3	99903	—	—	—	—	—	—	—	—	—
3,4	99933	—	—	—	—	—	—	—	—	—
3,5	99953	—	—	—	—	—	—	—	—	—
4,0	99994	—	—	—	—	—	—	—	—	—
5,0	99999	—	—	—	—	—	—	—	—	—

$S(t)$ в распределении Стьюдента

t \ n	1	2	3	4	5	6–7	8–10	11–15	16–25	25–30	∞
0,0	500	500	500	500	500	500	500	500	500	500	500,000
0,1	532	535	537	537	538	538	539	539	539	539	539,827
0,2	563	570	573	574	576	575	578	578	578	578	579,259
0,3	593	606	608	610	612	613	615	616	616	616	617,911
0,4	621	636	642	645	647	649	651	652	653	654	655,421
0,5	648	667	674	678	681	683	685	687	689	689	691,462
0,6	672	695	705	710	713	715	718	721	722	724	725,746
0,7	694	723	733	739	742	746	749	752	754	756	758,036
0,8	715	746	759	766	770	774	778	781	783	785	788,144
0,9	733	768	783	790	795	800	804	808	811	813	815,939
1,0	750	789	804	813	818	823	828	832	835	838	841,344
1,1	765	807	824	834	839	844	850	854	858	860	864,333
1,2	779	824	842	852	858	864	870	874	878	881	884,930
1,3	791	838	858	868	875	881	887	892	896	899	903,199
1,4	803	852	872	883	890	896	902	907	912	915	919,243
1,5	813	864	885	896	903	909	916	921	925	928	933,192
1,6	822	875	896	908	915	921	928	933	937	940	945,200
1,7	831	884	906	918	925	932	938	943	948	951	955,434
1,8	839	893	915	927	934	941	947	952	956	959	964,069
1,9	846	901	923	935	942	948	955	960	964	967	971,283
2,0	852	908	930	942	949	955	962	967	970	973	977,249
2,1	858	915	937	948	955	961	967	972	976	978	982,135
2,2	864	921	942	954	960	966	972	977	980	982	986,096
2,3	870	926	948	958	965	971	977	981	984	986	989,275
2,4	874	931	952	963	969	975	980	984	987	989	991,802
2,5	879	935	956	966	973	978	983	987	989	991	993,790
2,6	883	939	960	970	976	981	986	989	991	993	995,338
2,7	887	943	963	973	979	983	988	991	993	995	996,533
2,8	891	946	966	976	981	985	990	993	995	996	997,444
2,9	894	949	969	978	983	987	991	994	996	997	998,134
3,0	898	952	971	980	985	989	993	995	997	997	998,650
3,1	901	955	973	982	987	990	994	996	997	998	999,032
3,2	904	957	975	984	988	991	995	997	998	998	999,312
3,3	906	960	977	985	989	992	995	997	998	999	999,516
3,4	909	962	979	986	990	993	996	998	998	999	999,663
3,5	911	964	980	988	991	993	997	998	999	—	999,767
3,6	914	965	982	989	992	994	997	998	—	—	999,840
3,7	916	967	983	990	993	995	998	999	—	—	999,892
3,8	918	969	984	990	994	996	998	999	—	—	999,927
3,9	920	970	985	991	994	996	998	999	—	—	999,951
4,0	922	971	986	992	995	997	998	—	—	—	999,968
4,1	924	973	987	993	995	997	999	—	—	—	999,979
4,2	926	974	988	993	996	998	999	—	—	—	999,986
4,3	927	975	988	994	996	998	999	—	—	—	999,991
4,4	929	976	989	994	996	998	—	—	—	—	999,994
4,5	930	977	989	995	997	998	—	—	—	—	999,996

t \ n	1	2	3	4	5	6–7	8––10	11––15	16––25	25––30	∞
4,6	932	978	990	995	997	998	—	—	—	—	999,997
4,7	933	979	991	995	997	999	—	—	—	—	999,998
4,8	935	980	991	996	998	999	—	—	—	—	999,999
4,9	936	980	992	996	998	999	—	—	—	—	999,999
5,0	937	981	992	996	998	999	—	—	—	—	999,999
5,1	938	982	993	996	998	—	—	—	—	—	999,999
5,2	940	982	993	997	998	—	—	—	—	—	999,999
5,3	941	983	993	997	998	—	—	—	—	—	999,999
5,4	942	984	994	997	998	—	—	—	—	—	—
5,5	943	984	994	997	999	—	—	—	—	—	—
5,6	943	984	994	997	999	—	—	—	—	—	—
5,7	945	985	995	998	999	—	—	—	—	—	—
5,8	946	986	995	998	999	—	—	—	—	—	—
5,9	947	986	995	998	999	—	—	—	—	—	—
6,0	947	987	995	998	—	—	—	—	—	—	—

ПРИЛОЖЕНИЕ 4

Значение χ^2-критерия Пирсона при уровне значимости 0,10, 0,05, 0,01

df(ν)	0,10	0,05	0,01	*df*(ν)	0,10	0,05	0,01
1	2,71	3,84	6,63	21	29,62	32,67	38,93
2	4,61	5,99	9,21	22	30,81	33,92	40,29
3	6,25	7,81	11,34	23	32,01	34,17	41,64
4	7,78	9,49	13,28	24	33,20	36,42	42,98
5	9,24	11,07	15,09	25	34,38	37,65	44,31
6	10,64	12,59	16,81	26	35,56	38,89	45,64
7	12,02	14,07	18,48	27	36,74	40,11	46,96
8	13,36	15,51	20,09	28	37,92	41,34	48,28
9	14,68	16,92	21,67	29	39,09	42,56	49,59
10	15,99	18,31	23,21	30	40,26	43,77	50,89
11	17,28	19,68	24,72	40	51,80	55,76	63,69
12	18,55	21,03	26,22	50	63,17	67,50	76,15
13	19,81	22,36	27,69	60	74,40	79,08	88,38
14	21,06	23,68	29,14	70	85,53	90,53	100,42
15	22,31	25,00	30,58	80	96,58	101,88	112,33
16	23,54	26,30	32,00	90	107,56	113,14	124,12
17	24,77	27,59	33,41	100	118,50	124,34	135,81
18	25,99	28,87	34,81				
19	27,20	30,14	36,19				
20	28,41	31,41	37,57				

ПРИЛОЖЕНИЕ 5

Значения критерия Дурбина—Ватсона при 5%-ном уровне существенности (для положительной автокорреляции)*

Число наблюдений n	$v = 1$		$v = 2$		$v = 3$	
	d_1	d_2	d_1	d_2	d_1	d_2
15	1,08	1,36	0,95	1,54	0,82	1,75
16	1,10	1,37	0,98	1,54	0,86	1,73
17	1,13	1,38	1,02	1,54	0,90	1,71
18	1,16	1,39	1,05	1,53	0,93	1,69
19	1,18	1,40	1,08	1,53	0,97	1,68
20	1,20	1,41	1,10	1,54	1,00	1,68
30	1,35	1,49	1,28	1,57	1,21	1,65
50	1,50	1,59	1,46	1,63	1,42	1,67

* v – число переменных в уравнении регрессии.

ПРИЛОЖЕНИЕ 6

Значения функции P (λ)

λ	P	λ	P	λ	P
0,30	1,0000	0,80	0,5441	1,60	0,0120
0,35	0,9997	0,85	4653	1,70	0062
0,40	9972	0,90	3927	1,80	0032
0,45	9874	0,95	3275	1,90	0015
0,50	9639	1,00	2700	2,00	0007
0,55	9228	1,10	1777	2,10	0003
0,60	8643	1,20	1122	2,20	0001
0,65	7920	1,30	0681	2,30	0001
0,70	7112	1,40	0397	2,40	0000
0,75	6272	1,50	0222	2,50	0000

ПРИЛОЖЕНИЕ 7

Критические значения коэффициентов автокорреляции (r_e) при уровнях значимости $\alpha = 0,05$ и $\alpha = 0,01$

Объем выборки n	Положительные значения		Отрицательные значения	
	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,01$
5	0,253	0,297	−0,753	−0,798
6	0,345	0,447	−0,708	−0,863
7	0,370	0,510	−0,674	−0,799
8	0,371	0,531	−0,625	−0,764
9	0,366	0,533	−0,593	−0,737
10	0,360	0,525	−0,564	−0,705
11	0,353	0,515	−0,539	−0,679
12	0,348	0,505	−0,516	−0,655
13	0,341	0,495	−0,497	−0,634
14	0,335	0,485	−0,479	−0,615
15	0,328	0,475	−0,462	−0,597
20	0,299	0,432	−0,399	−0,524

Значение *F*-критерия Фишера при уровне значимости 0,05

df_2 (v_2)	df_1 (v_1)																
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**14**	**16**	**20**	**30**	∞
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	250	254
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41	19,42	19,43	19,44	19,46	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74	8,71	8,69	8,66	8,62	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91	5,87	5,84	5,80	5,74	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68	4,64	4,60	4,56	4,50	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,96	3,92	3,87	3,81	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57	3,52	3,49	3,44	3,38	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28	3,23	3,20	3,15	3,08	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07	3,02	2,98	2,93	2,86	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91	2,86	2,82	2,77	2,70	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79	2,74	2,70	2,65	2,57	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69	2,64	2,60	2,54	2,46	2,30
13	4,67	3,81	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60	2,55	2,51	2,46	2,38	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53	2,48	2,44	2,39	2,31	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48	2,43	2,39	2,33	2,25	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42	2,37	2,33	2,28	2,20	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38	2,33	2,29	2,23	2,15	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,29	2,25	2,19	2,11	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,34	2,31	2,26	2,21	2,15	2,07	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,31	2,28	2,23	2,18	2,12	2,04	1,84

df_2 (ν_2)	df_1 (ν_1)															
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**14**	**16**	**20**	**30**	**∞**
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,20	2,15	2,09	2,00	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	2,26	2,23	2,18	2,13	2,07	1,98	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28	2,24	2,20	2,14	2,10	2,04	1,96	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,22	2,18	2,13	2,09	2,02	1,94	1,73
25	4,24	3,88	2,99	2,76	2,60	2,49	2,41	2,34	2,28	2,24	2,20	2,16	2,11	2,06	2,00	1,92	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,10	2,05	1,99	1,90	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,30	2,25	2,20	2,16	2,13	2,08	2,03	1,97	1,88	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,19	2,15	2,12	2,06	2,02	1,96	1,87	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,05	2,00	1,94	1,85	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,16	2,12	2,09	2,04	1,99	1,93	1,84	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	2,04	2,00	1,95	1,90	1,84	1,74	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,98	1,95	1,90	1,85	1,78	1,69	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,86	1,81	1,75	1,65	1,39
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,88	1,85	1,79	1,75	1,68	1,57	1,28
∞	3,84	2,99	2,60	2,37	2,21	2,09	2,01	1,94	1,88	1,83	1,79	1,75	1,69	1,64	1,57	1,46	1,00

Примечание: df_1 (ν_1) — число степеней свободы для большей дисперсии; df_2 (ν_2) — число степеней свободы для меньшей дисперсии.

ПРИЛОЖЕНИЕ 9

Значение t-критерия Стьюдента при уровне значимости 0,10, 0,05, 0,01

df (ν)	α			df (ν)	α		
	0,10	0,05	0,01		0,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

ПРИЛОЖЕНИЕ 10

Таблица для расчета средних коэффициентов роста (снижения) по средней параболической:

$$\bar{k} + \bar{k}^2 + \bar{k}^3 + \ldots + \bar{k}^n = \frac{\sum_{1}^{n} y_i}{y_0}$$

$\bar{k}$ \ n	2	3	4	5	6	7	8	9	10
0,70	1,190	1,533	1,773	1,941	2,059	2,141	2,199	2,239	2,266
0,71	1,214	1,572	1,826	2,006	2,134	2,225	2,289	2,334	2,366
0,72	1,238	1,611	1,880	2,073	2,212	2,312	2,384	2,436	2,473
0,73	1,263	1,652	1,926	2,133	2,284	2,394	2,474	2,531	2,574
0,74	1,287	1,692	1,992	2,214	2,378	2,499	2,589	2,655	2,707
0,75	1,312	1,734	2,050	2,287	2,465	2,598	2,698	2,773	2,829
0,76	1,337	1,776	2,110	2,363	2,556	2,702	2,813	2,897	2,961
0,77	1,363	1,819	2,170	2,441	2,649	2,809	2,932	3,027	3,100
0,78	1,388	1,862	2,232	2,520	2,745	2,920	3,057	3,164	3,247
0,79	1,414	1,907	2,296	2,604	2,847	3,039	3,191	3,311	3,406
0,80	1,440	1,952	2,362	2,690	2,952	3,162	3,330	3,464	3,571
0,81	1,466	1,997	2,427	2,776	3,058	3,287	3,472	3,622	3,744
0,82	1,492	2,043	2,495	2,866	3,170	3,419	3,623	3,790	3,927

$\bar{k}$ \ n	2	3	4	5	6	7	8	9	10
0,83	1,519	2,091	2,566	2,960	3,287	3,558	3,783	3,970	4,125
0,84	1,546	2,139	2,637	3,055	3,406	3,701	3,949	4,157	4,332
0,85	1,572	2,186	2,708	3,152	3,529	3,849	4,121	4,352	4,549
0,86	1,600	2,236	2,783	3,253	3,658	4,006	4,350	4,563	4,784
0,87	1,627	2,285	2,858	3,356	3,790	4,167	4,495	4,781	5,030
0,88	1,654	2,335	2,935	3,463	3,927	4,391	4,799	5,157	5,475
0,89	1,682	2,384	3,011	3,572	4,069	4,511	4,905	5,255	5,567
0,90	1,710	2,439	3,095	3,685	4,216	4,694	5,124	5,511	5,859
0,91	1,738	2,491	3,177	3,801	4,369	4,886	5,356	5,784	6,173
0,92	1,776	2,544	3,260	3,919	4,525	5,083	5,596	6,068	6,502
0,93	1,795	2,599	3,347	4,043	4,690	5,292	5,852	6,373	6,857
0,94	1,824	2,655	3,436	4,170	4,860	5,508	6,117	6,690	7,228
0,95	1,852	2,709	3,523	4,297	5,032	5,730	6,394	7,024	7,624
0,96	1,881	2,766	3,615	4,430	5,212	5,963	6,684	7,376	8,041
0,97	1,911	2,823	3,708	4,566	5,399	6,207	6,990	7,751	8,488
0,98	1,940	2,881	3,803	4,707	5,593	6,461	7,311	8,145	8,962
0,99	1,970	2,940	3,900	4,851	5,792	6,724	7,647	8,560	9,464
1,000	2,000	3,000	4,000	5,000	6,000	7,000	8,000	9,000	10,00
1,005	2,015	3,030	4,050	5,076	6,106	7,142	8,182	9,288	10,28
1,01	2,030	3,060	4,101	5,152	6,214	7,286	8,368	9,462	10,57
1,015	2,045	3,091	4,152	5,230	6,323	7,433	8,560	9,703	10,86
1,02	2,060	3,122	4,204	5,308	6,434	7,583	8,754	9,949	11,17
1,025	2,076	3,152	4,256	5,388	6,547	7,736	8,954	10,200	11,48
1,03	2,091	3,184	4,309	5,468	6,662	7,892	9,159	10,46	11,81
1,035	2,106	3,215	4,362	5,550	6,779	8,052	9,368	10,73	12,14
1,04	2,122	3,246	4,416	5,633	6,898	8,214	9,583	11,010	12,49
1,045	2,137	3,278	4,471	5,717	7,019	8,380	9,802	11,29	12,84
1,05	2,152	3,310	4,526	5,802	7,142	8,549	10,03	11,58	13,21
1,055	2,168	3,342	4,581	5,888	7,267	8,721	10,26	11,88	13,58
1,06	2,184	3,375	4,637	5,975	7,394	8,898	10,49	12,18	13,97
1,065	2,199	3,407	4,694	6,064	7,523	9,076	10,73	12,49	14,37
1,07	2,215	3,440	4,751	6,153	7,654	9,260	10,98	12,82	14,78
1,075	2,231	3,473	4,808	6,244	7,788	9,447	11,23	13,15	15,21
1,08	2,246	3,506	4,867	6,336	7,923	9,636	11,49	13,49	15,64
1,085	2,262	3,540	4,925	6,418	8,061	9,831	11,75	13,84	16,10
1,09	2,278	3,573	4,985	6,502	8,200	10,03	12,02	14,19	16,56
1,095	2,294	3,607	5,044	6,619	8,342	10,23	12,30	14,56	17,04
1,1	2,310	3,641	5,105	6,715	8,487	10,44	12,58	14,94	17,53

Примечание. Таблица дана в сокращенном виде.

СОДЕРЖАНИЕ